Also published in this series

Moral Concepts edited by Joel Feinberg
Theories of Ethics edited by Philippa Foot
Knowledge and Belief edited by A. Phillips Griffiths
The Philosophy of Mathematics edited by Jaakko Hintikka
The Philosophy of Science edited by P. H. Nidditch
The Theory of Meaning edited by G. H. R. Parkinson
Political Philosophy edited by Anthony Quinton
The Philosophy of Language edited by John R. Searle
Philosophical Logic edited by P. F. Strawson
The Philosophy of Perception edited by G. J. Warnock
The Philosophy of Action edited by Alan R. White

REFERENCE AND MODALITY

EDITED BY

LEONARD LINSKY

OXFORD UNIVERSITY PRESS
1971

in view of the facts that

N ⌜the number of planets = the number of planets⌝

and

Δ('the number of planets', nine).

At this point Kaplan introduces an idea already hinted at by Mrs. Marcus and Føllesdal, standard names. These are names '. . . which are so intimately connected with what they name that they could not but name it. I shall say that such a name *necessarily denotes* its object, and I shall use "Δ_N" to symbolize this more discriminating form of denotation.' This leads to replacement of (23) with

$\exists\alpha(\Delta_N(\alpha, \text{nine})$ and N⌜α is greater than five⌝)

as the analysis of (24), and replacement of (26) by

(27) $\exists\alpha(\Delta_N(\alpha, \text{nine})$ and N⌜α = the number of planets⌝)

as the analysis of (25). (27) has the advantage of being false. Similar difficulties affect Kaplan's Fregean treatment of belief and similar 'solutions' are available, except that another notion is needed for the propositional attitudes, that of a *vivid* name. The analysis in terms of standard names and vivid names, of course, does not avoid essentialism. It embraces it.

I

REFERENCE AND MODALITY

W. V. O. Quine

I

One of the fundamental principles governing identity is that of *substitutivity*—or, as it might well be called, that of *indiscernibility of identicals*. It provides that, *given a true statement of identity, one of its two terms may be substituted for the other in any true statement and the result will be true*. It is easy to find cases contrary to this principle. For example, the statements:

(1) Giorgione = Barbarelli,

(2) Giorgione was so-called because of his size

are true; however, replacement of the name 'Giorgione' by the name 'Barbarelli' turns (2) into the falsehood:

Barbarelli was so-called because of his size.

Furthermore, the statements:

(3) Cicero = Tully,

(4) 'Cicero' contains six letters

are true, but replacement of the first name by the second turns (4) false. Yet the basis of the principle of substitutivity appears quite solid; whatever can be said about the person Cicero (or Giorgione) should be equally true of the person Tully (or Barbarelli), this being the same person.

In the case of (4), this paradox resolves itself immediately. The fact is that (4) is not a statement about the person Cicero, but simply about the word 'Cicero'. The principle of substitutivity should not be extended to contexts in which the name to be supplanted occurs without referring simply to the object. Failure of substitutivity reveals merely that the

From *From a Logical Point of View*, by W. Quine (New York: Harper and Row, 1961), pp. 139–57. Reprinted by permission of the publishers of the cloth edition, Cambridge, Mass.: Harvard University Press,

occurrence to be supplanted is not *purely referential*,[1] that is, that the statement depends not only on the object but on the form of the name. For it is clear that whatever can be affirmed about the object remains true when we refer to the object by any other name.

An expression which consists of another expression between single quotes constitutes a name of that other expression; and it is clear that the occurrence of that other expression or a part of it, within the context of quotes, is not in general referential. In particular, the occurrence of the personal name within the context of quotes in (4) is not referential, not subject to the substitutivity principle. The personal name occurs there merely as a fragment of a longer name which contains, beside this fragment, the two quotation marks. To make a substitution upon a personal name, within such a context, would be no more justifiable than to make a substitution upon the term 'cat' within the context 'cattle'.

The example (2) is a little more subtle, for it is a statement about a man and not merely about his name. It was the man, not his name, that was called so and so because of his size. Nevertheless, the failure of substitutivity shows that the occurrence of the personal name in (2) is not *purely* referential. It is easy in fact to translate (2) into another statement which contains two occurrences of the name, one purely referential and the other not:

(5) Giorgione was called 'Giorgione' because of his size.

The first occurrence is purely referential. Substitution on the basis of (1) converts (5) into another statement equally true:

Barbarelli was called 'Giorgione' because of his size.

The second occurrence of the personal name is no more referential than any other occurrence within a context of quotes.

It would not be quite accurate to conclude that an occurrence of a name within single quotes is *never* referential. Consider the statements:

(6) 'Giorgione played chess' is true,

(7) 'Giorgione' named a chess player,

each of which is true or false according as the quotationless statement:

(8) Giorgione played chess

[1] Frege spoke of *direct* (*gerade*) and *oblique* (*ungerade*) occurrences, and used substitutivity of identity as a criterion just as here. See his 'On Sense and Reference', trans. Max Black, in *Philosophical Writings of Gottlob Frege* (Oxford: Blackwell, 1952).

is true or false. Our criterion of referential occurrence makes the occurrence of the name 'Giorgione' in (8) referential, and must make the occurrences of 'Giorgione' in (6) and (7) referential by the same token, despite the presence of single quotes in (6) and (7). The point about quotation is not that it must destroy referential occurrence, but that it can (and ordinarily does) destroy referential occurrence. The examples (6) and (7) are exceptional in that the special predicates 'is true' and 'named' have the effect of undoing the single quotes—as is evident on comparison of (6) and (7) with (8).

To get an example of another common type of statement in which names do not occur referentially, consider any person who is called Philip and satisfies the condition:

(9) Philip is unaware that Tully denounced Catiline,

or perhaps the condition:

(10) Philip believes that Tegucigalpa is in Nicaragua.

Substitution on the basis of (3) transforms (9) into the statement:

(11) Philip is unaware that Cicero denounced Catiline,

no doubt false. Substitution on the basis of the true identity:

Tegucigalpa = capital of Honduras

transforms the truth (10) likewise into the falsehood:

(12) Philip believes that the capital of Honduras is in Nicaragua.

We see therefore that the occurrences of the names 'Tully' and 'Tegucigalpa' in (9)–(10) are not purely referential.

In this there is a fundamental contrast between (9), or (10), and:

Crassus heard Tully denounce Catiline.

This statement affirms a relation between three persons, and the persons remain so related independently of the names applied to them. But (9) cannot be considered simply as affirming a relation between three persons, nor (10) a relation between person, city, and country—at least not so long as we interpret our words in such a way as to admit (9) and (10) as true and (11) and (12) as false.

Some readers may wish to construe unawareness and belief as relations between persons and statements, thus writing (9) and (10) in the manner:

(13) Philip is unaware of 'Tully denounced Catiline',

(14) Philip believes 'Tegucigalpa is in Nicaragua',

in order to put within a context of single quotes every not purely referential occurrence of a name. Church argues against this. In so doing he exploits the concept of analyticity, concerning which we have felt misgivings[1a]; still his argument cannot be set lightly aside, nor are we required here to take a stand on the matter. Suffice it to say that there is certainly no *need* to reconstrue (9)–(10) in the manner (13)–(14). What *is* imperative is to observe merely that the contexts 'is unaware that . . .' and 'believes that . . .' *resemble* the context of the single quotes in this respect: a name may occur referentially in a statement *S* and yet not occur referentially in a longer statement which is formed by embedding *S* in the context 'is unaware that . . .' or 'believes that . . .'. To sum up the situation in a word, we may speak of the contexts 'is unaware that . . .' and 'believes that . . .' as *referentially opaque*.[2] The same is true of the contexts 'knows that . . .', 'says that . . .', 'doubts that . . .', 'is surprised that . . .', etc. It would be tidy but unnecessary to force all referentially opaque contexts into the quotational mold; alternatively we can recognize quotation as one referentially opaque context among many.

It will next be shown that referential opacity afflicts also the so-called *modal* contexts 'Necessarily . . .' and 'Possibly . . .', at least when those are given the sense of *strict* necessity and possibility as in Lewis's modal logic.[3] According to the strict sense of 'necessarily' and 'possibly', these statements would be regarded as true:

(15) 9 is necessarily greater than 7,

(16) Necessarily if there is life on the Evening Star then there is life on the Evening Star,

(17) The number of planets is possibly less than 7,

and these as false:

(18) The number of planets is necessarily greater than 7,

(19) Necessarily if there is life on the Evening Star then there is life on the Morning Star,

(20) 9 is possibly less than 7.

[1a] See *From a Logical Point of View*, pp. 23–37.

[2] This term is roughly the opposite of Russell's 'transparent' as he uses it in his Appendix C to *Principia*, 2nd edn., Vol. 1.

[3] Lewis, C. I., *A Survey of Symbolic Logic* (New York: Dover, 1918) Ch. 5; Lewis and Langford, *Symbolic Logic* (New York, 1932; 2nd printing New York: Dover, 1951) pp. 78–89, 120–66.

The general idea of strict modalities is based on the putative notion of *analyticity* as follows: a statement of the form 'Necessarily . . .' is true if and only if the component statement which 'necessarily' governs is analytic, and a statement of the form 'Possibly . . .' is false if and only if the negation of the component statement which 'possibly' governs is analytic. Thus (15)–(17) could be paraphrased as follows:

(21) '9 > 7' is analytic,

(22) 'If there is life on the Evening Star then there is life on the Evening Star' is analytic,

(23) 'The number of planets is not less than 7' is not analytic,

and correspondingly for (18)–(20).

That the contexts 'Necessarily . . .' and 'Possibly . . .' are referentially opaque can now be quickly seen; for substitution on the basis of the true identities:

(24) The number of planets = 9,

(25) The Evening Star = the Morning Star

turns the truths (15)–(17) into the falsehoods (18)–(20).

Note that the fact that (15)–(17) are equivalent to (21)–(23), and the fact that '9' and 'Evening Star' and 'the number of planets' occur within quotations in (21)–(23), would not of themselves have justified us in concluding that '9' and 'Evening Star' and 'the number of planets' occur irreferentially in (15)–(17). To argue thus would be like citing the equivalence of (8) to (6) and (7) as evidence that 'Giorgione' occurs irreferentially in (8). What shows the occurrences of '9', 'Evening Star', and 'the number of planets' to be irreferential in (15)–(17) (and in (18)–(20)) is the fact that substitution by (24)–(25) turns the truths (15)–(17) into falsehoods (and the falsehoods (18)–(20) into truths).

Some, it was remarked, may like to think of (9) and (10) as receiving their more fundamental expression in (13) and (14). In the same spirit, many will like to think of (15)–(17) as receiving their more fundamental expression in (21)–(23).[4] But this again is unnecessary. We would certainly not think of (6) and (7) as somehow more basic than (8), and we need not view (21)–(23) as more basic than (15)–(17). What is important is to appreciate that the contexts 'Necessarily . . .' and 'Possibly . . .' are,

[4] Cf. Carnap, *The Logical Syntax of Language* (New York: Harcourt Brace; London: Routledge & Kegan Paul, 1937), pp. 245–59.

like quotation and 'is unaware that . . .' and 'believes that . . .', referentially opaque.

II

The phenomenon of referential opacity has just now been explained by appeal to the behaviour of singular terms. But singular terms are eliminable, we know, by paraphrase. Ultimately the objects referred to in a theory are to be accounted not as the things named by the singular terms, but as the values of the variables of quantification. So, if referential opacity is an infirmity worth worrying about, it must show symptoms in connection with quantification as well as in connection with singular terms.[5] Let us then turn our attention to quantification.

The connection between naming and quantification is implicit in the operation whereby, from 'Socrates is mortal', we infer '$(\exists x)$ (x is mortal)', that is, 'Something is mortal'. This is the operation which was spoken of earlier[5a] as *existential generalization*, except that we now have a singular term 'Socrates' where we then had a free variable. The idea behind such inference is that whatever is true of the object named by a given singular term is true of something; and clearly the inference loses its justification when the singular term in question does not happen to name. From:

There is no such thing as Pegasus,

for example, we do not infer:

$(\exists x)$ (there is no such thing as x),

that is, 'There is something which there is no such thing as', or 'There is something which there is not'.

Such inference is of course equally unwarranted in the case of an irreferential occurrence of any substantive. From (2), existential generalization would lead to:

$(\exists x)$ (x was so-called because of its size),

that is, 'Something was so-called because of its size'. This is clearly meaningless, there being no longer any suitable antecedent for 'so-called'. Note, in contrast, that existential generalization with respect to the purely referential occurrence in (5) yields the sound conclusion:

$(\exists x)$ (x was called 'Giorgione' because of its size),

that is, 'Something was called 'Giorgione' because of its size'.

[5] Substantially this point was made by Church, in *Journal of Symbolic Logic*, 7 (1942), 100 ff.

[5a] *From a Logical Point of View*, p. 120.

The logical operation of *universal instantiation* is that whereby we infer from 'Everything is itself', for example, or in symbols '(x) $(x = x)$', the conclusion that Socrates = Socrates. This and existential generalization are two aspects of a single principle; for instead of saying that '(x) $(x = x)$' implies 'Socrates = Socrates', we could as well say that the denial 'Socrates ≠ Socrates' implies '$(\exists x)$ $(x \neq x)$'. The principle embodied in these two operations is the link between quantifications and the singular statements that are related to them as instances. Yet it is a principle only by courtesy. It holds only in the case where a term names and, furthermore, occurs referentially. It is simply the logical content of the idea that a given occurrence is referential. The principle is, for this reason, anomalous as an adjunct to the purely logical theory of quantification. Hence the logical importance of the fact that all singular terms, aside from the variables that serve as pronouns in connection with quantifiers, are dispensable and eliminable by paraphrase.[6]

We saw just now how the referentially opaque context (2) fared under existential generalization. Let us see what happens to our other referentially opaque contexts. Applied to the occurrence of the personal name in (4), existential generalization would lead us to:

(26) $(\exists x)$ ('x' contains six letters),

that is:

(27) There is something such that 'it' contains six letters,

or perhaps:

(28) 'Something' contains six letters.

Now the expression:

'x' contains six letters

means simply:

The 24th letter of the alphabet contains six letters.

In (26) the occurrence of the letter within the context of quotes is as irrelevant to the quantifier that precedes it as is the occurrence of the same letter in the context 'six'. (26) consists merely of a falsehood preceded by an irrelevant quantifier. (27) is similar; its part:

'it' contains six letters

[6] See *From a Logical Point of View*, pp. 7f, 13, and 166f. Note that existential generalization as of p. 120 does belong to pure quantification theory, for it has to do with free variables rather than singular terms. The same is true of a correlative use of universal instantiation, such as is embodied in R2 of Essay V, ibid.

is false, and the prefix 'there is something such that' is irrelevant. (28), again, is false—if by 'contains six' we mean 'contains exactly six'.

It is less obvious, and correspondingly more important to recognize, that existential generalization is unwarranted likewise in the case of (9) and (10). Applied to (9), it leads to:

($\exists x$) (Philip is unaware that x denounced Catiline),

that is:

(29) Something is such that Philip is unaware that it denounced Catiline.

What is this object, that denounced Catiline without Philip's having become aware of the fact? Tully, that is, Cicero? But to suppose this would conflict with the fact that (11) is false.

Note that (29) is not to be confused with:

Philip is unaware that ($\exists x$) (x denounced Catiline),

which, though it happens to be false, is quite straightforward and in no danger of being inferred by existential generalization from (9).

Now the difficulty involved in the apparent consequence (29) of (9) recurs when we try to apply existential generalization to modal statements. The apparent consequences:

(30) ($\exists x$) (x is necessarily greater than 7),

(31) ($\exists x$) (necessarily if there is life on the Evening Star then there is life on x)

of (15) and (16) raise the same questions as did (29). What is this number which, according to (30), is necessarily greater than 7? According to (15), from which (30) was inferred, it was 9, that is, the number of planets; but to suppose this would conflict with the fact that (18) is false. In a word, to be necessarily greater than 7 is not a trait of a number, but depends on the manner of referring to the number. Again, what is the thing x whose existence is affirmed in (31)? According to (16), from which (31) was inferred, it was the Evening Star, that is, the Morning Star; but to suppose this would conflict with the fact that (19) is false. Being necessarily or possibly thus and so is in general not a trait of the object concerned, but depends on the manner of referring to the object.

Note that (30) and (31) are not to be confused with:

Necessarily ($\exists x$) ($x > 7$),

Necessarily ($\exists x$) (if there is life on the Evening Star then there is life on x),

which present no problem of interpretation comparable to that presented by (30) and (31). The difference may be accentuated by a change of example: in a game of a type admitting of no tie it is necessary that some one of the players will win, but there is no one player of whom it may be said to be necessary that he win.

We had seen, in the preceding section, how referential opacity manifests itself in connection with singular terms; and the task which we then set ourselves at the beginning of this section was to see how referential opacity manifests itself in connection rather with variables of quantification. The answer is now apparent: if to a referentially opaque context of a variable we apply a quantifier, with the intention that it govern that variable from outside the referentially opaque context, then what we commonly end up with is unintended sense or nonsense of the type (26)–(31). In a word, we cannot in general properly *quantify into* referentially opaque contexts.

The context of quotation and the further contexts '. . . was so called', 'is unaware that . . .', 'believes that . . .', 'Necessarily . . .', and 'Possibly . . .' were found referentially opaque in the preceding section by consideration of the failure of substitutivity of identity as applied to singular terms. In the present section these contexts have been found referentially opaque by a criterion having to do no longer with singular terms, but with the miscarriage of quantification. The reader may feel, indeed, that in this second criterion we have not really got away from singular terms after all; for the discrediting of the quantifications (29)–(31) turned still on an expository interplay between the singular terms 'Tully' and 'Cicero', '9' and 'the number of planets', 'Evening Star' and 'Morning Star'. Actually, though, this expository reversion to our old singular terms is avoidable, as may now be illustrated by re-arguing the meaninglessness of (30) in another way. Whatever is greater than 7 is a number, and any given number x greater than 7 can be uniquely determined by any of various conditions, some of which have '$x > 7$' as a *necessary* consequence and some of which do not. One and the same number x is uniquely determined by the condition:

(32) $x = \sqrt{x} + \sqrt{x} + \sqrt{x} \neq \sqrt{x}$

and by the condition:

(33) There are exactly x planets,

but (32) has '$x > 7$' as a necessary consequence while (33) does not. *Necessary* greaterness than 7 makes no sense as applied to a *number* x; necessity attaches only to the connection between '$x > 7$' and the particular method (32), as opposed to (33), of specifying x.

Similarly, (31) was meaningless because the sort of thing x which fulfils the condition:

(34) If there is life on the Evening Star then there is life on x,

namely, a physical object, can be uniquely determined by any of various conditions, not all of which have (34) as a necessary consequence. *Necessary* fulfilment of (34) makes no sense as applied to a physical object x; necessity attaches, at best, only to the connection between (34) and one or another particular means of specifying x.

The importance of recognizing referential opacity is not easily overstressed. We saw in §1 that referential opacity can obstruct substitutivity of identity. We now see that it also can interrupt quantification: quantifiers outside a referentially opaque construction need have no bearing on variables inside it. This again is obvious in the case of quotation, as witness the grotesque example:

$(\exists x)$ ('*six*' contains 'x').

III

We see from (30)–(31) how a quantifier applied to a modal sentence may lead simply to nonsense. Nonsense is indeed mere absence of sense, and can always be remedied by arbitrarily assigning some sense. But the important point to observe is that granted an understanding of the modalities (through uncritical acceptance, for the sake of argument, of the underlying notion of analyticity), and given an understanding of quantification ordinarily so called, we do not come out automatically with any meaning for quantified modal sentences such as (30)–(31). This point must be taken into account by anyone who undertakes to work out laws for a quantified modal logic.

The root of the trouble was the referential opacity of modal contexts. But referential opacity depends in part on the ontology accepted, that is, on what objects are admitted as possible objects of reference. This may be seen most readily by reverting for a while to the point of view of §1, where referential opacity was explained in terms of failure of interchangeability of names which name the same object. Suppose now we were to repudiate all objects which, like 9 and the planet Venus, or Evening Star, are nameable by names which fail of interchangeability in modal contexts. To do so would be to sweep away all examples indicative of the opacity of modal contexts.

But what objects would remain in a thus purified universe? An object x must, to survive, meet this condition: if S is a statement containing a

referential occurrence of a name of x, and S' is formed from S by substituting any different name of x, then S and S' not only must be alike in truth value as they stand, but must stay alike in truth value even when 'necessarily' or 'possibly' is prefixed. Equivalently: putting one name of x for another in any analytic statement must yield an analytic statement. Equivalently: any two names of x must be synonymous.[7]

Thus the planet Venus as a material object is ruled out by the possession of heteronymous names 'Venus', 'Evening Star', 'Morning Star'. Corresponding to these three names we must, if modal contexts are not to be referentially opaque, recognize three objects rather than one—perhaps the Venus-concept, the Evening-Star-concept, and the Morning-Star-concept.

Similarly 9, as a unique whole number between 8 and 10, is ruled out by the possession of heteronymous names '9' and 'the number of the planets'. Corresponding to these two names we must, if modal contexts are not to be referentially opaque, recognize two objects rather than one; perhaps the 9-concept and the number-of-planets-concept. These concepts are not numbers, for the one is neither identical with nor less than nor greater than the other.

The requirement that any two names of x be synonymous might be seen as a restriction not on the admissible objects x, but on the admissible vocabulary of singular terms. So much the worse, then, for this way of phrasing the requirement; we have here simply one more manifestation of the superficiality of treating ontological questions from the vantage point of singular terms. The real insight, in danger now of being obscured, was rather this: necessity does not properly apply to the fulfilment of conditions by *objects* (such as the ball of rock which is Venus, or the number which numbers the planets), apart from special ways of specifying them. This point was most conveniently brought out by consideration of singular terms, but it is not abrogated by their elimination. Let us now review the matter from the point of view of quantification rather than singular terms.

From the point of view of quantification, the referential opacity of modal contexts was reflected in the meaninglessness of such quantifications as (30)–(31). The crux of the trouble with (30) is that a number x may be uniquely determined by each of two conditions, for example, (32) and (33), which are not necessarily, that is, analytically, equivalent to each other. But suppose now we were to repudiate all such objects and retain only objects x such that *any two conditions uniquely determining*

[7] See *From a Logical Point of View*, p. 32. Synonymy of names does not mean merely naming the same thing; it means that the statement of identity formed of the two names is analytic.

2

x are analytically equivalent. All examples such as (30)–(31), illustrative of the referential opacity of modal contexts, would then be swept away. It would come to make sense in general to say that there is an object which, independently of any particular means of specifying it, is necessarily thus and so. It would become legitimate, in short, to quantify into modal contexts.

Our examples suggest no objection to quantifying into modal contexts as long as the values of any variables thus quantified are limited to *intensional objects*. This limitation would mean allowing, for purposes of such quantification anyway, not classes but only class-concepts or attributes, it being understood that two open sentences which determine the same class still determine distinct attributes unless they are analytically equivalent. It would mean allowing, for purposes of such quantification, not numbers but only some sort of concepts which are related to the numbers in a many-one way. Further it would mean allowing, for purposes of such quantification, no concrete objects but only what Frege[7a] called senses of names, and Carnap[7b] and Church have called individual concepts. It is a drawback of such an ontology that the principle of individuation of its entities rests invariably on the putative notion of synonymy, or analyticity.

Actually, even granted these dubious entities, we can quickly see that the expedient of limiting the values of variables to them is after all a mistaken one. It does not relieve the original difficulty over quantifying into modal contexts; on the contrary, examples quite as disturbing as the old ones can be adduced within the realm of intensional objects. For, where A is any intensional object, say an attribute, and 'p' stands for an arbitrary true sentence, clearly

(35) $A = (\iota x)[p \,.\, (x = A)]$.

Yet, if the true sentence represented by 'p' is not analytic, then neither is (35), and its sides are no more interchangeable in modal contexts than are 'Evening Star' and 'Morning Star', or '9' and 'the number of the planets'.

Or, to state the point without recourse to singular terms, it is that the requirement lately italicized—'any two conditions uniquely determining x are analytically equivalent'—is not assured merely by taking x as an intensional object. For, think of 'Fx' as any condition uniquely determining x, and think of 'p' as any nonanalytic truth. Then '$p \,.\, Fx$' uniquely determines x but is not analytically equivalent to 'Fx', even though x be an intensional object.

[7a] 'On Sense and Reference.'
[7b] *Meaning and Necessity*, 2nd edn. (Chicago, Ill.: Univ. of Chicago Press, 1956).

It was in my 1943 paper that I first objected to quantifying into modal contexts, and it was in his review of it that Church proposed the remedy of limiting the variables thus quantified to intensional values. This remedy, which I have just now represented as mistaken, seemed all right at the time. Carnap adopted it in an extreme form, limiting the range of his variables to intensional objects throughout his system. He did not indeed describe his procedure thus; he complicated the picture by propounding a curious double interpretation of variables. But I have argued[8] that this complicating device has no essential bearing and is better put aside.

By the time Church came to propound an intensional logic of his own,[8a] he perhaps appreciated that quantification into modal contexts could not after all be legitimized simply by limiting the thus quantified variables to intensional values. Anyway his departures are more radical. Instead of a necessity operator attachable to sentences, he has a necessity predicate attachable to complex names of certain intensional objects called propositions. What makes this departure more serious than it sounds is that the constants and variables occurring in a sentence do not recur in Church's name of the corresponding proposition. Thus the interplay, usual in modal logic, between occurrences of expressions outside modal contexts and recurrences of them inside modal contexts, is ill reflected in Church's system. Perhaps we should not call it a system of modal logic; Church generally did not. Anyway let my continuing discussion be understood as relating to modal logics only in the narrower sense, where the modal operator attaches to sentences.

Church and Carnap tried—unsuccessfully, I have just argued—to meet my criticism of quantified modal logic by restricting the values of their variables. Arthur Smullyan took the alternative course of challenging my criticism itself. His argument depends on positing a fundamental division of names into proper names and (overt or covert) descriptions, such that proper names which name the same object are always synonymous. (Cf. (38) below.) He observes quite rightly on these assumptions, that any examples which, like (15)–(20) and (24)–(25), show failure of substitutivity of identity in modal contexts, must exploit some descriptions rather than just proper names. Then he undertakes to adjust matters by propounding, in connection with modal contexts, and alteration of

[8] In a criticism which Carnap generously included in his *Meaning and Necessity*, pp. 196f.
[8a] Church, A., 'A Formulation of the Logic of Sense and Denotation', in Henle, P., Kallen, H. M., and Langen, S. K. (eds.), *Structure, Method and Meaning: Essays in Honour of Henry M. Sheffer* (New York: Liberal Arts Press, 1951), pp. 3–24.

Russell's familiar logic of descriptions.[9] As stressed in the preceding section, however, referential opacity remains to be reckoned with even when descriptions and other singular terms are eliminated altogether.

Nevertheless, the only hope of sustaining quantified modal logic lies in adopting a course that resembles Smullyan's, rather than Church and Carnap, in this way: it must overrule my objection. It must consist in arguing or deciding that quantification into modal contexts makes sense even though any value of the variable of such a quantification be determinable by conditions that are not analytically equivalent to each other. The only hope lies in accepting the situation illustrated by (32) and (33) and insisting, despite it, that the object x is question is necessarily greater than 7. This means adopting an invidious attitude toward certain ways of uniquely specifying x, for example (33), and favouring other ways, for example (32), as somehow better revealing the 'essence' of the object. Consequences of (32) can, from such a point of view, be looked upon as necessarily true of the object which is 9 (and is the number of the planets), while some consequences of (33) are rated still as only contingently true of that object.

Evidently this reversion to Aristotelian essentialism[10] is required if quantification into modal contexts is to be insisted on. An object, of itself and by whatever name or none, must be seen as having some of its traits necessarily and others contingently, despite the fact that the latter traits follow just as analytically from some ways of specifying the object as the former traits do from other ways of specifying it. In fact, we can see pretty directly that any quantified modal logic is bound to show such favouritism among the traits of an object; for surely it will be held, for each thing x, on the one hand that

(36) necessarily $(x = x)$

and on the other hand that

(37) $\sim$ necessarily $[p \,.\, (x = x)]$,

where 'p' stands for an arbitrary contingent truth.

Essentialism is abruptly at variance with the idea, favoured by Carnap, Lewis, and others, of explaining necessity by analyticity[11]. For the appeal

[9] Russell's theory of descriptions, in its original formulation, involved distinctions of so-called 'scope'. Change in the scope of a description was indifferent to the truth value of any statement, however, unless the description failed to name. This indifference was important to the fulfilment, by Russell's theory, of its purpose as an analysis or surrogate of the practical idiom of singular description. On the other hand, Smullyan allows difference of scope to affect truth value even in cases where the description concerned succeeds in naming.

[10] *From a Logical Point of View*, p. 22.

[11] Ibid., p. 30.

to analyticity can pretend to distinguish essential and accidental traits of an object only relative to how the object is specified, not absolutely. Yet the champion of quantified modal logic must settle for essentialism.

Limiting the values of his variables is neither necessary nor sufficient to justify quantifying the variables into modal contexts. Limiting their values can, however, still have this purpose in conjunction with his essentialism: if he wants to limit his essentialism to special sorts of objects, he must correspondingly limit the values of the variables which he quantifies into modal contexts.

The system presented in Miss Barcan's pioneer papers on quantified modal logic differed from the system of Carnap and Church in imposing no special limitations on the values of variables. That she was prepared, moreover, to accept the essentialist presuppositions seems rather hinted in her theorem:

(38) $(x)(y)\{(x = y) \supset [\text{necessarily } (x = y)]\}$,

for this is as if to say that some at least (and in fact at most; cf. '$p \,.\, Fx$') of the traits that determine an object do so necessarily. The modal logic in Fitch, follows Miss Barcan on both points. Note incidentally that (38) follows directly from (36) and a law of substitutivity of identity for variables:

$(x)(y)[(x = y \,.\, Fx) \supset Fy]$.

The upshot of these reflections is meant to be that the way to do quantified modal logic, if at all, is to accept Aristotelian essentialism. To defend Aristotelian essentialism, however, is not part of my plan. Such a philosophy is as unreasonable by my lights as it is by Carnap's or Lewis's. And in conclusion I say, as Carnap and Lewis have not: so much the worse for quantified modal logic. By implication, so much the worse for unquantified modal logic as well; for, if we do not propose to quantify across the necessity operator, the use of that operator ceases to have any clear advantage over merely quoting a sentence and saying that it is analytic.

IV

The worries introduced by the logical modalities are introduced also by the admission of attributes (as opposed to classes). The idiom 'the attribute of being thus and so' is referentially opaque, as may be seen, for example, from the fact that the true statement:

(39) The attribute of exceeding 9 = the attribute of exceeding 9 goes over into the falsehood:

The attribute of exceeding the number of the planets = the attribute of exceeding 9 under substitution according to the true identity (24). Moreover, existential generalization of (39) would lead to:

(40) $(\exists x)$ (the attribute of exceeding x = the attribute of exceeding 9)

which resists coherent interpretation just as did the existential generalizations (29)–(31) of (9), (15), and (16). Quantification of a sentence which contains the variable of quantification within a context of the form 'the attribute of . . .' is exactly on a par with quantification of a modal sentence.

Attributes, as remarked earlier, are individuated by this principle: two open sentences which determine the same class do not determine the same attribute unless they are analytically equivalent. Now another popular sort of intensional entity is the *proposition*. Propositions are conceived in relation to statements as attributes are conceived in relation to open sentences: two statements determine the same proposition just in case they are analytically equivalent. The foregoing strictures on attributes obviously apply to propositions. The truth:

(41) The proposition that $9 > 7$ = the proposition that $9 > 7$ goes over into the falsehood:

The proposition that the number of the planets > 7 = the proposition that $9 > 7$.

under substitution according to (24). Existential generalization of (41) yields a result comparable to (29)–(31) and (40).

Most of the logicians, semanticists, and analytical philosophers who discourse freely of attributes, propositions, or logical modalities betray failure to appreciate that they thereby imply a metaphysical position which they themselves would scarcely condone. It is noteworthy that in *Principia Mathematica*, where attributes were nominally admitted as entities, all actual contexts occurring in the course of formal work are such as could be fulfilled as well by classes as by attributes. All actual contexts are *extensional* in the sense already explained.[12] The authors of *Principia Mathematica* thus adhered in practice to a principle of extensionality which they did not espouse in theory. If their practice had been otherwise, we might have been brought sooner to an appreciation of the urgency of the principle.

We have seen how modal sentences, attribute terms, and proposition terms conflict with the nonessentialist view of the universe. It must be

[12] *From a Logical Point of View*, p. 30.

kept in mind that those expressions create such conflict only when they are quantified into, that is, when they are put under a quantifier and themselves contain the variable of quantification. We are familiar with the fact (illustrated by (26) above) that a quotation cannot contain an effectively free variable, reachable by an outside quantifier. If we preserve a similar attitude toward modalities, attribute terms, and proposition terms, we may then make free use of them without any misgivings of the present urgent kind.

What has been said of modality in these pages relates only to strict modality. For other sorts, for example, physical necessity and possibility, the first problem would be to formulate the notions clearly and exactly. Afterwards we could investigate whether such modalities, like the strict ones, cannot be quantified into without precipitating an ontological crisis. The question concerns intimately the practical use of language. It concerns, for example, the use of the contrary-to-fact conditional within a quantification; for it is reasonable to suppose that the contrary-to-fact conditional reduces to the form 'Necessarily, if p then q' in some sense of necessity. Upon the contrary-to-fact conditional depends in turn, for example, this definition of solubility in water: To say that an object is soluble in water is to say that it would dissolve if it were in water. In discussions of physics, naturally, we need quantifications containing the clause 'x is soluble in water', or the equivalent in words; but, according to the definition suggested, we should then have to admit within quantifications the expression 'if x were in water then x would dissolve', that is, 'necessarily if x is in water then x dissolves'. Yet we do not know whether there is a suitable sense of 'necessarily' into which we can so quantify.[13]

Any way of imbedding statements within statements, whether based on some notion of 'necessity' or, for example, on a notion of 'probability' as in Reichenbach, must be carefully examined in relation to its susceptibility to quantification. Perhaps the only useful modes of statement composition susceptible to unrestricted quantification are the truth functions. Happily, no other mode of statement composition is needed, at any rate, in mathematics; and mathematics, significantly, is the branch of science whose needs are most clearly understood.

Let us return, for a final sweeping observation, to our first test of referential opacity, namely, failure of substitutivity of identity; and let us suppose that we are dealing with a theory in which (a) *logically* equivalent formulas are interchangeable in all contexts *salva veritate* and (b) the logic of classes

[13] For a theory of disposition terms, like 'soluble', see Carnap, 'Testability and Meaning', *Philosophy of Science*, 3 (1936), 419–71.

is at hand.[14] For such a theory it can be shown that *any* mode of statement composition, other than the truth functions, is referentially opaque. For, let ϕ and ψ be any statements alike in truth value, and let $\Phi(\phi)$ be any true statement containing ϕ as a part. What is to be shown is that $\Phi(\psi)$ will also be true, unless the context represented by 'Φ' is referentially opaque. Now the class named by $\hat{\alpha}\phi$ is either V or Λ, according as ϕ is true or false; for remember that ϕ is a statement, devoid of free α. (If the notation $\hat{\alpha}\phi$ without recurrence of α seems puzzling, read it as $\hat{\alpha}(\alpha = \alpha \,.\, \phi)$.) Moreover ϕ is logically equivalent to $\hat{\alpha}\phi = V$. Hence, by (a), since $\Phi(\phi)$ is true, so is $\Phi(\hat{\alpha}\phi = V)$. But $\hat{\alpha}\phi$ and $\hat{\alpha}\psi$ name one and the same class, since ϕ and ψ are alike in truth value. Then, since $\Phi(\hat{\alpha}\phi = V)$ is true, so is $\Phi(\hat{\alpha}\psi = V)$ unless the context represented by 'Φ' is referentially opaque. But if $\Phi(\hat{\alpha}\psi = V)$ is true, then so in turn is $\Phi(\psi)$, by (a).

[14] See *From a Logical Point of View*, pp. 27, 87.

II

MODALITY AND DESCRIPTION

ARTHUR F. SMULLYAN

THERE are logicians who maintain that modal logic violates Leibniz's principle that if x and y are identical, then y has every property of x. The alleged difficulty is illustrated in the following example due to Quine.[1]

(a) It is logically necessary that 9 is less than 10.
(b) 9 = the number of the planets.
(c) Therefore, it is logically necessary that the number of the planets is less than 10.

The premisses of this argument are true, the conclusion is false, and yet the conclusion appears to be derived by means of the logical precept that if x is y then any property of x is a property of y. Such is the paradox of modal logic. But the difficulty is obviated if we draw a distinction. We must distinguish between statements of the following forms:

(d) The so-and-so satisfies the condition that it is necessary that *Fx*.
and
(e) It is necessary that the so-and-so satisfies the condition that *Fx*.

The reader at this stage is bound to feel as though he were being asked to distinguish between Tweedledum and Tweedledee. Possibly, it will be of assistance to him to remark that statements of type (d) are sometimes synthetic, whereas those of type (e) are never synthetic. I will ask the reader to believe that James is now thinking of the number 3. If, now, some one were to remark, 'there is one and only one integer which James is now thinking of and that integer is necessarily odd', then he would be stating a contingent truth. For that there is just one integer which James now thinks of, is only an empirical fact. This statement could just as well be expressed in the form, (d), 'The integer, which James is now

[1] W. Quine, 'Notes on Existence and Necessity', *Journal of Philosophy*, XL (1943), 113–27.
We assume, at this stage of the discussion, that '9' and '10', as they occur in the illustration, are *names* of familiar logical *properties*. This is in order to simplify the introductory discussion. In due course we shall consider the problem in a more general way.

III

EXTENSIONALITY

RUTH B. MARCUS

THE continued development of intensional logics, and concern with problems of their interpretation has had a rather curious effect. It has reinforced the notion, unjustifiable in my opinion, that extensionality is an unambiguous concept. This presumed clarity is usually singled out as the virtue of extensional systems, to say nothing of their metaphysical advantages. The assertion that in mathematics and empirical science one does not *need* to traffic in non-extensional notions which are fuzzy and troublesome, has become a virtual platitude. Yet a cursory examination of the literature does not reveal any well-defined theory of extensionality, although it is possible to find a core of agreement. Indeed, there are differences as to (a) what are the principles of extensionality, (b) which objects are or ought to be extensional, and (c) which formal systems are extensional.

My purpose in this paper is to arrive at a characterization of extensionality in terms of these differences which may be helpful in connection with some familiar problems of interpreting intensional systems.

Principles of extensionality. Consider first some unspecified system of material implication L with theory of types. On the propositional level, extensionality takes the form of a substitution principle:

(1) If p is equivalent$_1$ to q then A is equivalent$_2$ to B,

where B is the result of replacing one or more occurrences of p in A by q.

As stated, (1) is of course ambiguous. The ambiguity concerns the meaning of 'equivalence$_1$' and 'equivalence$_2$'. A minimal requirement of an equivalence relation is that it be reflexive, transitive and symmetrical. These conditions are met by a variety of relations ranging from identity to having the same weight, and further interpretation is required. Our concern is with logically definable relations of equivalence.

Using the abbreviations 'eq$_1$' and 'eq$_2$', let us first consider principles

From *Mind*, n.s., 69 (1960), 55–62. Reprinted by permission of the author and the Editor of *Mind*.

in which eq$_1$ and eq$_2$ have the same meaning. If they are taken as identity, (1) becomes

(1.1) If pIq then AIB (where 'I' names the identity relation)

and is merely explicative of the notion of identity. Suppose what is intended is

(1.2) If '$p \equiv q$' is a tautology then '$A \equiv B$' is a tautology.

In what sense is (1.2) an extensionality principle? Only in that it eliminates as possible predicates of propositions, certain intensional predicates such as 'believed by John'. Not all intensional predicates are precluded by (1.2). In particular, modal predicates such as 'logically necessary' would not falsify (1.2). Ordinarily, variables which range over predicates of propositions are dispensable, and consequently (1.1) is often provable as a strong form of the substitution theorem.

Most commonly, eq$_1$ and eq$_2$ are interpreted as material equivalence without the modifying condition of (1.2):

(1.3) If $p \equiv q$ then $A \equiv B$.

Here (1) is taken to mean that if p and q have the same truth value, whether contingently or necessarily, then A and B have the same truth value. As contrasted with (1.2), (1.3) is a strongly extensional principle for it disallows all intensional predicates of propositions. Here again, where variables which range over propositional predicates are not introduced, (1.3) is provable as the substitution theorem.

Consider, again on the level of propositions, principles in which eq$_1$ and eq$_2$ are not the same. If eq$_1$ is taken as identity and eq$_2$ material equivalence, then (1) becomes

(1.4) If pIq then $A \equiv B$.

(1.4) like (1.1) is explicative of the identity relation. [The converse of (1.4) is of course another matter involving as it does the assumption of Leibniz's law, in addition to being strongly extensional.]

If we take eq$_1$ as material equivalence and eq$_2$ as identity, we have

(1.5) If $p \equiv q$ then AIB

which in the first instance, where A is p, becomes

(1.51) If $p \equiv q$ then pIq.

In contrast to (1.4), (1.5) is very strongly extensional since it not only eliminates intensional predicates of propositions but assimilates propositions to truth values.

What I am trying to make apparent by this necessarily crude and informal analysis, is that even on the level of propositions, we cannot talk of *the* thesis of extensionality but only of stronger and weaker extensionality principles. I will call a principle *extensional* if it either (a) *directly or indirectly imposes restrictions on the possible values of the functional variables such that some intensional functions are prohibited or* (b) *it has the consequence of equating identity with a weaker form of equivalence.* Obviously (a) and (b) are interdependent. On the basis of this characterization, (1.2), (1.3), and (1.5) are all principles of extensionality, in order of increasing strength. It should now be clear why there is often disagreement as to whether a given formal system is or is not extensional. There is, for example, a literature of tiresome arguments as to whether the formal system of *Principia* is extensional. It is all a matter of deciding how extensional a system must be to be properly so-called. There are by contrast, logicians such as Alonzo Church who does talk in terms of degrees of extensionality. A more reasonable approach would be to assert, in connection with *Principia*, that the formal system as interpreted in the first edition is less extensional than the interpretation proposed by the second edition, since the latter assumes an analogue of (1.5) which is stronger than (1.2).

Consider next another set of principles which are more frequently associated with the theory of extensionality. Principles which relate the equivalence of classes (or attributes) to the equivalence of their defining functions.

(2) If $(x)(F(x) \text{ eq}_1\, G(x))$ then $F \text{ eq}_2\, G$.

In addition to the interpretation of eq_1 and eq_2, one must specify whether F and G are predicate variables, class variables, or non-committal functional variables. I cannot give an exhaustive account of the many possible variations of (2). In a weakly extensional system, eq_1 might be taken as tautological equivalence, eq_2 as identity, F and G functional variables, as follows:

(2.1) If $(x)(F(x) \equiv G(x))$ is tautological, then *FIG*.

By the criterion of extensionality stated above, (2.1) is weakly extensional since it precludes some intensional contexts. On the other hand (2.1) permits us to state an identity between the terms '9' and '3^2' but not '9' and (on the assumption that it can be construed as an expression of proper type level) 'the number of planets'. A stronger alternative (referred to most often as *the* extensionality principle) asserts identity of functions as a consequence of formal equivalence.

(2.2) If $(x)(F(x) \equiv G(x))$ then *FIG*.

In languages which distinguish classes from attributes, the distinction is sometimes maintained by postulating (2.2) for classes and perhaps (2.1) for attributes. This has the effect of eliminating intensional contexts involving class names but allowing such contexts for attribute names. Such is the interpreted procedure of *Principia.* The concept of identity in *Principia* is systematically ambiguous not only as prescribed by the theory of types, but on the same type level. In the second order predicate calculus, 'identity' means something different for classes than for attributes, and has still another import for individuals. My preference is for the alternative procedure of giving uniform meaning to 'identity' and to talk of attributes and classes as being equal, but not identical. Functional *equality* would be defined as

(3) $(F = G) \rightarrow_{\text{df}} (x)(F(x) \equiv G(x))$

where $F = G$ is *not* equated to *FIG*. On the basis of known substitution theorems, the substitution of F for G in strongly extensional contexts is still permissible and in such contexts $F = G$ is *like FIG*. This permits us to say that the class of mermaids and the class of Greek gods are equal but not identical and that in strongly extensional contexts (arithmetic ones for example) we are concerned only with their equality, so that the name of one may be substituted for the other.

It seems to me that much of the discussion these past few years concerning apparent breakdowns of substitutivity principles in intensional contexts and its presumably devastating results for logic and mathematics are largely terminological. I am not (as Quine[1] insists in his review of two of my papers on quantified modal logic) proposing that there be more than one kind of identity, but only that the distinctions between stronger and weaker equivalences be made explicit before, for one avowed purpose or another, they are obliterated.

The usual reason given for reducing identity to equality [(3) (2.2)] is that it provides a simpler base for mathematics, mathematics being concerned with aggregates discussed in truth functional contexts, not with predicates in intensional contexts. Under such restrictive conditions, the substitution theorem can generally be proved for equal (formally equivalent) classes, with the result that equality functions *as* identity.

Establishing the foundations of mathematics is not the only purpose of logic, particularly if the assumptions deemed convenient for mathematics do violence to both ordinary and philosophical usage. I am not

[1] W. V. O. Quine, *Journal of Symbolic Logic*, 12 (1947), 95–6.

disturbed by the possibility of equal, non-identical classes or attributes, *e.g.* man and featherless biped. To me it seems reasonable that there are many empty classes of the same type, *e.g.* mermaids and Greek gods, equal but not identical. And why should the non-identity of the numbers n and $n+1$ depend on the enumeration of different things in the world? To subsume mathematics under logic is not to equate them. A much broader base is indicated in the direction of intensional systems such as the modal logics. I will try to show that the apparent difficulties of interpreting such systems are not genuine, but analogous to a rejection of a non-Euclidean geometry because it allows parallel lines to meet.

To complete this analysis, I must consider another set of modifications of identity involving Leibniz's law.

The identity of indiscernibles. Sometimes identity is introduced as

(4) If $(\phi)(\phi(x)$ eq $\phi(y))$ then xIy.

The usual interpretation of eq is as material equivalence:

(4.1) If $(\phi)(\phi(x) \equiv \phi(y))$ then xIy.

Another possibility is in terms of tautological equivalence.

(4.2) If $(\phi)(\phi(x) \equiv \phi(x))$ is a tautology, then xIy.

The converse of (4) is of course explicative of identity. The status of (4) itself is not quite so clear. It has also been accepted as a truism and merely explicative. However, I do not regard Ramsey's[2] reservations about (4) as entirely spurious. He objected to taking (4.1) as definitive of identity on the ground that it is logically possible for two things to have all their properties in common and still be two. Such a possibility is excluded by (4.1). To argue that if they are two then they are distinguishable as having two different names will not do for Ramsey, since they may be unknown, unnamed, and still two.

According to our characterization of extensionality, instances of (4) may therefore be interpreted as extensionality principles in that they equate identity with the slightly weaker relation of indiscernibility which requires that to be distinct means to be *discernibly* distinct.[3]

[2] F. P. Ramsey, *The Foundations of Mathematics* (London: Routledge & Kegan Paul; and New York: Harcourt Brace, 1931), pp. 30–2.

[3] I am aware that interpreting (4) in this way is somewhat paradoxical since (4) has the effect of establishing a logical priority of the concept of 'property' over that of 'class' whereas in an extensional system the emphasis is held to be on the class concept. If such usage is intolerable the characterization of extensionality can be appropriately modified.

Interpreting intensional systems. Quine[4] states: 'When modal logic is extended (as by Miss Barcan) to include quantification theory, . . . serious obstacles to interpretation are encountered.' These difficulties revolve about the substitution of equivalences in contexts involving 'knows that', 'is aware that', and in particular 'is necessary that', and 'is possible that'. Quine describes such contexts as being referentially opaque. It is the point of this paper to show that the opacity lies with Quine's use of such terms as 'identity', 'true identity', 'equality'. The above analysis leads to the dissolution of at least some of the problems of interpretation associated with intensional contexts.

Among the equivalence relations which can be introduced into L are identity, indiscernibility, tautological equivalence, material equivalence. On the functional level, these are listed in order of decreasing strength, for there is some model, some permissible range of values, which prevents our equating them except by explicit postulate. It should be noted that for variables of lowest type, there are only identity and indiscernibility. Indeed a recent paper of Bergmann[5] on *Individuals* may be understood as an attempt to explicate the notion of individuals as those entities for which only the strongest equivalence relation holds.

Consider now modal functional calculi such as my[6] extension of the Lewis systems. In such languages (2.1) can be stated directly as

(5) If $N((x)(F(x) \equiv G(x)))$ then FIG (where N is interpreted as logical necessity)

(4.2) becomes

(6) If $N((\phi)(\phi(x) \equiv \phi(y)))$ then xIy

and (1.1) is

(7) If $N(P \equiv Q)$ then PIQ.

[4] W. Quine, 'The Problem of Interpreting Modal Logic', *Journal of Symbolic Logic*, 12 (1947), 43–8.

[5] G. Bergmann, 'Individuals', *Philosophical Studies*, IX (1958), 78–85. (My interpretation of this paper rests on the assumption that the statement of (Ext) involves a typographical error.)

[6] R. C. Barcan (Marcus): 'A Functional Calculus of First Order Based on Strict Implication', *Journal of Symbolic Logic*, 11 (1946), 1–16; 'The Deduction Theorem in a Functional Calculus of First Order Based on Strict Implication', ibid. 115–18; 'The Identity of Individuals in a Strict Functional Calculus of Second Order', ibid. 12, 12–15.

Within these extended systems, I have been able to prove theorems which relate different kinds of equivalence. It is possible to show

(8) Given $P \equiv Q$, P is not everywhere interchangeable with Q, but only in restricted non-modal contexts. Given $N(P \equiv Q)$, then the substitution theorem is unrestricted.

This theorem has the effect of prohibiting the substitution of 'Socrates is a featherless biped' for 'Socrates is a man' in 'It is necessary that if Socrates is a man then Socrates is a man'. That the substitution theorem for strict equivalence differs from the theorem for material equivalence, is not paradoxical, but a more adequate formalization of a known distinction.

It is when Quine[7] refers to

(9) The number of planets equals nine

as a 'true identity', without hint of ambiguity that we become aware that his fundamental criticism is directed not toward presumed paradoxes but toward the intensional point of view. As indicated above (9) is *not* unambiguous except in a strongly extensional language.

Let us assume for the moment that '9' and 'the number of planets' are expressions of the same type level and can meaningfully be equated. Quite apart from interpretation in terms of the theory of descriptions, within the modal language the problem revolves about substituting 'the number of planets' for '9' in

(10) $N(9 > 7)$.

But such a substitution is prohibited by (8), for (9) does not assert tautological equivalence, and the substitution would have to be made within the scope of a modal operator. The paradox evaporates. By the same token, since

(11) $N(9 = (5+4))$, '5+4' can replace '9' in (10).

The problem of the Morning Star and the Evening Star is resolved in an analogous way.[8] For, like (9),

[7] W. Quine, *From a Logical Point of View* (Cambridge, Mass.: Harvard Univ. Press, 1953), p. 144. [See above, p. 21].

[8] The paragraph which follows restates a point made by F. B. Fitch in 'The Problem of the Morning Star and the Evening Star', *Philosophy of Science*, XVI (1949), 137–40.

(12) The Evening Star equals the Morning Star

is not unambiguous. If (12) involves proper names of individuals then 'the Evening Star' may replace 'the Morning Star' without paradox in

(13) It is necessary that the Evening Star is the Evening Star

for the only equivalence relation between individuals are identity and indiscernibility. Indeed, although it appears as if (4.1) and (4.2) express two kinds of indiscernibility, they can be *proved* strictly equivalent within a modal system. Quine's[9] failure to note the latter in his review of my paper had the unfortunate result of perpetuating a non-existent paradox.

If, on the other hand, (12) is about classes or properties, then it states a non-tautological equality, not an identity, and consequently, the conditions of the substitution theorem (8) prevent the substitution of 'the Morning Star' for one of the occurrences of 'the Evening Star' in (13). At the risk of too much repetition, we are not asserting that the substitution *ought* not to be made on the basis of some pre-formal analysis, but that they are prohibited by the theorems provable[10] in such extended systems.

I have tried in this brief paper, to characterize the theory of extensionality, and to show that logical systems are more or less extensional. Their extensionality depends on the kinds of contexts and predicates which are prohibited, and the degree to which the relation of identity is equated to weaker forms of equivalence. I also tried to show that a more broadly based logic in the direction of modalities need not do violence to the foundations of mathematics, and the supposed paradoxes involved in interpreting such intensional systems are not genuine.

[9] See n. 4 above. Quine's failure to notice that (4.1) and (4.2) are materially equivalent in $s2^2$ and strictly equivalent in $s4^2$ and $s5^2$ leads him to conclude that modal logic must deal with individual concepts rather than individuals. In a recent letter Quine forwarded copies of a note to the editor of *The Journal of Symbolic Logic*, and his publisher correcting the error.

[10] The substitution principle (8) is a rough restatement of substitution theorems for some of the extended modal calculi. The theorems are proved at the end of the first paper listed in n. 6 p. 49 above.

objects, that is that '$\boxdot F$', if true of an object, is true of it regardless of the way in which it is referred to, then one should better not quantify into causal contexts; one should avoid contrary-to-fact conditionals, scientific law-statements, confirmation statements, and many types of probability statements and disposition terms—if one wants to make sense.

V

SEMANTICAL CONSIDERATIONS ON MODAL LOGIC

SAUL A. KRIPKE

THIS paper gives an exposition of some features of a semantical theory of modal logics.[1] For a certain quantified extension of S5, this theory was presented in 'A Completeness Theorem in Modal Logic',[2] and it has been summarized in 'Semantical Analysis of Modal Logic'.[3] The present paper will concentrate on one aspect of the theory—the introduction of quantifiers—and it will restrict itself in the main to one method of achieving this end. The emphasis of the paper will be purely semantical, and hence it will omit the use of semantic tableaux, which is essential to a full presentation of the theory.[4] Proofs, also, will largely be suppressed.

We consider four modal systems. Formulae A, B, C, . . . are built out of atomic formulae P, Q, R, . . ., using the connectives $\wedge$, $\sim$, and $\Box$. The system M has the following axiom schemes and rules:

A0. Truth-functional tautologies
A1. $\Box A \supset A$
A2. $\Box(A \supset B) \supset . \Box A \supset \Box B$
R1. $A, A \supset B/B$
R2. $A/\Box A$

If we add the following axiom scheme, we get S4:

$$\Box A \supset \Box\Box A$$

From *Acta Philosophica Fennica*, 16 (1963), 83–94. Reprinted by permission of the author and the publishers, Societas Philosophica Fennica, Helsinki.

[1] The theory given here has points of contact with many authors: For lists of these, see S. Kripke, 'Semantical Analysis of Modal Logic', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9 (1963), 67–96, and J. Hintikka, 'Modality and Quantification' *Theoria*, 27 (1961) 119–28. The authors closest to the present theory appear to be Hintikka and Kanger. The present treatment of quantification, however, is unique as far as I know, although it derives some inspiration from acquaintance with the very different methods of Prior and Hintikka.

[2] *Journal of Symbolic Logic*, 24 (1959), 1–15.

[3] Ibid., pp. 323–4 (Abstract).

[4] For these see 'A Completeness Theorem in Modal Logic', *Journal of Symbolic Logic*, 24 (1959), 1–15 and 'Semantical Analysis of Modal Logic', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9, 67–96.

We get the *Brouwersche* system if we add to M:

$A \supset \Box\Diamond A$
S5, if we add:
$\Diamond A \supset \Box\Diamond A$

Modal systems whose theorems are closed under the rules R1 and R2, and include all theorems of M, are called 'normal'. Although we have developed a theory which applies to such non-normal systems as Lewis's S2 and S3, we will restrict ourselves here to normal systems.

To get a semantics for modal logic, we introduce the notion of a (normal) *model structure*. A model structure (m.s.) is an ordered triple $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$ where $\mathbf{K}$ is a set, $\boldsymbol{R}$ is a reflexive relation on $\mathbf{K}$, and $\mathbf{G} \varepsilon \mathbf{K}$. Intuitively, we look at matters thus: $\mathbf{K}$ is the set of all 'possible worlds'; $\mathbf{G}$ is the 'real world'. If $\mathbf{H}_1$ and $\mathbf{H}_2$ are two worlds, $\mathbf{H}_1\boldsymbol{R}\,\mathbf{H}_2$ means intuitively that $\mathbf{H}_2$ is 'possible relative to' $\mathbf{H}_1$; *i.e.*, that every proposition *true* in $\mathbf{H}_2$ is *possible* in $\mathbf{H}_1$. Clearly, then, the relation $\boldsymbol{R}$ should indeed be reflexive; every world $\mathbf{H}$ is *possible* relative to itself, since every proposition *true* in $\mathbf{H}$ is, *a fortiori*, possible in $\mathbf{H}$. Reflexivity is thus an intuitively natural requirement. We may impose additional requirements, corresponding to various 'reduction axioms' of modal logic: If $\boldsymbol{R}$ is transitive, we call $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$ an S4-m.s.; if $\boldsymbol{R}$ is symmetric, $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$ is a *Brouwersche* m.s.; and if $\boldsymbol{R}$ is an equivalence relation, we call $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$ an S5-m.s. A model structure without restriction is also called an M-model structure.

To complete the picture, we need the notion of *model*. Given a model structure $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$, a *model* assigns to each atomic formula (propositional variable) P a truth-value $\mathbf{T}$ or $\mathbf{F}$ in each world $\mathbf{H} \varepsilon \mathbf{K}$. Formally, a *model* φ on a m.s. $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$ is a binary function $\varphi(P, \mathbf{H})$, where P varies over atomic formulae and $\mathbf{H}$ varies over elements of $\mathbf{K}$, whose range is the set $\{\mathbf{T}, \mathbf{F}\}$. Given a model, we can define the assignments of truth-values to non-atomic formulae by induction. Assume $\varphi(A, \mathbf{H})$ and $\varphi(B, \mathbf{H})$ have already been defined for all $\mathbf{H} \varepsilon \mathbf{K}$. Then if $\varphi(A, \mathbf{H}) = \varphi(B, \mathbf{H}) = \mathbf{T}$, define $\varphi(A \wedge B, \mathbf{H}) = \mathbf{T}$; otherwise, $\varphi(A \wedge B, \mathbf{H}) = \mathbf{F}$. $\varphi(\sim A, \mathbf{H})$ is defined to be $\mathbf{F}$ iff $\varphi(A, \mathbf{H}) = \mathbf{T}$; otherwise, $\varphi(\sim A, \mathbf{H}) = \mathbf{T}$. Finally, we define $\varphi(\Box A, \mathbf{H}) = \mathbf{T}$ iff $\varphi(A, \mathbf{H}') = \mathbf{T}$ for every $\mathbf{H}' \varepsilon \mathbf{K}$ such that $\mathbf{H}\,\boldsymbol{R}\,\mathbf{H}'$; otherwise, $\varphi(\Box A, \mathbf{H}) = \mathbf{F}$. Intuitively, this says that A is necessary in $\mathbf{H}$ iff A is true in all worlds $\mathbf{H}'$ possible relative to $\mathbf{H}$.

Completeness theorem. $\vdash A$ in M (S4, S5, the *Brouwersche* system) if and only if $\varphi(A, \mathbf{G}) = \mathbf{T}$ for every model φ on an M-(S4-, S5-, *Brouwersche*) model structure $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$.[5]

[5] For a proof, see 'Semantical Analysis . . .,' *Zeitschrift* . . ., 9.

This completeness theorem equates the syntactical notion of *provability* in a modal system with a semantical notion of *validity*.

The rest of this paper concerns, with the exception of some concluding remarks, the introduction of quantifiers. To do this, we must associate with each world a domain of individuals, the individuals that exist in that world. Formally, we define a *quantificational model structure* (q.m.s.) as a model structure $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$, together with a function ψ which assigns to each $\mathbf{H} \varepsilon \mathbf{K}$ a set $\psi(\mathbf{H})$, called the *domain* of $\mathbf{H}$. Intuitively $\psi(\mathbf{H})$ is the set of all individuals existing in $\mathbf{H}$. Notice, of course, that $\psi(\mathbf{H})$ need not be the same set for different arguments $\mathbf{H}$, just as, intuitively, in worlds other than the real one, some actually existing individuals may be absent, while new individuals, like Pegasus, may appear.

We may then add, to the symbols of modal logic, an infinite list of individual variables $x, y, z, \ldots$, and, for each nonnegative integer n, a list of n-adic predicate letters $P^n, Q^n, \ldots$, where the superscripts will sometimes be understood from the context. We count propositional variables (atomic formulae) as '0-adic' predicate letters. We then build up well-formed formulae in the usual manner, and can now prepare ourselves to define a quantificational *model*.

To define a quantificational model, we must extend the original notion, which assigned a truth-value to each atomic formula in each world. Analogously, we must suppose that in each world a given n-adic predicate letter determines a certain set of ordered n-tuples, its *extension* in that world. Consider, for example, the case of a monadic predicate letter $P(x)$. We would like to say that, in the world $\mathbf{H}$, the predicate $P(x)$ is true of some individuals in $\psi(\mathbf{H})$ and false of others; formally, we would say that, relative to certain assignments of elements of $\psi(\mathbf{H})$ to x, $\varphi(P(x), \mathbf{H}) = \mathbf{T}$ and relative to others $\varphi(P(x), \mathbf{H}) = \mathbf{F}$. The set of all individuals of which P is true is called the *extension* of P in $\mathbf{H}$. But there is a problem: should $\varphi(P(x), \mathbf{H})$ be given a truth-value when x is assigned a value in the domain of some *other* world $\mathbf{H}'$ and not in the domain of $\mathbf{H}$? Intuitively, suppose $P(x)$ means 'x is bald'—are we to assign a truth-value to the substitution instance 'Sherlock Holmes is bald'? Holmes does not exist, but in other states of affairs, he would have existed. Should we assign a definite truth-value to the statement that he is bald, or not? Frege[6] and Strawson[7]

[6] G. Frege 'Uber Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik*, 100 (1892), 25–50. English translations in Geach and Black, *Translations from the Philosophical Writings of Gottlob Frege*, (Oxford: Blackwell, 1952), and in Feigl and Sellars (eds.), *Readings in Philosophical Analysis* (New York: Appleton Century Crofts, 1949).
[7] P. F. Strawson, 'On referring', *Mind*, n.s., 59 (1950), 320–44.

would not assign the statement a truth-value; Russell would.[8] For the purposes of modal logic we hold that different answers to this question represent alternative *conventions*. All are tenable. The only existing discussions of this problem I have seen—those of Hintikka[9] and Prior[10]—adopt the Frege-Strawson view. This view necessarily must lead to some modification of the usual modal logic. The reason is that the semantics for modal propositional logic, which we have already given, assumed that every formula must take a truth-value in each world; and now, for a formula $A(x)$ containing a free variable x, the Frege-Strawson view requires that it not be given a truth-value in a world **H** when the variable x is assigned an individual not in the domain of that world. We thus can no longer expect that the original laws of modal propositional logic hold for statements containing free variables, and are faced with an option: either revise modal propositional logic or restrict the rule of substitution. Prior does the former, Hintikka the latter. There are further alternatives the Frege-Strawson choice involves: Should we take $\Box A$ (in **H**) to mean that A is *true* in all possible worlds (relative to **H**), or just *not false* in any such world? The second alternative merely demands that A be either true or lack a truth-value in each world. Prior, in his system Q, in effect admits both types of necessity, one as 'L' and the other as 'NMN'. A similar question arises for conjunction: if A is false and B has no truth-value, should we take $A \wedge B$ to be false or truth-valueless?

In a full statement of the semantical theory, we would explore all these variants of the Frege-Strawson view. Here we will take the other option, and assume that a statement containing free variables has a truth-value in each world for every assignment to its free variables.[11] Formally, we state the matter as follows: Let $\mathbf{U} = \bigcup_{\mathbf{H} \in \mathbf{K}} \psi(\mathbf{H})$. $\mathbf{U}^n$ is the nth Cartesian product of **U** with itself. We define a quantificational *model* on a q.m.s. $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$ as a binary function $\varphi(P^n, \mathbf{H})$, where the first variable ranges over n-adic predicate letters, for arbitrary n, and **H** ranges over elements of **K**. If $n = 0$, $\varphi(P^n, \mathbf{H}) = \mathbf{T}$ or **F**; if $n \geq 1$, $\varphi(P^n, \mathbf{H})$ is a subset of $\mathbf{U}^n$. We now define, inductively, for every formula A and $\mathbf{H} \varepsilon \mathbf{K}$, a truth-value $\varphi(A, \mathbf{H})$, relative to a given assignment of elements of **U** to the free variables of A. The case of a propositional variable is obvious. For an atomic formula $P^n(x_1, \ldots, x_n)$, where P^n is an n-adic predicate letter and $n \geq 1$, given an assignment of elements $a_1, \ldots, a_n$ of **U** to $x_1, \ldots, x_n$, we define $\varphi(P^n(x_1, \ldots, x_n), \mathbf{H}) = \mathbf{T}$ if the n-tuple $(a_1, \ldots, a_n)$ is a member of $\varphi(P^n, \mathbf{H})$; otherwise, $\varphi(P^n(x_1, \ldots, x_n), \mathbf{H}) = \mathbf{F}$, relative to the given assignment. Given these assignments for atomic formulae, we can build up the assignments for complex formulae by induction. The induction steps for the propositional connectives $\wedge$, $\sim$, $\Box$, have already been given. Assume we have a formula $A(x, y_1, \ldots, y_n)$, where x and the y_i are the only free variables present, and that a truth-value $\varphi(A(x, y_1, \ldots, y_n), \mathbf{H})$ has been defined for each assignment to the free variables of $A(x, y_1, \ldots, y_n)$. Then we define $\varphi((x)A(x, y_1, \ldots, y_n), \mathbf{H}) = \mathbf{T}$ relative to an assignment of $b_1, \ldots, b_n$ to $y_1, \ldots, y_n$ (where the b_i are elements of **U**), if $\varphi((A(x, y_1, \ldots, y_n), \mathbf{H}) = \mathbf{T}$ for every assignment of $a, b_1, \ldots, b_n$ to $x, y_1, \ldots, y_n$, respectively, where $a \varepsilon \psi(\mathbf{H})$; otherwise, $\varphi((x)A(x, y_1, \ldots, y_n), \mathbf{H}) = \mathbf{F}$ relative to the given assignment. Notice that the restriction $a \varepsilon \psi(\mathbf{H})$ means that, in **H**, we quantify only over the objects actually existing in **H**.

To illustrate the semantics, we give counterexamples to two familiar proposals for laws of modal quantification theory—the 'Barcan formula' $(x)\Box A(x) \supset \Box(x)A(x)$ and its converse $\Box(x)A(x) \supset (x)\Box A(x)$. For each we consider a model structure $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$, where $\mathbf{K} = \{\mathbf{G}, \mathbf{H}\}$, $\mathbf{G} \neq \mathbf{H}$, and $\boldsymbol{R}$ is simply the Cartesian product $\mathbf{K}^2$. Clearly $\boldsymbol{R}$ is reflexive, transitive, and symmetric, so our considerations apply even to S5.

For the Barcan formula, we extend $(\mathbf{G}, \mathbf{K}, \boldsymbol{R})$ to a quantificational model structure by defining $\psi(\mathbf{G}) = \{a\}$, $\psi(\mathbf{H}) = \{a, b\}$, where a and b are distinct. We then define, for a monadic predicate letter P, a model φ in which $\varphi(P, \mathbf{G}) = \{a\}$, $\varphi(P, \mathbf{H}) = \{a\}$. Then clearly $\Box P(x)$ is true in **G** when x is assigned a; and since a is the only object in the domain of **G**, so is $(x)\Box P(x)$. But, $(x)P(x)$ is clearly false in **H** (for $\varphi(P(x), \mathbf{H}) = \mathbf{F}$ when x is assigned b), and hence $\Box(x)P(x)$ is false in **G**. So we have a counterexample to the Barcan formula. Notice that this counterexample is quite independent of whether $P(x)$ is assigned a truth-value in **G** or not when x is assigned 6, so also it applies to the systems of Hintikka and Prior. Such counterexamples can be disallowed, and the Barcan formula reinstated, only if we require a model structure to satisfy the condition that $\psi(\mathbf{H}') \subseteq \psi(\mathbf{H})$ whenever $\mathbf{H}\boldsymbol{R}\mathbf{H}'$ $(\mathbf{H}, \mathbf{H}' \varepsilon \mathbf{K})$.

For the converse of the Barcan formula, set $\psi(\mathbf{G}) = \{a, b\}$, $\psi(\mathbf{H}) = \{a\}$,

[8] Bertrand Russell, 'On denoting', *Mind*, n.s., 14 (1905), 479–93.

[9] 'Modality and Quantification'.

[10] A. N. Prior, *Time and Modality* (Oxford: Clarendon Press, 1957, VIII+148 pp.)

[11] It is natural to assume that an *atomic* predicate should be *false* in a world **H** of all those individuals not existing in that world; that is, that the extension of a predicate letter must consist of actually existing individuals. We can do this by requiring semantically that $\varphi(P^n, \mathbf{H})$ be a subset of $[\psi(\mathbf{H})]^n$; the semantical treatment below would otherwise suffice without change. We would have to add to the axiom system below all closures of formulae of the form $P^n(x_1, \ldots, x_n) \wedge (y)A(y) . \supset . A(x_i)$ $(1 \leq i \leq n)$. We have chosen not to do this because the rule of substitution would no longer hold; theorems would hold for atomic formulae which would not hold when the atomic formulae are replaced by arbitrary formulae. (This answers a question of Putnam and Kalmar.)

where again $a \neq b$. Define $\varphi(P, \mathbf{G}) = \{a, b\}$, $\varphi(P, \mathbf{H}) = \{a\}$, where P is a given monadic predicate letter. Then clearly $(x)P(x)$ holds in both $\mathbf{G}$ and $\mathbf{H}$, so that $\varphi(\Box(x)P(x), \mathbf{G}) = \mathbf{T}$. But $\varphi(P(x), \mathbf{H}) = \mathbf{F}$ when x is assigned b, so that, when x is assigned b, $\varphi(\Box P(x), \mathbf{G}) = \mathbf{F}$. Hence $\varphi((x)\Box P(x), \mathbf{G}) = \mathbf{F}$, and we have the desired counterexample to the converse of the Barcan formula. This counterexample, however, depends on asserting that, in $\mathbf{H}$, $P(x)$ is actually *false* when x is assigned b; it might thus disappear if, for this assignment, $P(x)$ were declared to lack truth-value in $\mathbf{H}$. In this case, we will still have a counterexample if we require a necessary statement to be *true* in all possible worlds (Prior's 'L'), but not if we merely require that it never be false (Prior's 'NMN'). On our present convention, we can eliminate the counterexample only by requiring, for each q.m.s., that $\psi(\mathbf{H}) \subseteq \psi(\mathbf{H}')$ whenever $\mathbf{H}R\mathbf{H}'$.

These counterexamples lead to a peculiar difficulty: We have given countermodels, in quantified S5, to both the Barcan formula and its converse. Yet Prior appears to have shown[12] that the Barcan formula is derivable in quantified S5; and the converse seems derivable even in quantified M by the following argument:

(A) $(x)A(x) \supset A(y)$ (by quantification theory)
(B) $\Box((x)A(x) \supset A(y))$ (by necessitation)
(C) $\Box((x)A(x) \supset A(y)) \supset \Box(x)A(x) \supset \Box A(y)$ (Axiom A2)
(D) $\Box(x)A(x) \supset \Box A(y)$ (from (B) and (C))
(E) $(y)(\Box(x)A(x) \supset \Box A(y))$ (generalizing on (D))
(F) $\Box(x)A(x) \supset (y)\Box A(y)$ (by quantification theory, and (E))

We seem to have derived the conclusion using principles that should all be valid in the model-theory. Actually, the flaw lies in the application of necessitation to (A). In a formula like (A), we give the free variables the

[12] See 'Modality and Quantification in s5', *Journal of Symbolic Logic*, 21 (1956), 60–2.

[13] It is not asserted that the generality interpretation of theorems with free variables is the only possible one. One might wish a formula A to be provable iff, for each model φ, $\varphi(A, \mathbf{G}) = \mathbf{T}$ for every assignment to the free variables of A. But then $(x)A(x) \supset A(y)$ will not be a theorem; in fact, in the countermodel above to the Barcan formula, $\varphi((x)P(x) \supset P(y), \mathbf{G}) = \mathbf{F}$ if y is assigned b. Thus quantification theory would have to be revised along the lines proposed by Hintikka (in 'Existential Presuppositions and Existential Commitments', *Journal of Philosophy*, 56 (1959), 125–37) and by H. Leblanc and T. Hailperin (in 'Nondesignating Singular Terms', *Philosophical Review*, 68 (1959), 239–43). This procedure has much to recommend it, but we have not adopted it since we wished to show that the difficulty can be solved without revising quantification theory or modal propositional logic.

generality interpretation:[13] When (A) is asserted as a theorem, it abbreviates assertion of its ordinary universal closure

(A′) $(y)((x)A(x) \supset A(y))$
Now if we applied necessitation to (A′), we would get
(B′) $\Box(y)((x)A(x) \supset A(y))$
On the other hand, (B) itself is interpreted as asserting
(B″) $(y)\Box((x)A(x) \supset A(y))$

To infer (B″) from (B′), we would need a law of the form $\Box(y)C(y) \supset (y)\Box C(y)$, which is just the converse Barcan formula that we are trying to prove. In fact, it is readily checked that (B″) fails in the countermodel given above for the converse Barcan formula, if we replace $A(x)$ by $P(x)$.

We can avoid this sort of difficulty if, following Quine[14] we formulate quantification theory so that only *closed* formulae are asserted. Assertion of formulae containing free variables is at best a convenience; assertion of $A(x)$ with free x can always be replaced by assertion of $(x)A(x)$.

If A is a formula containing free variables, we define a *closure* of A to be any formula without free variables obtained by prefixing universal quantifiers and necessity signs, in any order, to A. We then define the axioms of quantified M to be the closures of the following schemata:

(0) Truth-functional tautologies
(1) $\Box A \supset A$
(2) $\Box(A \supset B) . \supset . \Box A \supset \Box B$
(3) $A \supset (x)A$, where x is not free in A
(4) $(x)(A \supset B) . \supset . (x)A \supset (x)B$
(5) $(y)((x)A(x) \supset A(y))$

The rule of inference is detachment for material implication. Necessitation can be obtained as a derived rule.

To obtain quantified extensions of S4, S5, the *Brouwersche* system, simply add to the axiom schemata all closures of the appropriate reduction axiom.

The systems we have obtained have the following properties: They are a straightforward extension of the modal propositional logics, without the modifications of Prior's Q; the rule of substitution holds without restriction, unlike Hintikka's presentation; and nevertheless neither the Barcan formula nor its converse is derivable. Further, all the laws of quantification theory—modified to admit the empty domain—hold.

[14] W. Quine, *Mathematical Logic* (Cambridge, Mass.: Harvard Univ. Press, 1940; 2nd edn., rev., 1951, xii+346 pp.).

The semantical completeness theorem we gave for modal propositional logic can be extended to the new systems.

We can introduce *existence as a predicate* in the present system if we like. Semantically, existence is a monadic predicate $E(x)$ satisfying, for each model φ on a m.s. (**G**, **K**, $\boldsymbol{R}$), the identity $\varphi(E, \mathbf{H}) = \psi(\mathbf{H})$ for every $\mathbf{H} \varepsilon \mathbf{K}$. Axiomatically, we can introduce it through the postulation of closures of formulae of the form: $(x)A(x) \wedge E(y) . \supset . A(y)$, and $(x)E(x)$. The predicate P used above in the counter-example to the converse Barcan formula can now be recognized as simply existence. This fact shows how existence differs from the tautological predicate $A(x) \vee \sim A(x)$ even though $\Box(x)E(x)$ is provable. For although $(x)\Box(A(x) \vee \sim A(x))$ is valid, $(x)\Box E(x)$ is not; although it is necessary that every thing exists, it does not follow that everything has the property of necessary existence.

We can introduce identity semantically in the model theory by defining $x = y$ to be true in a world **H** when x and y are assigned the same value and otherwise false; existence could then be defined in terms of identity, by stipulating that $E(x)$ means $(\exists y)(x = y)$. For reasons not given here, a broader theory of identity could be obtained if we complicated the notion of quantificational model structure.

We conclude with some brief and sketchy remarks on the 'provability' interpretations of modal logics, which we give in each case for propositional calculus only. The reader will have obtained the main point of this paper if he omits this section. Provability interpretations are based on a desire to adjoin a necessity operator to a formal system, say Peano arithmetic, in such a way that, for any formula A of the system, $\Box A$ will be interpreted as true iff A is provable in the system. It has been argued that such 'provability' interpretations of a model operator are dispensable in favour of a provability *predicate*, attaching to the Gödel number of A; but Professor Montague's contribution to the present volume casts at least some doubt on this viewpoint.

Let us consider the formal system **PA** of Peano arithmetic, as formalized in Kleene.[15] We adjoin to the formation rules operators $\wedge$, $\sim$, and $\Box$ (the conjunction and negation adjoined are to be distinct from those of the original system), operating on closed formulae only. In the model theory we gave above, we took atomic formulae to be propositional variables, or predicate letters followed by parenthesized individual variables; here we take them to be simply the closed well-formed formulae of **PA** (*not* just the atomic formulae of **PA**). We define a model structure (**G**, **K**, $\boldsymbol{R}$), where **K** is the set of all distinct (non-isomorphic) countable models of

[15] S. C. Kleene, *Introduction to Metamathematics* (New York: D. Van Nostrand, 1952, x+550 pp.).

PA, **G** is the standard model in the natural numbers, and $\boldsymbol{R}$ is the Cartesian product $\mathbf{K}^2$. We define a model φ by requiring that, for any atomic formula P and $\mathbf{H} \varepsilon \mathbf{K}$, $\varphi(P, \mathbf{H}) = \mathbf{T}(\mathbf{F})$ iff P is true (false) in the model **H**. (Remember, P is a wff of **PA**, and **H** is a countable model of **PA**.) We then build up the evaluation for compound formulae as before.[16] To say that A is true is to say it is true in the real world **G**; and, for any atomic P, $\varphi(\Box P, \mathbf{G}) = \mathbf{T}$ iff P is provable in **PA**. (Notice that $\varphi(P, \mathbf{G}) = \mathbf{T}$ iff P is true in the intuitive sense). Since (**G**, **K**, $\boldsymbol{R}$) is an S5-m.s., all the laws of S5 will be valid on this interpretation; and we can show that *only* the laws of S5 are generally valid. (For example, if P is Gödel's undecidable formula, $\varphi(\Box P \vee \Box \sim P, \mathbf{G}) = \mathbf{F}$, which is a counterexample to the 'law' $\Box A \vee \Box \sim A$.)

Another provability interpretation is the following: Again we take the atomic formulae to be the closed wffs of **PA**, and then build up new formulae using the adjoined connectives $\wedge$, $\sim$, and $\Box$. Let **K** be the set of all ordered pairs (**E**, α), where **E** is a consistent extension of **PA**, and α is a (countable) model of the system **E**. Let **G** = (**PA**, α_0), where α_0 is the standard model of **PA**. We say (**E**, α) $\boldsymbol{R}$ (**E**′, α'), where (**E**, α) and (**E**′, α') are in **K**, iff **E**′ is an extension of **E**. For atomic P, define $\varphi(P, (\mathbf{E}, \alpha)) = \mathbf{T}(\mathbf{F})$ iff P is true (false) in α. Then we can show, for atomic P, that $\varphi(\Box P, (\mathbf{E}, \alpha)) = \mathbf{T}$ iff P is provable in **E**; in particular, $\varphi(\Box P, \mathbf{G}) = \mathbf{T}$ iff P is provable in **PA**. Since (**G**, **K**, $\boldsymbol{R}$) is an S4-m.s., all the laws of S4 hold. But not all the laws of S5 hold; if P is Gödel's undecidable formula, $\varphi((\sim \Box P \supset \Box \sim \Box P), \mathbf{G}) = \mathbf{F}$. But some laws are valid which are not provable in S4; in particular, we can prove for any A, $\varphi(\Box \sim \Box(\Diamond A \wedge \Diamond \sim A), \mathbf{G}) = \mathbf{T}$, which yields the theorems of McKinsey's S4.1.[17] By suitable modifications this difficulty could be removed; but we do not go into the matter here.

Similar interpretations of M and the *Brouwersche* system could be stated; but, in the present writer's opinion, they have less interest than those given above. We mention one more class of provability interpretations, the 'reflexive' extensions of **PA**. Let **E** be a formal system containing

[16] It may be protested that **PA** already contain symbols for conjunction and negation, say '&' and '¬'; so why do we adjoin new symbols '$\wedge$' and '$\sim$'? The answer is that if P and Q are atomic formulae, then P & Q is *also* atomic in the present sense, since it is well-formed in **PA**; but $P \wedge Q$ is not. In order to be able to apply the previous theory, in which the conjunction of atomic formulae is not atomic, we need '$\wedge$'. Nevertheless, for any $\mathbf{H} \varepsilon \mathbf{K}$ and atomic P and Q, $\varphi(P \& Q, \mathbf{H}) = \varphi(P \wedge Q, \mathbf{H})$, so that confusion of '&' with '$\wedge$' causes no harm in practice. Similar remarks apply to negation, and to the provability interpretation of s4 in the next paragraph.

[17] See J. C. C. McKinsey, 'On the Syntactical Construction of Systems of Modal Logic', *Journal of Symbolic Logic*, 10 (1945), 83–94.

PA, and whose well-formed formulae are formed out of the closed formulae of **PA** by use of the connectives &, $\neg$, and $\Box$. (I say '&' and '$\neg$' to indicate that I am using the same conjunction and negation as in **PA** itself, not introducing new ones. See footnote 16, p. 71.) Then **E** is called a reflexive extension of **PA** iff: (1) It is an inessential extension of **PA**; (2) $\Box A$ is provable in **E** iff A is; (3) there is a valuation α, mapping the closed formulae of **E** into the set {**T**, **F**}, such that conjunction and negation obey the usual truth tables, all the true closed formulae of **PA** get the value **T**, $\alpha(\Box A) =$ **T** iff A is provable in **E**, and all the theorems of **E** get the value **T**. It can be shown that there are reflexive extensions of **PA** containing the axioms of S4 or even S4.1, but none containing S5.

Finally, we remark that, using the usual mapping of intuitionistic logic into S4, we can get a model theory for the intuitionistic predicate calculus. We will not give this model theory here, but instead will mention, for propositional calculus only, a particular useful interpretation of intuitionistic logic that results from the model theory. Let **E** be any consistent extension of **PA**. We say a formula P of **PA** is *verified* in **E** iff it is provable in **E**. We take the closed wffs P of **PA** as atomic, and build formulae out of them using the intuitionistic connectives $\wedge$, $\vee$, $\neg$, and $\supset$. We then stipulate inductively: $A \wedge B$ is verified in **E** iff A and B are; $A \vee B$ is verified in **E** iff A or B is; $\neg A$ is verified in **E** iff there is no consistent extension of **E** verifying A; $A \supset B$ is verified in **E** iff every consistent extension **E**$'$ of **E** verifying A also verifies B.

Then every instance of a law of intuitionistic logic is verified in **PA**; but, e.g., $A \vee \neg A$ is not, if A is the Gödel undecidable formula. In future work, we will extend this interpretation further, and show that using it we can find an interpretation for Kreisel's system FC of absolutely free choice sequences.[18] It is clear, incidentally, that **PA** can be replaced in the provability interpretations of S4 and S5 by any truth functional system (i.e., by any system whose models determine each closed formula as true or false); while the interpretation of intuitionism applies to any formal system whatsoever.

[18] G. Kreisel, 'A Remark on Free Choice Sequences and the Topological Completeness' Proofs, *Journal of Symbolic Logic*, 23 (1958), 369–88.

VI

ESSENTIALISM AND QUANTIFIED MODAL LOGIC[1]

TERENCE PARSONS

PROBLEMS involving essentialism are now receiving a great deal of attention from modal logicians and philosophers. Even a cursory glance at work in this field, however, soon reveals that there are many doctrines which go by this title. I will isolate and discuss one such doctrine. In particular, after isolating one version of essentialism (Sections I and II), I will argue that work in quantified modal logic can be and is independent of the acceptance of the truth of this doctrine (Sections III–V). In the last section (Section VI) I will attempt to show, on the basis of facts established in Sections III–V, just why this particular form of essentialism is a philosophically suspect doctrine. I will also argue that work in quantified modal logic need not even presuppose the *meaningfulness* of essentialist claims *in any objectionable sense*.

My arguments aim at (a) a clarification of one sort of essentialism, and (b) a partial vindication of quantified modal logic.

I. PRELIMINARY CLARIFICATION

To begin, let us dichotomize essentialist doctrines into two kinds. One kind has to do with what I shall call *individual* essences and the other with what I shall call *general* essences. The former doctrine makes some claim to the effect that some or all objects have characteristics (or properties) which are so intimately associated with the object that nothing else *could* (with emphasis on the 'could') have precisely those characteristics without being that object. This is meant to be a stronger thesis than the Identity of Indiscernibles, which holds merely that no two objects can simultaneously exist while sharing all properties. It is stronger in two ways: (1) it prohibits the simultaneous existence of two objects which share the same individual essence (even when they could differ in other of their properties), and (2) it makes a claim about what *might have been*: had

From *The Philosophical Review*, LXXVIII. 1 (January 1969), 35–52. Reprinted by permission of the author and *The Philosophical Review*.

[1] In addition to the authors cited in the paper, I am particularly indebted to John Vickers and to Kathryn Pyne Parsons for comments on earlier drafts, and to the referee of *The Philosophical Review* for help in improving the final draft.

plain matter of fact that we do succeed in making reidentifications of both kinds.

Suppose someone tells us 'I did not miss this morning's lecture, but I might have'. The intelligibility to us of this statement depends upon our ability to make sense of the idea that the subject of the statement is identical with an individual in another possible world: one in which he misses this morning's lecture. Indeed, the statements, 'I did not miss this morning's lecture, but I might have' and 'I did not miss this morning's lecture, but there is a possible world in which I did,' are full paraphrases of each other. The latter statement explicitly identifies its subject with an individual in another possible world; if this makes no sense to us, neither does the former statement. But the former statement does make perfect sense. To the extent that we understand such assertions we are able to make sense of the idea of identical individuals in different possible worlds; and we may exploit this understanding to give intuitive meaning to the statements of quantified modal logic. So much the better for quantified modal logic; for this is good enough for modal logicians to be going on with.

VIII

QUANTIFIERS AND PROPOSITIONAL ATTITUDES[1]

W. V. O. QUINE

I

THE incorrectness of rendering 'Ctesias is hunting unicorns' in the fashion:

$(\exists x)(x$ is a unicorn. Ctesias is hunting $x)$

is conveniently attested by the non-existence of unicorns, but is not due simply to that zoological lacuna. It would be equally incorrect to render 'Ernest is hunting lions' as:

(1) $(\exists x)(x$ is a lion . Ernest is hunting $x)$,

where Ernest is a sportsman in Africa. The force of (1) is rather that there is some individual lion (or several) which Ernest is hunting; stray circus property, for example.

The contrast recurs in 'I want a sloop'. The version:

(2) $(\exists x)(x$ is a sloop . I want $x)$

is suitable in so far only as there may be said to be a certain sloop that I want. If what I seek is mere relief from slooplessness, then (2) gives the wrong idea.

The contrast is that between what may be called the *relational* sense of lion-hunting or sloop-wanting, viz. (1)–(2), and the likelier or *notional* sense. Appreciation of the difference is evinced in Latin and Romance languages by a distinction of mood in subordinate clauses; thus '*Procuro un perro que habla*' has the relational sense:

$(\exists x)(x$ is a dog . x talks . I seek $x)$

as against the notional '*Procuro un perro que hable*':

I strive that $(\exists x)(x$ is a dog . x talks . I find $x)$.

From *The Ways of Paradox*, by W. Quine (New York: Random House, 1966), pp. 183–94. Reprinted by permission of the author and the *Journal of Philosophy*.

[1] This paper appeared in the *Journal of Philosophy*, 53 (1956), summing up some points which I had made in lectures at Harvard and Oxford from 1952 onwards. It is reprinted here minus fifteen lines.

Pending considerations to the contrary in later pages, we may represent the contrast strikingly in terms of permutations of components. Thus (1) and (2) may be expanded (with some violence to both logic and grammar) as follows:

(3) $(\exists x)(x$ is a lion . Ernest strives that Ernest finds $x)$,

(4) $(\exists x)(x$ is a sloop . I wish that I have $x)$,

whereas 'Ernest is hunting lions' and 'I want a sloop' in their notional senses may be rendered rather thus:

(5) Ernest strives that $(\exists x)(x$ is a lion . Ernest finds $x)$,

(6) I wish that $(\exists x)(x$ is a sloop . I have $x)$.

The contrasting versions (3)–(6) have been wrought by so paraphrasing 'hunt' and 'want' as to uncover the locutions 'strive that' and 'wish that', expressive of what Russell has called *propositional attitudes*. Now of all examples of propositional attitudes, the first and foremost is *belief*; and, true to form, this example can be used to point up the contrast between relational and notional senses still better than (3)–(6) do. Consider the relational and notional senses of believing in spies.

(7) $(\exists x)$(Ralph believes that x is a spy),

(8) Ralph believes that $(\exists x)(x$ is a spy).

Both may perhaps be ambiguously phrased as 'Ralph believes that someone is a spy', but they may be unambiguously phrased respectively as 'There is someone whom Ralph believes to be a spy' and 'Ralph believes there are spies'. The difference is vast; indeed, if Ralph is like most of us, (8) is true and (7) false.

In moving over to propositional attitudes, as we did in (3)–(6), we gain not only the graphic structural contrast between (3)–(4) and (5)–(6) but also a certain generality. For we can now multiply examples of striving and wishing, unrelated to hunting and wanting. Thus we get the relational and notional senses of wishing for a president:

(9) $(\exists x)$(Witold wishes that x is president),

(10) Witold wishes that $(\exists x)(x$ is president).

According to (9), Witold has his candidate; according to (10) he merely wishes the appropriate form of government were in force. Also we open other propositional attitudes to similar consideration—as witness (7)–(8).

However, the suggested formulations of the relational senses—viz., (3), (4), (7), and (9)—all involve quantifying into a propositional-attitude idiom from outside. This is a dubious business, as may be seen from the following example.

There is a certain man in a brown hat whom Ralph has glimpsed several times under questionable circumstances on which we need not enter here; suffice it to say that Ralph suspects he is a spy. Also there is a grey-haired man, vaguely known to Ralph as rather a pillar of the community, whom Ralph is not aware of having seen except once at the beach. Now Ralph does not know it, but the men are one and the same. Can we say of this *man* (Bernard J. Ortcutt, to give him a name) that Ralph believes him to be a spy? If so, we find ourselves accepting a conjunction of the type:

(11) w sincerely denies '. . .' . w believes that . . .

as true, with one and the same sentence in both blanks. For, Ralph is ready enough to say, in all sincerity, 'Bernard J. Ortcutt is no spy'. If, on the other hand, with a view to disallowing situations of the type (11), we rule simultaneously that

(12) Ralph believes that the man in the brown hat is a spy,

(13) Ralph does not believe that the man seen at the beach is a spy,

then we cease to affirm any relationship between Ralph and any man at all. Both of the component 'that'-clauses are indeed about the man Ortcutt; but the 'that' must be viewed in (12) and (13) as sealing those clauses off, thereby rendering (12) and (13) compatible because not, as wholes, about Ortcutt at all. It then becomes improper to quantify as in (7); 'believes that' becomes, in a word, referentially opaque.[2]

No question arises over (8); it exhibits only a quantification *within* the 'believes that' context, not a quantification *into* it. What goes by the board, when we rule (12) and (13) both true, is just (7). Yet we are scarcely prepared to sacrifice the relational construction 'There is someone whom Ralph believes to be a spy', which (7) as against (8) was supposed to reproduce.

The obvious next move is to try to make the best of our dilemma by distinguishing two senses of belief: *belief*$_1$, which disallows (11), and *belief*$_2$, which tolerates (11) but makes sense of (7). For belief$_1$, accordingly, we sustain (12)–(13) and ban (7) as nonsense. For belief$_2$, on the other hand, we sustain (7); and for *this* sense of belief we must reject (13) and acquiesce in the conclusion that Ralph believes$_2$ that the man at the beach is a

[2] See *From a Logical Point of View*, pp. 142–59 [pp. 19–34 above]; also 'Three grades of modal involvement', Essay 13, *The Ways of Paradox*.

spy even though he *also* believes$_2$ (and believes$_1$) that the man at the beach is not a spy.

II

But there is a more suggestive treatment. Beginning with a single sense of belief, viz., belief$_1$ above, let us think of this at first as a relation between the believer and a certain *intension*, named by the 'that'-clause. Intensions are creatures of darkness, and I shall rejoice with the reader when they are exorcised, but first I want to make certain points with the help of them. Now intensions named thus by 'that'-clauses, without free variables, I shall speak of more specifically as intensions of degree 0, or propositions. In addition I shall (for the moment) recognize intensions of degree 1, or attributes. These are to be named by prefixing a variable to a sentence in which it occurs free; thus z (z is a spy) is spyhood. Similarly we may specify intensions of higher degrees by prefixing multiple variables.

Now just as we have recognized a dyadic relation of belief between a believer and a proposition, thus:

(14) Ralph believes that Ortcutt is a spy,

so we may recognize also a triadic relation of belief among a believer, an object, and an attribute, thus:

(15) Ralph believes z(z is a spy) of Ortcutt.

For reasons which will appear, this is to be viewed not as dyadic belief between Ralph and the proposition *that* Ortcutt has z(z is a spy), but rather as an irreducibly triadic relation among the three things Ralph, z(z is a spy), and Ortcutt. Similarly there is tetradic belief:

(16) Tom believes yz(y denounced z) of Cicero and Catiline,

and so on.

Now we can clap on a hard and fast rule against quantifying into propositional-attitude idioms; but we give it the form now of a rule against quantifying into names of intensions. Thus, though (7) as it stands becomes unallowable, we can meet the needs which prompted (7) by quantifying rather into the triadic belief construction, thus:

(17) ($\exists x$)(Ralph believes z(z is a spy) of x).

Here then, in place of (7), is our new way of saying that there is someone whom Ralph believes to be a spy.

Belief$_1$ was belief so construed that a proposition might be believed when an object was specified in it in one way, and yet not believed when the same object was specified in another way; witness (12)–(13). Hereafter we can adhere uniformly to this narrow sense of belief, both for the dyadic case and for triadic and higher; in each case the term which names the intension (whether proposition or attribute or intension of higher degree) is to be looked on as referentially opaque.

The situation (11) is thus excluded. At the same time the effect of belief$_2$ can be gained, simply by ascending from dyadic to triadic belief as in (15). For (15) does relate the men Ralph and Ortcutt precisely as belief$_2$ was intended to do. (15) does remain true of Ortcutt under any designation; and hence the legitimacy of (17).

Similarly, whereas from:

Tom believes that Cicero denounced Catiline

we cannot conclude:

Tom believes that Tully denounced Catiline,

on the other hand we can conclude from:

Tom believes y(y denounced Catiline) of Cicero

that

Tom believes y(y denounced Catiline) of Tully,

and also that

(18) ($\exists x$)(Tom believes y(y denounced Catiline) of x).

From (16), similarly, we may infer that

(19) ($\exists w$)($\exists x$)(Tom believes yz(y denounced z) of w and x).

Such quantifications as:

($\exists x$)(Tom believes that x denounced Catiline),

($\exists x$)(Tom believes y(y denounced x) of Cicero)

still count as nonsense, along with (7); but such legitimate purposes as these might have served are severed by (17)–(19) and the like. Our names of intensions, and these only, are what count as referentially opaque.

Let us sum up our findings concerning the seven numbered statements about Ralph. (7) is now counted as nonsense, (8) as true, (12)–(13) as true, (14) as false, and (15) and (17) as true. Another that is true is:

(20) Ralph believes that the man seen at the beach is not a spy,

which of course must not be confused with (13).

The kind of exportation which leads from (14) to (15) should doubtless be viewed in general as implicative. Under the terms of our illustrative story, (14) happens to be false; but (20) is true, and it leads by exportation to:

(21) Ralph believes *z*(*z* is not a spy) of the man seen at the beach.

The man at the beach, hence Ortcutt, does not receive reference in (20), because of referential opacity; but he does in (21), so we may conclude from (21) that

(22) Ralph believes *z*(*z* is not a spy) of Ortcutt.

Thus (15) and (22) both count as true. This is not, however, to charge Ralph with contradictory beliefs. Such a charge might reasonably be read into:

(23) Ralph believes *z*(*z* is a spy . *z* is not a spy) of Ortcutt,

but this merely goes to show that it is undesirable to look upon (15) and (22) as implying (23).

It hardly needs be said that the barbarous usage illustrated in (15)–(19) and (21)–(23) is not urged as a practical reform. It is put forward by way of straightening out a theoretical difficulty, which, summed up, was as follows: Belief contexts are referentially opaque; therefore it is *prima facie* meaningless to quantify into them; how then to provide for those indispensable relational statements of belief, like 'There is someone whom Ralph believes to be a spy'?

Let it not be supposed that the theory which we have been examining is just a matter of allowing unbridled quantification into belief contexts after all, with a legalistic change of notation. On the contrary, the crucial choice recurs at each point: quantify if you will, but pay the price of accepting near-contraries like (15) and (22) at each point at which you choose to quantify. In other words: distinguish as you please between referential and non-referential positions, but keep track, so as to treat each kind appropriately. The notation of intensions, of degree one and higher, is in effect a device for inking in a boundary between referential and non-referential occurrences of terms.

III

Striving and wishing, like believing, are propositional attitudes and referentially opaque. (3) and (4) are objectionable in the same way as (7), and our recent treatment of belief can be repeated for these propositional

attitudes. Thus, just as (7) gave way to (17), so (3) and (4) give way to:

(24) (∃*x*)(*x* is a lion . Ernest strives *z*(Ernest finds *z*) of *x*),

(25) (∃*x*(*x* is a sloop . I wish *z*(I have *z*) of *x*),

a certain breach of idiom being allowed for the sake of analogy in the case of 'strives'.

These examples came from a study of hunting and wanting. Observing in (3)–(4) the quantification into opaque contexts, then, we might have retreated to (1)–(2) and forborne to paraphrase them into terms of striving and wishing. For (1)–(2) were quite straightforward renderings of lion-hunting and sloop-wanting in their relational senses; it was only the notional senses that really needed the breakdown into terms of striving and wishing, (5)–(6).

Actually, though, it would be myopic to leave the relational senses of lion-hunting and sloop-wanting at the unanalyzed stage (1)–(2). For, whether or not we choose to put these over into terms of wishing and striving, there are other relational cases of wishing and striving which require our consideration anyway—as witness (9). The untenable formulations (3)–(4) may indeed be either corrected as (24)–(25) or condensed back into (1)–(2); on the other hand we have no choice but to correct the untenable (9) on the pattern of (24)–(25), viz., as:

(∃*x*)(Witold wishes *y*(*y* is president) of *x*).

The untenable versions (3)–(4) and (9) all had to do with wishing and striving in the relational sense. We see in contrast that (5)–(6) and (10), on the notional side of wishing and striving, are innocent of any illicit quantification into opaque contexts from outside. But now notice that exactly the same trouble begins also on the notional side, as soon as we try to say not just that Ernest hunts lions and I want a sloop, but that *someone* hunts lions or wants a sloop. This move carries us, ostensibly, from (5)–(6) to:

(26) (∃*w*)(*w* strives that (∃*x*)(*x* is a lion . *w* finds *x*)),

(27) (∃*w*)(*w* wishes that (∃*x*)(*x* is a sloop . *w* has *x*)),

and these do quantify unallowably into opaque contexts.

We know how, with help of the attribute apparatus, to put (26)–(27) in order; the pattern, indeed, is substantially before us in (24)–(25). Admissible versions are:

(∃*w*)(*w* strives *y*(∃*x*)(*x* is a lion . *y* finds *x*) of *w*),

(∃*w*)(*w* wishes *y*(∃*x*)(*x* is a sloop . *y* has *x*) of *w*), or briefly:

(28) $(\exists w)(w$ strives $y(y$ finds a lion) of $w)$,

(29) $(\exists w)(w$ wishes $y(y$ has a sloop) of $w)$.

Such quantification of the subject of the propositional attitude can of course occur in belief as well; and, if the subject is mentioned in the belief itself, the above pattern is the one to use. Thus 'Someone believes he is Napoleon' must be rendered:

$(\exists w)(w$ believes $y(y =$ Napoleon) of $w)$.

For concreteness I have been discussing belief primarily, and two other propositional attitudes secondarily: striving and wishing. The treatment is, we see, closely parallel for the three; and it will pretty evidently carry over to other propositional attitudes as well—e.g., hope, fear, surprise. In all cases my concern is, of course, with a special technical aspect of the propositional attitudes: the problem of quantifying in.

IV

There are good reasons for being discontent with an analysis that leaves us with propositions, attributes, and the rest of the intensions. Intensions are less economical than extensions (truth values, classes, relations), in that they are more narrowly individuated. The principle of their individuation, moreover, is obscure.

Commonly logical equivalence is adopted as the principle of individuation of intensions. More explicitly: if S and S' are any two sentences with $n(\geqq 0)$ free variables, the same in each, then the respective intensions which we name by putting the n variables (or 'that', if $n = 0$) before S and S' shall be one and the same intension if and only if S and S' are logically equivalent. But the relevant concept of logical equivalence raises serious questions in turn.[3] The intensions are at best a pretty obscure lot.

Yet it is evident enough that we cannot, in the foregoing treatment of propositional attitudes, drop the intensions in favour of the corresponding extensions. Thus, to take a trivial example, consider 'w is hunting unicorns'. On the analogy of (28), it becomes:

w strives $y(y$ finds a unicorn) of w.

Correspondingly for the hunting of griffins. Hence, if anyone w is to hunt unicorns without hunting griffins, the attributes

$y(y$ finds a unicorn),

$y(y$ finds a griffin)

[3] See my 'Two Dogmas of Empiricism', in *From a Logical Point of View*; also 'Carnap and logical truth', which is Essay 10 in *The Ways of Paradox*.

must be distinct. But the corresponding classes are identical, being empty. So it is indeed the attributes, and not the classes, that were needed in our formulation. The same moral could be drawn, though less briefly, without appeal to empty cases.

But there is a way of dodging the intensions which merits serious consideration. Instead of speaking of intensions we can speak of sentences, naming these by quotation. Instead of:

w believes that . . .

we may say:

w believes-true '. . .'.

Instead of:

(30) w believes $y(\ldots y \ldots)$ of x

we may say:

(31) w believes '. . . y . . .' satisfied by x.

The words 'believes satisfied by' here, like 'believes of' before, would be viewed as an irreducibly triadic predicate. A similar shift can be made in the case of the other propositional attitudes, of course, and in the tetradic and higher cases.

This semantical reformulation is not, of course, intended to suggest that the subject of the propositional attitude speaks the language of the quotation, or any language. We may treat a mouse's fear of a cat as his fearing true a certain English sentence. This is unnatural without being therefore wrong. It is a little like describing a prehistoric ocean current as clockwise.

How, where, and on what grounds to draw a boundary between those who believe or wish or strive that p, and those who do not quite believe or wish or strive that p, is undeniably a vague and obscure affair. However, if anyone does approve of speaking of belief of a proposition at all and of speaking of a proposition in turn as meant by a sentence, then certainly he cannot object to our semantical reformulation 'w believes-true S' on any special grounds of obscurity; for, 'w believes-true S' is explicitly definable in *his* terms as 'w believes the proposition meant by S'. Similarly for the semantical reformulation (31) of (30); similarly for the tetradic and higher cases; and similarly for wishing, striving, and other propositional attitudes.

Our semantical versions do involve a relativity to language, however, which must be made explicit. When we say that w believes-true S, we need to be able to say what language the sentence S is thought of as belonging to; not because *w* needs to understand *S*, but because *S* might by coincidence exist (as a linguistic form) with very different meanings in two languages.[4] Strictly, therefore, we should think of the dyadic 'believes-true *S*' as expanded to a triadic '*w* believes-true *S* in *L*'; and correspondingly for (31) and its suite.

As noted two paragraphs back, the semantical form of expression:

(32) *w* believes-true '. . .' in *L*

can be explained in intensional terms, for persons who favour them, as:

(33) *w* believes the proposition meant by '. . .' in *L*,

thus leaving no cause for protest on the score of relative clarity. Protest may still be heard, however, on a different score: (32) and (33), though equivalent to each other, are not strictly equivalent to the '*w* believes that . . .' which is our real concern. For, it is argued, in order to infer (33) we need not only the information about *w* which '*w* believes that . . .' provides, but also some extraneous information about the language *L*. Church[5] brings the point out by appeal to translations, substantially as follows. The respective statements:

w believes that there are unicorns,
w believes the proposition meant by 'There are unicorns' in English

go into German as:

(34) *w glaubt, dass es Einhörne gibt,*

(35) *w glaubt diejenige Aussage, die* „There are unicorns" *auf Englisch bedeutet,*

and clearly (34) does not provide enough information to enable a German ignorant of English to infer (35).

The same reasoning can be used to show that 'There are unicorns' is not strictly or analytically equivalent to:

'There are unicorns' is true in English.

Nor, indeed, was Tarski's truth paradigm intended to assert analytic equivalence. Similarly, then, for (32) in relation to '*w* believes that . . .'; a systematic agreement in truth value can be claimed, and no more. This limitation will prove of little moment to persons who share my scepticism about analyticity.

[4] This point is made by Church in 'On Carnap's analysis' [Essay XI below].
[5] Ibid., with an acknowledgement to Langford.

What I find more disturbing about the semantical versions, such as (32), is the need of dragging in the language concept at all. What is a language? What degree of fixity is supposed? When do we have one language and not two? The propositional attitudes are dim affairs to begin with, and it is a pity to have to add obscurity to obscurity by bringing in language variables too. Only let it not be supposed that any clarity is gained by restituting the intensions.

in question, may be required in order to produce an appropriately discriminating form of Δ which will yield results in conformity with our intuitive demands. Indeed, such an investigation may well lead far beyond the philosophy of language proper into metaphysics and epistemology. I know of no earlier source than 'Quantifiers and Propositional Attitudes' in which relational uses of intermediate contexts are so clearly identified throughout an area of concern more urgent than modal logic. In that article Quine early expressed his remarkable insights into the pervasiveness of the relational forms and the need for a special analysis of their structure. And in fact following Quine's outlook and attempting to refine the conditions for valid applications of exportation, one might well arrive at the same metaphysical and epistemological insights as those obtained in attempting to refine Δ. What is important is that we should achieve some form of analysis of these contexts without recourse to the very idioms we are attempting to analyse.

The problem of interpreting the most interesting form of quantification in, appears in various guises: as the problem of making trans-world identifications, as the problem of finding favoured names, and as the problem of distinguishing 'essential' from 'accidental' properties.

The present paper suggests two polar techniques for finding favoured names. It is curious and somehow satisfying that they so neatly divide the objects between them, the one applying only to objects capable of being perceived (or at least of initiating causal chains), the other applying only to purely abstract objects. I am well aware of obscurities and difficulties in my formulations of the two central notions—that of a standard name and that of a name being *of* an object for a particular user. Yet both seem to me promising and worthy of further investigation.

X

SEMANTICS FOR PROPOSITIONAL ATTITUDES

JAAKKO HINTIKKA

I. THE CONTRAST BETWEEN THE THEORY OF REFERENCE AND THE THEORY OF MEANING IS SPURIOUS

IN the philosophy of logic a distinction is often made between the *theory of reference* and the *theory of meaning*.[1] In this paper I shall suggest (*inter alia*) that this distinction, though not without substance, is profoundly misleading. The theory of reference is, I shall argue, the theory of meaning for certain simple types of language. The only entities needed in the so-called theory of meaning are, in many interesting cases and perhaps even in all cases, merely what is required in order for the expressions of our language to be able to refer in certain more complicated situations. Instead of the theory of reference and the theory of meaning we perhaps ought to speak in some cases of the theory of simple and of multiple reference, respectively. Quine has regretted that the term 'semantics', which etymologically ought to refer to the theory of meaning, has come to mean the theory of reference.[1] I submit that this usage is happier than Quine thinks, and that large parts of the theory of meaning in reality are—or ought to be—but semantical theories for notions transcending the range of certain elementary types of concepts.

It seems to me in fact that the usual reasons for distinguishing between meaning and reference are seriously mistaken. Frequently, they are formulated in terms of a first-order (i.e., quantificational) language. In such a language, it is said, knowing the mere references of individual constants, or knowing the extensions of predicates, cannot suffice to specify their meanings because the references of two individual constants or the extensions of two predicate constants 'obviously' can coincide without there being any identity of meaning.[2] Hence, it is often concluded, the theory

From *Philosophical Logic*, J. W. Davis *et al.* (ed.), (Dordrecht-Holland: D. Reidel Publishing Co., 1969), pp. 21–45. Reprinted by permission of the publishers.

[1] See e.g. W. Quine, *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press, 1953, 2nd edn., 1961), pp. 130–2.

[2] For a simple recent argument of this sort (without a specific reference to first-order theories), see e.g. William P. Alston, *Philosophy of Language* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964), p. 13. Cf. also Quine, op. cit. pp. 21–2.

of reference for first-order languages will have to be supplemented by a theory of the 'meanings' of the expressions of these languages.

The line of arguments is not without solid intuitive foundation, but its implications are different from what they are usually taken to be. This whole concept of meaning (as distinguished from reference) is very unclear and usually hard to fathom. However it is understood, it seems to me in any case completely hopeless to try to divorce the idea of the meaning of a sentence from the idea of the *information* that the sentence can convey to a hearer or reader, should someone truthfully address it to him.[3] Now what is this information? Clearly it is just information to the effect that the sentence is true, that the world is such as to meet the truth-conditions of the sentence.

Now in the case of a first-order language these truth-conditions cannot be divested from the references of singular terms and from the extensions of its predicates. In fact, these references and estensions are precisely what the truth-conditions of quantified sentences turn on. The truth-value of a sentence is a function of the references (extensions) of the terms it contains, not of their 'meanings'. Thus it follows from the above principles that a theory of reference is for genuine first-order languages the basis of a theory of meaning. Recently, a similar conclusion has in effect been persuasively argued for (from entirely different premises and in an entirely different way) by Donald Davidson.[4] The references, not the alleged meanings, of our primitive terms are thus what determine the meanings (in the sense explained) of first-order sentences. Hence the introduction of the 'meanings' of singular terms and predicates is strictly useless: In any theory of meaning which serves to explain the information which first-order sentences convey, these 'meanings' are bound to be completely idle.

What happens, then, to our intuitions concerning the allegedly obvious difference between reference and meaning in first-order languages? If these intuitions are sound, and if the above remarks are to the point, then the only reasonable conclusion is that our intuitions do not really

[3] In more general terms, it seems to me hopeless to try to develop a theory of sentential meaning which is not connected very closely with the idea of the information which the sentence can convey to us, or a theory of meaning for individual words which would not show how understanding them contributes to appreciating the information of the sentences in which they occur. There are of course many nuances in the actual use of words and sentences which are not directly explained by connecting meaning and information in this way, assuming that this can be done. However, there do not seem to be any obstacles in principle to explaining these nuances in terms of pragmatic, contextual, and other contingent pressures operating on a language-user. For remarks on this methodological situation, see my paper 'Epistemic Logic and the Methods of Philosophical Analysis', *Australasian Journal of Philosophy*, 46 (1968), 37–51.

[4] Donald Davidson, 'Truth and Meaning', *Synthèse*, 17 (1967), 304–23.

pertain to first-order discourse. The 'ordinary language' which we think of when we assert the obviousness of the distinction cannot be reduced to the canonical form of an applied first-order language without violating these intuitions. How these other languages enable us to appreciate the real (but frequently misunderstood) force of the apparently obvious difference between reference and meaning I shall indicate later (see Section VI *infra*).

II. FIRST-ORDER LANGUAGES

I conclude that the traditional theory of reference, suitably extended and developed, is all we need for a full-scale theory of meaning in the case of an applied first-order language. All that is needed to grasp the information that a sentence of such a language yields is given by the rules that determine the references of its terms, in the usual sense of the word. For the purposes of first-order languages, to specify the meaning of a singular term is therefore nearly tantamount to specifying its reference, and to specify the meaning of a predicate is for all practical purposes to specify its extension. As long as we can restrict ourselves to first-order discourse, the theory of truth and satisfaction will therefore be the central part of the theory of meaning.

A partial exception to this statement seems to be the theory of so-called 'meaning postulates' or 'semantical rules' which are supposed to catch non-logical synonymies.[5] However, I would argue that whatever non-logical identities of meaning there might be in our discourse ought to be spelled out, not in terms of definitions of terms, but by developing a satisfactory semantical theory for the terms which create these synonymies. In those cases in which meaning postulates are needed, this enterprise no longer belongs to the theory of first-order logic.

In more precise terms, one may thus say that to understand a sentence of first-order logic is to know its interpretation in the actual world. To know this is to know the interpretation function ϕ. This can be characterized as a function which does the following things:

(1.1) For each individual constant a of our first-order language, $\phi(a)$ is a member of the domain of individuals **I**.

The domain of individuals **I** is of course to be thought of as the totality of objects which our language speaks of.

(1.2) For each constant predicate Q (say of n terms), $\phi(Q)$ is a set of n-tuples of the members of **I**.

[5] See Quine, op. cit., pp. 32–7.

If we know ϕ and if we know the usual rules holding of satisfaction (truth), we can in principle determine the truth-values of all the sentences of our-order language. This is the cash value of the statement made above that the extensions of our individual constants and constant predicates are virtually all that we need in the theory of meaning in an applied first-order language.[6]

These conditions may be looked upon in slightly different ways. If ϕ is considered as an arbitrary function in (1.1)–(1.2), instead of that particular function which is involved in one's understanding of a language, and if I is likewise allowed to vary, we obtain a characterization of the concept of interpretation in the general model-theoretic sense.

III. PROPOSITIONAL ATTITUDES

We have to keep in mind the possibility that ϕ might be only a partial function (as applied to free singular terms), i.e., that some of our singular terms are in fact empty. This problem is not particularly prominent in the present paper, however.[7] If what I have said so far is correct, then the emphasis philosophers have put on the distinction between reference and meaning (e.g. between *Bedeutung* and *Sinn*) is motivated only in so far as they have implicitly or explicitly considered concepts which go beyond the expressive power of first-order languages.[8] Probably the most important type of such concept is a propositional attitude.[9] One purpose of this paper is to sketch some salient features of a semantical theory of such concepts. An interesting problem will be the question as to what extent we have to assume entities other than the usual individuals (the members of I) in order to give a satisfactory account of the meaning of propositional attitudes. As will be seen, what I take to be the true answer to this question is surprisingly subtle, and cannot be formulated by a simple 'yes' or 'no'.

What I take to be the distinctive feature of all use of propositional attitudes is the fact that in using them we are considering more than one possibility concerning the world.[10] (This consideration of different possibilities is precisely what makes propositional attitudes propositional, it seems to me.) It would be more natural to speak of different possibilities concerning our 'actual' world than to speak of several possible worlds. For the purpose of logical and semantical analysis, the second locution is much more appropriate than the first, however, although I admit that it sounds somewhat weird and perhaps also suggests that we are dealing with something much more unfamiliar and unrealistic than we are actually doing. In our sense, whoever has made preparations for more than one course of events has dealt with several 'possible courses of events' or 'possible worlds'. Of course, the possible courses of events he considered were from his point of view so many alternative courses that the actual events might take. However, only one such course of events (at most) became actual. Hence there is a sense in which the others were merely 'possible courses of events', and this is the sense on which we shall try to capitalize.

[6] The main reason why the truth of these observations is not appreciated more widely seems to be the failure to consider realistically what the actual use of a first-order language (say for the purpose of conveying information to another person) would look like.

[7] The basic problems as to what happens when this possibility is taken seriously are discussed in my paper, 'Studies in the Logic of Existence and Necessity I', *The Monist*, 50 (1966), 55–76.

[8] This is certainly true of Frege. His very interest in oblique contexts seems to have been kindled by the realization that they cannot be handled by means of the ideas he had successfully applied to first-order logic.

[9] The term seems to go back to Bertrand Russell, *An Inquiry into Meaning and Truth* (London: George Allen and Unwin, 1940).

[10] An important qualification here is that for deep logical reasons one cannot usually distinguish effectively between what is 'really' a logically possible world and what merely 'appears' on the face of one's language (or thinking) to be a possibility. This, in a sufficiently sharp analysis, is what destroys the pleasant invariance of propositional attitudes with respect to logical equivalence. Even though p and q are equivalent, i.e. even though the 'real' possibilities concerning the world that they admit and exclude are the same,

$$a \left\{\begin{matrix}\text{knows}\\ \text{believes}\\ \text{remembers}\\ \text{hopes}\\ \text{strives}\end{matrix}\right\} \text{ that } p$$

and

$$a \left\{\begin{matrix}\text{knows}\\ \text{believes}\\ \text{remembers}\\ \text{hopes}\\ \text{strives}\end{matrix}\right\} \text{ that } q$$

need not be equivalent, for the apparent (to a) possibilities admitted by p and q need not be identical.

I have studied this concept of an 'apparent' possibility and its consequences at some length elsewhere (especially in the second and third paper printed in *Deskription, Analytizität und Existenz*, ed. by Paul Weingartner (Salzburg and Munich: Pustet, 1966); in 'Are Logical Truths Analytic?', *The Philosophical Review*, 74 (1965), 178–203; in 'Surface Information and Depth Information', forthcoming in *Information and Inference*, ed. by K. J. J. Hintikka and P. Suppes (Dordrecht: D. Reidel Publishing Co., 1969), and in 'Are Mathematical Truths Synthetic A Priori?', *Journal of Philosophy* LXV (1968), 640–51.

It is an extremely interesting concept to study and to codify. However, it is not directly relevant to the concerns of the present paper, and would in any case break its confines. Hence it will not be taken up here, except by way of this *caveat*.

Let us assume for simplicity that we are dealing with only one propositional attitude and that we are considering a situation in which it is attributed to one person only. Once we can handle this case, a generalization to the others is fairly straightforward. Since the person in question remains constant throughout the first part of our discussion, we need not always indicate him explicitly.

IV. PROPOSITIONAL ATTITUDES AND 'POSSIBLE WORLDS'

My basic assumption (slightly oversimplified) is that an attribution of any propositional attitude to the person in question involves a division of all the possible worlds (more precisely, all the possible worlds which we can distinguish in the part of language we use in making the attribution) into two classes: into those possible worlds which are in accordance with the attitude in question and into those which are incompatible with it. The meaning of the division in the case of such attitudes as knowledge, belief, memory, perception, hope, wish, striving, desire, etc., is clear enough. For instance, if what we are speaking of are (say) *a*'s memories, then, these possible worlds are all the possible worlds compatible with everything he remembers.

There are propositional attitudes for which this division is not possible. Some such attitudes can be defined in terms of attitudes for which the assumptions do hold, and thus in a sense can be 'reduced' to them. Others may fail to respond to this kind of attempted reduction to those 'normal' attitudes which we shall be discussing here. If there really are such recalcitrant propositional attitudes, I shall be glad to restrict the scope of my treatment so as to exclude them. Enough extremely important notions will still remain within the purview of my methods.

There is a sense in which in discussing a propositional attitude, attributed to a person, we can even restrict our attention to those possible worlds which are in accordance with this attitude.[11] This may be brought out, e.g. by paraphrasing statements about propositional attitudes in terms

[11] There is a distinction here which is not particularly relevant to my concerns in the present paper but important enough to be noted in passing, especially as I have not made it clear in my earlier work. What precisely are the worlds 'alternative to' a given one, say μ? A moment's reflection on the principles underlying my discussion will show, I trust, that they must be taken to be worlds compatible with a certain person's having a definite propositional attitude in μ, and not just compatible with the content of his attitude, for instance, compatible with someone's knowing something in μ and not just compatible with what he knows. I failed to spell this out in my *Knowledge and Belief* (Ithaca, N.Y.: Cornell Univ. Press, 1962), as R. Chisholm in effect pointed out in his review article, 'The Logic of Knowing', *Journal of Philosophy*, LX (1963), 773–95.

of this restricted class of all possible worlds. The following examples will illustrate these approximate paraphrases:

a believes that p = in all the possible worlds compatible with what *a* believes, it is the case that p;

a does not believe that p (in the sense 'it is not the case that *a* believes that p') = in at least one possible world compatible with what *a* believes it is not the case that p.

V. SEMANTICS FOR PROPOSITIONAL ATTITUDES

What kind of semantics is appropriate for this mode of treating propositional attitudes? Clearly what is involved is a set Ω of possible worlds or of models in the usual sense of the word. Each of them, say $\mu \in \Omega$, is characterized by a set of individuals $I(\mu)$ existing in that 'possible world'. An interpretation of individual constants and predicates will now be a two-argument function $\phi(a, \mu)$ or $\phi(Q, \mu)$ which depends also on the possible world μ in question. Otherwise an interpretation works in the same way as in the pure first-order case, and the same rules hold for propositional connectives as in this old case.

Simple though this extension of the earlier semantical theory is, it is in many ways illuminating. For instance, it is readily seen that in many cases earlier semantical rules are applicable without changes. *Inter alia*, in so far as no words for propositional attitudes occur inside the scope of a quantifier, this quantifier is subject to the same semantical rules (satisfaction conditions) as before.

VI. MEANING AND THE DEPENDENCE OF REFERENCE ON 'POSSIBLE WORLDS'

A new aspect of the situation is the fact that the reference $\phi(a, \mu)$ of a singular term now depends on μ—on what course the events will take, one might say. This enables us to appreciate an objection which you probably felt like making earlier when it was said that in a first-order language the theory of meaning is the theory of reference. What really determines the meaning of a singular term, you felt like saying, is not whatever reference it happens to have, but rather the way in which this reference is determined. But in order for this to make any difference, we must consider more than one possibility as to what the reference is, depending on the circumstances (i.e. depending on the course events will take). This dependence is just what is expressed by $\phi(a, \mu)$ when it is considered as a function of μ. (This function *is* the meaning of *a*, one is tempted to say.) Your objection thus has a point. However, it does not show that more is involved in the theory of meaning for first-order languages than the references of its terms. Rather, what is shown is that in order to spell out the idea that the meaning of a term is the way in which its reference is determined we have to consider how the reference varies in different possible worlds, and therefore go beyond first-order languages, just as I suggested above. Analogous remarks apply of course to the extensions of predicates.

Another novelty here is the need of picking out one distinguished possible world from among them all, viz. the world that happens to be actualized ('the actual world').

VII. DEVELOPING AN EXPLICIT SEMANTICAL THEORY: ALTERNATIVENESS RELATIONS

How are these informal observations to be incorporated into a more explicit semantical theory? According to what I have said, understanding attributions of the propositional attitude in question (let us assume that this is expressed by 'B') means being able to make a distinction between two kinds of possible worlds, according to whether they are compatible with the relevant attitudes of the person in question. The semantical counterpart to this is of course a function which to a given individual person assigns a set of possible worlds.

However, a minor complication is in order here. Of course, the person in question may himself have different attitudes in the different worlds we are considering. Hence this function in effect becomes a relation which to a given individual *and to a given possible world* μ associates a number of possible worlds which we shall call the *alternatives* to μ. The relation will be called the alternativeness relation. (For different propositional attitudes, we have to consider different alternativeness relations.) Our basic apparatus does not impose many restrictions on it. The obvious requirement that ensues from what has been said is the following:

(S.B.) $B_a p$ is true in a possible world μ if and only if p is true in all the alternatives to μ.

$B_a p$ may here be thought of as a shorthand for 'a believes that p'. We can write this condition in terms of an interpretation function ϕ. What understanding B means is to have a function ϕ_B which to a given possible world μ and to a given individual a associates a set of possible worlds $\phi_B(a, \mu)$, namely, the set of all alternatives to μ.[12] Intuitively, they are

[12] As the reader will notice, I am misusing (in the interest of simplicity) my terminology systematically by speaking elliptically of 'the person a' etc. when 'the person referred to by a' or some such thing is meant. I do not foresee any danger of confusion resulting from this, however.

the possible worlds compatible with the presence of the attitude expressed by B in the person a in the possible world μ.

In terms of this extended interpretation function, (S.B.) can be written as follows:

$B_a p$ is true in μ if and only if p is true in every member of $\phi_B(a, \mu)$.

VIII. RELATION TO QUINE'S CRITERION OF COMMITMENT

The interesting and important feature of this truth-condition is that it involves quantification over a certain set of possible worlds. By Quine's famous criterion, we allegedly are ontologically committed to whatever we quantify over.[13] Thus my semantical theory of propositional attitudes seems to imply that we are committed to the existence of possible worlds as a part of our ontology.

This conclusion seems to me false, and I think that it in fact constitutes a counter-example to Quine's criterion of commitment *qua* a criterion of ontological commitment. Surely we must in some sense be *committed* to whatever we quantify over. To this extent Quine seems to be entirely right. But why call this a criterion of *ontological* commitment? One's ontology is what one assumes to exist in one's world, it seems to me. It is, as it were, one's census of one's universe. Now such a census is meaningful only in some particular possible world. Hence Quine's criterion can work as a criterion of *ontological* commitment only if the quantification it speaks of is a quantification over entities belonging to some one particular world. To be is perhaps to be a value of a bound variable. But to exist in an ontologically relevant sense, to be a part of the furniture of the world, is to be a value of a special kind of a bound variable, namely one whose values all belong to the same possible world. Thus the notion of a possible world serves to clarify considerably the idea of ontological commitment so as to limit the scope of Quine's dictum.

Clearly, our quantification over possible worlds does not satisfy this extra requirement. Hence there is a perfectly good sense in which we are not ontologically committed to possible worlds, however important their role in our semantical theory may be.

Quine's distinction between *ontology* and *ideology*, somewhat modified and put to a new use, is handy here.[14] We have to distinguish between

[13] e.g. Quine, op. cit., pp. 1–14; W. Quine, *Word and Object* (Cambridge, Mass.: The MIT Press, and New York and London: Wiley and Sons, 1960), pp. 241–3. It is not quite clear from Quine's exposition, however, precisely how much emphasis is to be put on the word 'ontology' in his criterion of ontological commitment. My discussion which focuses on this word may thus have to be taken as a qualification to Quine's criterion rather than as outright criticism.

[14] Quine, *From a Logical Point of View*, pp. 130–2.

what we are committed to in the sense that we believe it to exist in the actual world or in some other possible world, and what we are committed to as a part of our ways of dealing with the world conceptually, committed to as a part of our conceptual system. The former constitute our ontology, the latter our 'ideology'. What I am suggesting is that the possible worlds we have to quantify over are a part of our ideology but not of our ontology.

The general criterion of commitment is a generalization of this. Quantification over the members of one particular world is a measure of ontology, quantification that crosses possible worlds is often a measure of ideology. Quine's distinction thus ceases to mark a difference between two different types of studies or two different kinds of entities within one's universe. It now marks, rather, a distinction between the object of reference and certain aspects of our own referential apparatus. Here we can perhaps see what the so-called distinction between theory of reference and theory of meaning really amounts to.

It follows, incidentally, that if we could restrict our attention to *one* possible word only, Quine's restriction would be true without qualifications. Of course, the restriction is one which Quine apparently would very much like to make; hence he has a legitimate reason for disregarding the qualifications for his own purposes.

Our 'ideological' commitment to possible worlds other than the actual one is neither surprising nor disconcerting. If what we are dealing with are the things people do—more specifically, the concepts they use—in order to be prepared for more than one eventuality, it is not at all remarkable that in order to describe these concepts fully we have to speak of courses of events other than the actual one.

IX. SINGULAR TERMS AND QUANTIFICATION IN THE CONTEXT OF PROPOSITIONAL ATTITUDES

Let us return to the role of individual constants (and other singular terms). Summing up what was said before, we can say that what the understanding of an individual constant amounts to in a first-order language is knowing which individual it stands for. Now it is seen that in the presence of propositional attitudes this statement has to be expanded to say that one has to know what the singular term stands for in the different possible worlds we are considering.

Furthermore, in the same way as these individuals (or perhaps rather the method of specifying them) may be said to be what is 'objectively given' to us when we understand the constant, in the same way what is involved in

the understanding of a propositional attitude is precisely that distinction which in our semantical apparatus is expressed by the function which serves to define the alternativeness relation. This function is what is 'objectively given' to us with the understanding of a word for a propositional attitude.

These observations enable us to solve almost all the problems that relate to the use of identity in the context of propositional attitudes. For instance, we can at once see why the familiar principle of the substitutivity of identity is bound to fail in the presence of propositional attitudes when applied to arbitrary singular terms.[15] Two such terms, say a and b, may refer to one and the same individual in the actual world ($\phi(a, \mu_0) = \phi(b, \mu_0)$ for the world μ_0 that happens to be actualized), thus making the identity '$a = b$' true, and yet fail to refer to the same individual in some other (alternative) possible world (i.e., we may have $\phi(a, \mu_1) \neq \phi(b, \mu_1)$ for some $\mu_1 \in \phi_B(c, \mu_0)$ where c is the individual whose attitudes are being discussed and B the relevant attitude). Since the presence of propositional attitudes means (if I am right) that these other possible worlds have to be discussed as well, in their presence the truth of the identity '$a = b$' does not guarantee that the same things can be said of the references of a and b without qualification, i.e., does not guarantee the intersubstitutivity of the terms a and b.

Our observations also enable us to deal with quantification in contexts governed by words for propositional attitudes as long as we do not quantify *into* them. However, as soon as we try to do so, all the familiar difficulties which have been so carefully and persuasively presented by Quine and others will apply with full force.[16] An individual constant occurring within the scope of an operator like B which expresses a propositional attitude does not specify a unique individual. Rather, what it does is to specify an individual in each of the possible worlds we have to consider. Replace it by an individual variable, and you do not get anything that you could describe by speaking of the individuals over which this variable ranges. There are (it seems) simply no uniquely defined individuals here at all.

It is perhaps thought that the way out is simply to deny that one can ever quantify into a non-extensional context. However, this way out does

[15] For a discussion of the problems connected with the substitutivity principle, see my exchange with Føllesdal: D. Føllesdal, 'Knowledge, Identity, and Existence', *Theoria*, 33 (1967), 1–27; J. Hintikka, 'Existence and Identity in Epistemic Contexts', ibid. pp. 138–47.

[16] See Quine, *From a Logical Point of View*, Ch. 8 [Essay I above]; *Word and Object*, Ch. 6; *The Ways of Paradox and Other Essays* (New York: Random House, 1966), Chs. 13–15.

not work.[17] As a matter of fact, in our ordinary language we often quantify into a grammatical construction governed by an expression for a propositional attitude. Locutions like 'knows who', 'sees what', 'has an opinion concerning the identity of' are cases in point, and so is almost any (other) construction in which pronouns are allowed to mix with words for propositional attitudes. Beliefs about 'oneself' and 'himself' yield further examples, and an account of their peculiarities leads to an interesting reconstruction of the traditional distinction between so-called modalities *de dicto* and *de re*.[18]

Another general fact is that we obviously have beliefs about definite individuals and not just about whoever happens to meet a certain description. I want to suggest that such beliefs (and the corresponding attitudes in the case of other propositional attitudes) are precisely what one half of the *de dicto*—*de re* distinction amounts to.[19]

Furthermore, it does not do to try to maintain that in these constructions the propositional attitude itself has to be taken in an unusual extensional or 'referentially transparent' sense. Such senses can in fact be defined in terms of the normal senses of propositional attitudes. However, these definitions already involve the objectionable quantification into opaque contexts, and if one tries to postulate the defined senses as irreducible primitive senses, they do not have the properties which they ought to have in order to provide the resulting quantified statements with the logical powers they in fact have in ordinary language. For instance, Quine's attempt to postulate a sense of (say) knowledge in which one is allowed to quantify into a context governed by a transparently construed construction 'knows that' has the paradoxical result that

$(\exists x)$ Jones knows that $(x = a)$

is implied by any (transparently interpreted) statement of the form

Jones knows that $(b = a)$

[17] Some arguments to this effect were given in *Knowledge and Belief* (ref. n. 11 above), pp. 142–6. The only informed criticism of this criticism that I have seen has been presented by R. L. Sleigh, in a paper entitled 'A Note on an Argument of Hintikka's', *Philosophical Studies*, 18 (1967), 12–14. As I point out in my reply, 'Partially Transparent Senses of Knowing' (forthcoming), Sleigh's argument turns on an ambiguity in my original formulation which is easily repaired. Neither the ambiguity nor its elimination provides any solace to the adherents of the view I have criticized, however.

[18] One thing at which this old distinction aims is obviously the distinction (which I am about to explain) between statements about whoever or whatever meets a description and statements about the individual who in fact does so. For the distinction, cf. J. Hintikka, 'Individuals, Possible Worlds, and Epistemic Logic', *Nous*, I (1967), 32–62, especially 46–9, as well as ' "Knowing Oneself" and Other Problems in Epistemic Logic', *Theoria*, 32 (1966), 1–13.

[19] Cf. below (Section XII).

and even by a similarly interpreted sentence

Jones knows that $(a = a)$.

This I take to show that the first of these three sentences can scarcely serve as a formulation of 'Jones knows who (or what) *a* is' in our canonical idiom. Yet no other paraphrase of this ubiquitous locution has been proposed, and none is likely to be forthcoming. (For what else can there be to Jones' knowing who *a* is than his knowing of some well-defined individual that Jones is that very individual?) And it is Quine who always insists as strongly as anyone else that the values of bound variables have to be well-defined individuals. It is not much more helpful to try to maintain that no true sentences of the form

Jones knows that $(b = a)$

(with the transparent sense of 'knows') are forthcoming whenever Jones fails to know who *a* is. The transparent sense in which this would be the case has never been explained in a satisfactory way, and I do not see how it can be done in a reasonable way without falling back to my own analysis. (What can it conceivably mean e.g. for Jones not to know in the transparent sense that an *a*, whom he knows to exist, is not self-identical? Can this self-identity fail to be true in a possible world compatible with everything Jones knows?)

Hence we have to countenance quantification into a context governed by an expression for an (opaquely construed) propositional attitude. Our semantical theory at once suggests a way of handling these problems. For instance, in order for existential generalization to be applicable to a singular term *b* occurring, say, in a context where *a*'s belief are being discussed, it has to be required that *b* refers to the *same* individual in the different possible worlds compatible with what *a* believes (plus, possibly, in the actual world). This, naturally, will be expressed by a statement of the form

(*) $(\exists x)[B_a(x = b) \;\&\; (x = b)]$

or, if we do not have to consider the actual world, of the form

$(\exists x)B_a(x = b)$.

X. METHODS OF CROSS-IDENTIFICATION

This solution is simple, straightforward, and workable. It generalizes easily to other propositional attitudes. However, it hides certain interesting conceptual presuppositions. With what right do we speak of individuals in the different possible worlds as being *identical*? This is the problem to which we have to address ourselves.

It is not difficult to see what more there is given to us with our ordinary understanding of propositional attitudes that we have not yet dealt with. For instance, consider a man who has a number of beliefs as to what will happen tomorrow to himself and to his friends. Consider, on his behalf, a number of possible courses of events tomorrow. If I know what our man believes, I can sort these into those which are compatible with his beliefs as distinguished from those which are incompatible with them. But this is not all that is involved. Surely the same or largely the same individuals must figure in these different sequences of events. Under different courses of events a given individual may undergo different experiences, entertain different beliefs and hopes and fears; he may behave rather differently and perhaps even look somewhat different. Nevertheless our man can be (although he need not be) and usually is completely confident that, whatever may happen, he is going to be able to recognize (re-identify) his friends under these various courses of events, at least in principle. He may admit that courses of events are perhaps logically possible under which he would fail to do so; but these would not be compatible with his beliefs as to what will happen. Given full descriptions of two different courses of events tomorrow, both compatible with what our man believes ('believes possible', we sometimes say with more logical than grammatical justification), he will be able to recognize which individuals figuring in one of these descriptions are identical with which individual in the other, even if their names are being withheld. (Of course our man need not believe all this but my point is merely that he *can* and very often *does* believe it.)

The logical moral of this story is that together with the rest of our beliefs we are often given something more than we have so far incorporated into our semantical theory. We are given ways of *cross-identifying* individuals, that is to say, ways of understanding questions as to whether an individual figuring in one possible world is or is not identical with an individual figuring in another world.[20]

This is one point at which the obviousness of my claim may be partially obscured by my terminology. Let us recall what these 'possible worlds' are in the case of propositional attitude. They are normally possible states of affairs or courses of events compatible with the attitude in question in some specified person. Now normally these attitudes may be

[20] Cf. here my paper, 'On the Logic of Perception', forthcoming in *Perception and Personal Identity*, ed. by N. Care and R. Grimm (Cleveland: Case Western Reserve Univ. Press, 1969).

attitudes towards definite persons or definite physical objects. But how is it that we may be sure, sight unseen, that the attitudes are directed towards the right persons or objects? Only if in all the possible worlds compatible with the attitude in question we can pick out the recipient of this attitude, i.e. the individual at its receiving end. Although in many concrete situations the possibility of doing so is obvious, it has not been built into our semantical apparatus so far. There is so far nothing in our semantical theory which enables us to relate to each other the members of the different domains of individuals $\mathbf{I}(\mu)$. In many, though not necessarily all, applications of such relations are given to us as a part of our understanding of the concepts involved. For such cases, we have to build a richer semantical theory.

The way to do so is to postulate a method of making cross-identifications. One possible way to do so is to postulate a set of functions **F** each member f of which picks out at most one individual $f(\mu)$ from the domain of individuals $\mathbf{I}(\mu)$ of each given model μ. We must allow that there is no such value for some models μ. In other words, $f \in \mathbf{F}$ may be a partial function. Furthermore, we must often require that, given $f_1, f_2 \in \mathbf{F}$, if $f_1(\mu) = f_2(\mu)$ then $f_1(\lambda) = f_2(\lambda)$ for all alternatives λ to μ. In other words, an individual cannot 'split' when we move from a world to its alternatives. This question may seem to be a mere matter of detail, but it is easily seen that the question whether an individual can split in the sense just explained is tantamount to the question whether the substitutivity of identity can fail for bound (individual) variables, i.e. to the question whether a sentence

$$(x)(y)(x = y \supset B_a(x = y))$$

can fail to be logically true. This, again, is tantamount to the question whether a sentence of the form

$$(x)(y)(x = y \supset (Q(x) \supset Q(y)))$$

(with just one layer of operators for propositional attitudes in Q) can fail to be logically true.

In terms of the set **F**, the question whether $a \in \mathbf{I}(\mu)$ is identical with $b \in \mathbf{I}(\lambda)$ amounts to the question whether there is a function of $f \in \mathbf{F}$ such that $f(\mu) = a, f(\lambda) = b$.

XI. THE ROLE OF INDIVIDUATING FUNCTIONS

Instead of speaking of a set of functions correlating to each other the individuals existing in the different possible worlds, it is often more appropriate to speak of these domains of individuals as being partly identical (overlapping). Then there would be no need to speak of correlations at all. This point of view is useful in that it illustrates the fact that the apparently different individuals which are correlated by one of the functions $f \in \mathbf{F}$ is just what we ordinarily mean by one and the same individual. It is the concrete individual which we speak about, which we give a name to, etc. In fact, the members of **F** might in fact be thought of as names or individual constants of a certain special kind, namely those having a unique reference in all the different worlds we are speaking of and hence satisfying formulas of the form (*). Indeed, I shall assume in the sequel that a constant of this kind can be associated with each function $f \in \mathbf{F}$.

However, emphasizing the role of the functions $f \in \mathbf{F}$ is useful for several purposes. First and foremost, it highlights an extremely important non-trivial part of our native conceptual skills, namely, our capacity to recognize one and the same individual under different circumstances and under different courses of events. What the set **F** of functions embodies is just the totality of ways of doing this. The non-trivial character of the possibility of this recognition would be lost if we should simply speak of the members of the different possible worlds as being partly identical.

For another thing, the structure formed by the relations of cross-world identity (David Kaplan calls them 'trans-world heir lines') may be so complex as to be indescribable by speaking simply of partial identities between the domains of individuals of the different possible worlds. Above, it was said that in the case of many propositional attitudes an individual cannot 'split' when we move from a world to its alternatives. Although this seems to me to be the case with all the propositional attitudes I have studied in any detail, it is not quite clear to me precisely why this should always be the case. At any rate, there seem to be reasons for suspecting that the opposite 'irregularity' can occasionally take place with some modalities: individuals can 'merge together' when we move from a world to its alternatives. An analogy with temporal modalities may be instructive here.[21] If we presuppose some suitable system of cross-identifications between individuals existing at different times which turn on continuity, it seems possible in principle that a singular term should refer to the same physical system at all the different moments of

[21] For temporal modalities, see e.g. A. N. Prior, *Past, Present and Future* (Oxford: Clarendon Press, 1967). I am not saying that our actual methods of cross-identification in the case of temporal modalities (i.e. on ordinary methods of re-identification) turn on continuity quite as exclusively as I am about to suggest. It suffices for my purposes to present an example of methods of cross-identification that allows both 'branching' and 'merging', and it seems to me at least conceivably that temporal modalities might under suitable circumstances create such a situation.

time we are considering although this system 'merges' with others at times and occasionally 'splits up' into several. Some of these complications seem to be impossible to rule out completely in the case of some propositional attitudes, and because of them the idea of partly overlapping domains seems to me seriously oversimplified.

An extremely important further reason why we cannot reify the members of **F** into ordinary individuals is the possibility of having two different methods of cross-identification between the members of the same possible worlds, i.e. two different sets of 'individuating functions' although we are dealing with precisely the same sets of possible worlds. I have argued elsewhere that this kind of situation is not only possible to envisage but is actually present in our own ways with perceptual contexts.[22] It would take us too far to show precisely what is involved in such cases. Suffice it to point out that this claim, if true, would strikingly demonstrate the dependence of our methods of cross-indentification on our own conceptual schemes and hence on things of our own creation. The apparent simplicity of our idea of an 'ordinary' individual, safe as it may seem in its solid commonplace reality, is thus seen to be merely a reflection of the familiarity and relatively deep customary entrenchment of one particular method of cross-identification, which *sub specie aeternitatis* (i.e. *sub specie logicae*) nevertheless enjoys but a relative privilege as against a host of others.

The methods of cross-identification represented by the set **F** of 'individuating functions', as we might call them, also call for several further comments.

The main function of this part of our semantical apparatus is to make sense of quantification into contexts of propositional attitudes. The truth-conditions of statements in which this happens can be spelled out in terms of membership in **F**. As an approximation we can say the following: A sentence of the form $(\exists x)Q(x)$ is true in μ if an only if there is an individual constant (say b) associated with some $f \in \mathbf{F}$ such that $Q(b)$ is true in μ. This approximation shows, incidentally, how close we can stick to the simple-minded idea that an existentially quantified sentence is true if and only if it has a true substitution instance. The only additional requirement we need is that the substitution-value of the bound variable has to be of the right sort, to wit, has to specify the same individual in all the possible worlds we are speaking of in the existential sentence in question. This is what is meant by the requirement that b has to be associated with one of the functions $f \in \mathbf{F}$.

This approximation, although not unrepresentative of the general situation, requires certain modifications in order to work in all cases. The

[22] This is argued in 'On the Logic of Perception' (ref. n. 20 above).

set **F** has to be relativized somewhat in the same way the unrestricted notion of a possible world was replaced by the notion of an alternative in the truth-criterion (S.B.) above. (Not everyone is in all situations 'familiar with' all the relevant methods of individuation, it might be said.) I shall not discuss the ensuing complications here, however, for they do not change the overall picture in those respects which are relevant in the rest of this paper.

XII. STATEMENTS ABOUT DEFINITE INDIVIDUALS VS. STATEMENTS ABOUT WHOEVER OR WHATEVER IS REFERRED TO BY A TERM

The possibility of quantifying across an operator which expresses a propositional attitude enables us to explicate the logic of the locutions in which we need this possibility in the first place. Perhaps the most important thing we can do here is to make a distinction between propositional attitudes directed to whoever (whatever) happens to be referred to by a term and attitudes directed towards a certain individual, independently of how he happens to be referred to. This distinction was hinted at above. Now it is time to explain it more fully. For instance, someone may have a belief concerning the next Governor of California, whoever he is or may be, say that he will be a Democrat. This is different from believing something about the individual who, so far unbeknownst to all of us, in fact is the next Governor of California.

In formal terms, the distinction is illustrated by the pair of statements

$B_a(g$ is a Democrat)

$(\exists x)((x = g)$ & $B_a(x$ is a Democrat)).

Notice, incidentally, that my way of drawing this distinction implies that one can have (say) a belief concerning the individual who in fact is a only if such individual actually exists, whereas one can in principle have a belief concerning a, 'whoever he is', even though there is no such person. This, of course, is just as it ought to be.

The naturalness of our semantical conditions, and their close relation to the realities of actual usage, can be illustrated by applying them to what I have called a statement about a definite individual. As an example, we can use

'a believes of the man who in fact is Mr. Smith that he is a thief',

in brief,

$(\exists x)(x =$ Smith & $B_a(x$ is a thief)).

In order for this to be true, there has to be some $f \in \mathbf{F}$ such that the value of f in the actual world (call it μ_0) exists and is Smith and that $f(\mu)$ has the property of being a thief whenever $\mu \in \phi_B(a, \mu_0)$, i.e. in all the alternatives to the actual world.

What the requirement of the existence of f amounts to is clear enough. If it is true to say that a has a belief about *the particular individual* who in fact is Smith, then a clearly must believe that he can characterize this individual uniquely. In other words, he must have some way of referring to or characterizing this individual in such a way that one and the same individual is in fact so characterized in all the worlds compatible with what he believes. This is precisely what the existence of f amounts to. If no such function existed, a would not be able to pick out the individual who in fact is Smith under all the courses of events he believes possible, and there would not be any sense in saying that a's belief is *about* the particular individual in question.

XIII. INDIVIDUATING FUNCTIONS VS. INDIVIDUAL CONCEPTS

One important consequence of my approach is that not every function which from each μ picks out an individual can be said to specify a unique individual. In fact, many perfectly good free singular terms fail to do so in the context of many propositional attitudes. Even proper names fail to do so in epistemic contexts, for one may fail to know who the bearer of a given proper name is.

Such arbitrary functions may be important for many purposes. They are excellent approximations in our theory to the 'individual concepts' which many philosophers have postulated.[23] (In Section VI above we already met a number of such 'individual concepts' in the form of the functions $\phi(a, \mu)$ with a fixed a.) Each such individual concept specifies or 'contains', as Frege would say, not just a reference (in the actual world) but also the way in which this reference is given to us. Each of them would thus qualify for a sense (*Sinn*) of a singular term *à la* Frege.[24] However, we do not need the totality of such arbitrary functions in the semantics which I am building up and which (I want to argue) is largely implicit in our native conceptual apparatus. Quine's criterion, however misleading it may be as a criterion of ontological commitment, still works as a criterion of commitment. If it is applied here, it shows that we are not committed (ontologically or 'ideologically') to these arbitrary functions,

[23] Cf. e.g. R. Carnap, *Meaning and Necessity* (Chicago: University of Chicago Press, 1947, 2nd edn.: 1956) pp. 41, 180–1, and Section VI, above.

[24] Cf. Gottlob Frege, 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik*, 100 (1892), 25–50, especially p. 26, last few lines.

since we do not have to quantify over them, only over the members of the much narrower class **F**.

The other side of the coin is that in our semantical apparatus we do have to quantify over the members of **F**. Does it follow that they 'exist' or 'are part of our ontology'? An answer to this question can be given along the same lines as to the corresponding question concerning 'possible worlds'. The members of **F** are not members of any possible world; they are not part of anybody's count of 'what there is'. They may 'subsist' or perhaps 'exist', and they are certainly 'objective', but they do not have any ontological role to play. The need to distinguish between ontology and 'ideology' is especially patent here.

The functions that belong to **F** may of course be considered special cases of the 'individual concepts' postulated by some philosophers of logic or as special cases of Frege's 'senses' (*Sinne*). No identification is possible between the two classes, however, for we saw earlier that not every arbitrary singular term (say b) which picks out an individual from each $\mathbf{I}(\mu)$ we are considering goes together with an $f \in \mathbf{F}$, although every such term is certainly meaningful and hence has a Fregean 'sense' and perhaps even gives us an 'individual concept'. As I have put it elsewhere, members of **F** do not only involve a 'way of being given' as Frege's senses do, but also *a way of being individuated*.[25] The primary care is in our approach devoted to ordinary concrete individuals. Singular terms merit a special honorary mention only if they succeed in picking out a unique individual of this sort.

Let us say that an $f \in \mathbf{F}$ is (gives us) an individuating concept, and let us say that a term b does *individuate* (in the context of discussing a's beliefs) in so far as

$$(\exists x)B_a(x = b)$$

is true. Then we could have individuation without reference and reference without individuation: Both

$$(\exists x)B_a(x = b) \ \& \sim(\exists x)(x = b)$$

and

$$(\exists x)(x = b) \ \& \sim(\exists x)B_a(x = b)$$

can be true. We could even have both, but without matching:

$$(\exists x)(x = b) \ \& \ (\exists x)B_a(x = b) \ \& \sim(\exists x)((x = b) \ \& \ B_a(x = b))$$

[25] Cf. 'On the Logic of Perception' (ref. n. 20 above).

is satisfiable. Only if

$$(\exists x)((x = b) \ \& \ B_a(x = b))$$

is true does the successful individuation give us the individual which the term b actually refers to.

XIV. THEORY OF REFERENCE AS REPLACING THE THEORY OF MEANING

Here we are perhaps beginning to see what I meant when I said in the beginning of this paper that what is often called the theory of meaning is better thought of as the theory of reference for certain more complicated conceptual situations. Some of the most typical concepts used in the theory of meaning, such as Frege's *Sinne* and the 'individual concepts' of certain other philosophers of logic were in the first place introduced to account for such puzzles as the failure of the substitutivity of identity and the difficulty of quantifying into opaque contexts (e.g. into a context governed by a word for a propositional attitude). I have argued, however, that a satisfactory semantical theory which clears up these puzzles can be built up without using Frege's *Sinne* and without any commitment to individual concepts in any ordinary sense of the word. Instead, what we need are the individuating functions, i.e., the members of **F**. And what these functions do is not connected with the ideas of the traditional theory of meaning. What they do is precisely to give us the individuals which we naively think our singular statements to be about and which we think our singular terms as referring to. This naive point of view is essentially correct, it seems to me. The functions of $f \in \mathbf{F}$ are the prime vehicles of our references to individuals when we discuss propositional attitudes. What is not always realized, however, is how much goes into our ordinary concepts of an individual and a reference. These are not specified in a way which works only under one particular course of events. They are in fact specified in a way which works under a wide variety of possible courses of events. But, in order to spell out this idea, we are led to consider several possible worlds, with all the problems with which we have dealt in this paper, including very prominently the problem of cross-identifying individuals.

The function of our 'individuating functions', i.e. the members of the set **F**, is to bring out these hidden—or perhaps merely overlooked—aspects of our concept of an individual (definite individual). This close connection between the set **F** and the concept of an individual appears in a variety of ways. One may for instance think of the role which the membership in **F** plays in the truth-conditions which we set up above for quantification into modal contexts. When it is asked in such a context whether there exists an individual of a certain kind, a singular term specifies such an individual only if its references match the values of a unique member of **F** in all the relevant possible worlds. Thus it is these functions that in effect give us the individuals which can serve as values of bound variables. As we saw above, it is mainly the possible subtlety and multiplicity of relations of cross-identity that prevent us from simply making the domains of the different possible worlds partly identical and thus hypostatizing my individuating functions into commonplace individuals.

This connection between individuating functions and the concept of an individual is part of what justifies us in thinking that in the traditional dichotomy their theory would belong primarily to the theory of reference, in spite of the fact that their main function in our semantical theory is to solve some of the very problems which the traditional theory of meaning was calculated to handle. This role is perhaps especially clear in connection with the substitutivity of identity. As we have seen, this principle does not hold for arbitrary singular terms *a*, *b*. However, if it is required in addition that both of these terms specify a well-defined individual, i.e. satisfy expressions like (*), depending on the context, then the substitutivity of identicals is easily seen to hold, presupposing of course here the prohibition against merging that was mentioned above. What this observation shows is clear enough. The failure of the substitutivity of identity poses one of the most typical problems for the treatment of which meanings, individual concepts and other paraphernalia of the theory of meaning were introduced in the first place. If the substitutivity of identity fails, clearly we cannot be dealing with ordinary commonplace individuals, it was alleged, for if two such individuals are in fact identical, surely precisely the same things can be said of them. This is what prompts the quest for individuals of some non-ordinary sort, capable of restoring the substitutivity principle when used as references of our terms. (This is almost precisely Frege's strategy.) We have seen, however, that the (apparent) failure of the substitutivity is due simply to the failure of some free singular terms to specify the same individual in the different 'possible worlds' we have to consider. Moreover, we have seen that this apparent failure is automatically corrected in precisely those cases in which it ought to be corrected, viz. in the cases where the two terms in question really do specify a *unique* individual. (That this depends on certain specific requirements concerning our methods of cross-identification, viz. on a prohibition against 'splitting', does not affect my point.) Substitutivity of identity is restored, in belief, not by requiring that our singular terms refer to the entities postulated

by the so-called theory of meaning, but by requiring (in the form of an explicit premise) that they really succeed in specifying uniquely the kind of ordinary individual with which the theory of reference typically deals. One can scarcely hope to find a more striking example of the breakdown of the distinction between a theory of meaning and a theory of reference.[26]

XV. TOWARDS A SEMANTIC NEOKANTIANISM

The aspect of my observations most likely to upset many contemporary philosophers is the ensuing implicit dependence of our concept of an individual on our ways of cross-identifying members of different 'possible worlds'. These 'possible worlds' and the supply of individuating functions which serve to interrelate their respective members may enjoy, and in my view do enjoy, some sort of objective reality. However, their existence is not a 'natural' thing. They may be as solidly objective as houses or books, but they are as certainly as these created by men (however unwittingly) for the purpose of facilitating their transactions with the reality they have to face. Hence my reasoning ends on a distinctly Kantian note. Whatever we say of the world is permeated throughout with concepts of our own making. Even such *prima facie* transparently simple notions as that of an individual turn out to depend on conceptual assumptions dealing with different possible states of affairs. As far as our thinking is concerned, reality cannot be in principle wholly disentangled from our concepts. A *Ding an sich*, which could be described or even as much as individuated without relying on some particular conceptual framework, is bound to remain an illusion.

[26] Views closely resembling some of those which I am putting forward here (and in some cases anticipating them) have been expressed by David Kaplan, Richard Montague, Dagfinn Føllesdal, Stig Kanger, Saul Kripke, and others. Here I am not trying to relate my own ideas to theirs. It is only fair, however, to emphasize my direct and indirect debts to these writers.

XI

ON CARNAP'S ANALYSIS OF STATEMENTS OF ASSERTION AND BELIEF

ALONZO CHURCH

I

FOR statements such as (1) *Seneca said that man is a rational animal* and (A) *Columbus believed the world to be round*, the most obvious analysis makes them statements about certain abstract entities which we shall call 'propositions' (though this is not the same as Carnap's use of the term), namely the proposition that man is a rational animal and the proposition that the world is round; and these propositions are taken as having been respectively the object of an assertion by Seneca and the object of a belief by Columbus. We shall not discuss this obvious analysis here except to admit that it threatens difficulties and complications of its own, which appear as soon as the attempt is made to formulate systematically the syntax of a language in which statements like (1) and (A) are possible. But our purpose is to point out what we believe may be an insuperable objection against alternative analyses that undertake to do away with propositions in favour of such more concrete things as sentences.

As attempts which have been or might be made to analyse (1) in terms of sentences we cite: (2) *Seneca wrote the words 'Man is a rational animal'*; (3) *Seneca wrote the words 'Rationale enim animal est homo'*; (4) *Seneca wrote words whose translation from Latin into English is 'Man is a rational animal'*; (5) *Seneca wrote words whose translation from some Language S′ into English is 'Man is a rational animal'*; (6) *There is a language S′ such that Seneca wrote as sentence of S′ words whose translation from S′ into English is 'Man is a rational animal'*. In each case, 'wrote' is to be understood in the sense, 'wrote with assertive intent'. And to simplify the discussion, we ignore the existence of spoken languages, and treat all languages as written.

Of these proposed analyses of (1), we must reject (2) on the ground that it is no doubt false although (1) is true. And each of (3)–(6), though having the same truth-value as (1), must be rejected on the ground that it

From *Analysis*, 10, 5 (1950), 97–9. Reprinted by permission of the author, *Analysis*, and Basil Blackwell.

does not convey the same information as (1). Thus (1) conveys the content of what Seneca said without revealing his actual words, while (3) reproduces Seneca's words without saying what meaning was attached to them. In (4) the crucial information is omitted (without which (1) is not even a consequence) that Seneca intended his words as a Latin sentence, rather than as a sentence of some other language in which conceivably the identical words 'Rationale enim animal est homo' might have some quite different meaning. To (5) the objection is the same as to (4), and indeed if we take 'language' in the abstract sense of Carnap's 'semantical system' (so that it is not part of the concept of a language that a language must have been used in historical fact by some human kindred or tribe), then (5) is L-equivalent merely to the statement that Seneca once wrote something.

(5) and (6) are closely similar to the analysis of belief statements which is offered by Carnap in 'Meaning and Necessity', and although he does not say so explicitly it seems clear that Carnap must have intended also such an analysis as this for statements of assertion. However, (6) is likewise unacceptable as an analysis of (1). For it is not even possible to infer (1) as a consequence of (6), on logical grounds alone—but only by making use of the item of factual information, not contained in (6), that 'Man is a rational animal' means in English that man is a rational animal.

Following a suggestion of Langford[1] we may bring out more sharply the inadequacy of (6) as an analysis of (1) by translating into another language, say German, and observing that the two translated statements would obviously convey different meanings to a German (whom we may suppose to have no knowledge of English). The German translation of (1) is (1′) *Seneca hat gesagt, dass der Mensch ein vernünftiges Tier sei.* In translating (6), of course 'English' must be translated as 'Englisch' (not as 'Deutsch') and 'Man is a rational animal' must be translated as 'Man is a rational animal' (not as 'Der Mensch ist ein vernünftiges Tier').

Replacing the use of translation (as it appears in (6)) by the stronger requirement of intensional isomorphism, Carnap would analyse the belief statement (A) as follows: (B) *There is a sentence* $\mathfrak{S}_i$ *in a semantical system S′ such that* (*a*) $\mathfrak{S}_i$ *is intensionally isomorphic to 'The world is round' and* (*b*) *Columbus was disposed to an affirmative response to* $\mathfrak{S}_i$. However, intentional isomorphism, as appears from Carnap's definition of it, is a relation between ordered pairs consisting each of a sentence and a semantical system. Hence (B) must be rewritten as: (C) *There is a sentence* $\mathfrak{S}_i$ *in a semantical system S′ such that* (*a*) $\mathfrak{S}_i$ *as sentence of S′ is intensionally*

[1] In *Journal of Symbolic Logic*, 2 (1937), 53.

isomorphic to 'The world is round' as English sentence and (b) Columbus was disposed to an affirmative response to $\mathfrak{S}_i$ *as sentence of S'.*

For the analysis of (1), the analogue of (C) would seem to be: (7) *There is a sentence* $\mathfrak{S}_i$ *in a semantical system S' such that (a)* $\mathfrak{S}_i$ *as sentence of S' is intensionally isomorphic to 'Man is a rational animal' as English sentence and (b) Seneca wrote* $\mathfrak{S}_i$ *as sentence of S'.*

Again Langford's device of translation makes evident the untenability of (C) as an analysis of (A), and of (7) as an analysis of (1).

II

The foregoing assumes that the word 'English' in English and the word 'Englisch' in German have a sense which includes a reference to matters of pragmatics (in the sense of Morris and Carnap)—something like, e.g., 'the language which was current in Great Britain and the United States in A.D 1949.'

As an alternative we might consider taking the sense of these words to be something like 'the language for which such and such semantical rules hold', a sufficient list of rules being given to ensure that there is only one language satisfying the description. The objection would then be less immediate that (1) is not a logical consequence of (6) or (7), and it is possible that it would disappear.

In order to meet this latter alternative without discussing in detail the list of semantical rules which would be required, we modify as follows the objection to (7) as an analysis of (1). Analogous to the proposal, for English, to analyse (1) as (7), we have, for German, the proposal to analyse (1') as (7'') *Es gibt einen Satz* $\mathfrak{S}_i$ *auf einem semantischen System S', so dass (a)* $\mathfrak{S}_i$ *als Satz von S' intensional isomorph zu 'Der Mensch ist ein vernünftiges Tier' als deutscher Satz ist, und (b) Seneca* $\mathfrak{S}_i$ *als Satz von S' geschrieben hat.* Because of the exact parallelism between them, the two proposals stand or fall together. Yet (7'') in German and (7) in English are not in any acceptable sense translations of each other. In particular, they are not intensionally isomorphic. And if we consider the English sentence (*a*) *John believes that Seneca said that man is a rational animal* and its German translation (*a'*), we see that the sentences to which we are led as supposed analyses of (*a*) and (*a'*) may even have opposite truth-values in their respective languages; for John, though knowing the semantical rules of both English and German, may nevertheless fail to draw certain of their logical (or other) consequences.

NOTES ON THE CONTRIBUTORS

WILLARD V. O. QUINE is Professor of Philosophy at Harvard, and is widely regarded as America's leading philosopher. Among his many publications are *A System of Logistic*, *Mathematical Logic*, *Elementary Logic*, *Methods of Logic*, *From a Logical Point of View*, *Word and Object*, *Set Theory and its Logic*, *The Ways of Paradox*, *Selected Logic Papers*, and *Ontological Relativity*.

ARTHUR F. SMULLYAN is Professor at Rutgers University. He is the author of *Fundamentals of Logic*.

RUTH B. MARCUS is Professor at Northwestern University. She is one of the principal founders of quantified modal logic, and she has written extensively on the philosophical foundations of the subject.

DAGFINN FØLLESDAL is Professor at the University of Oslo and has been a regular visitor at Stanford University. His main work has been on the foundations of epistemic and causal logic, though he has also written on Husserl.

SAUL KRIPKE is Professor at Rockefeller University, New York. He has published widely on modal logic and in other areas of logic as well.

TERENCE PARSONS is Associate Professor at the University of Illinois at Chicago Circle. He has written on the problem of essentialism, and on ontological commitment.

LEONARD LINSKY, the editor of this volume, is Professor and Chairman of the Department of Philosophy at the University of Chicago. He is editor of *Semantics and the Philosophy of Language* and the author of *Referring*.

DAVID KAPLAN is Professor at the University of California at Los Angeles. His main work has been on intensional languages, referential opacity, and belief contexts. He has been influential through his participation in colloquia on these topics at philosophical meetings.

JAAKKO HINTIKKA is Professor at Helsinki and at Stanford University. He is the principal founder of epistemic logic, though he has published widely in other areas such as inductive logic, perception, and the philosophy of mathematics. He is the author of *Knowledge and Belief*, and edited *The Philosophy of Mathematics* in the present series.

ALONZO CHURCH is Professor of Philosophy at the University of California at Los Angeles. He has been an editor of the *Journal of Symbolic Logic* since its inception in 1936. He is the author of *Introduction to Logic*. Church's main work has been in logic, although he has also written much on philosophy, a good deal of it in the form of reviews in the journal which he edits. He has written on Frege's semantics.