

## THE PROBLEM OF INTERPRETING MODAL LOGIC

W. V. QUINE

There are logicians, myself among them, to whom the ideas of modal logic (e. g. Lewis's) are not intuitively clear until explained in non-modal terms. But so long as modal logic stops short of quantification theory, it is possible (as I shall indicate in §2) to provide somewhat the type of explanation desired. When modal logic is extended (as by Miss Barcan<sup>1</sup>) to include quantification theory, on the other hand, serious obstacles to interpretation are encountered—particularly if one cares to avoid a curiously idealistic ontology which repudiates material objects. Such are the matters which it is the purpose of the present paper to set forth.

**1. Analytic statements from the standpoint of non-modal logic.** All true statements which (like ' $(x) (x = x)$ ') contain only logical signs are naturally to be classified as logically true. But there are also other logically true statements (e. g. 'Socrates is mortal  $\supset$  Socrates is mortal') which contain extra-logical signs. Now every logical truth of the latter kind is (if for the present we disregard modal logic) either *provable within* the familiar logic of truth-functions and quantification alone, or else *deducible by* the logic of truth-functions and quantification *from* logical truths of the kind which contain only logical signs. So *every* logical truth is either:

- (a) a true statement containing only logical signs, or
- (b) provable in the logic of truth-functions and quantification, or
- (c) *deducible by the logic of truth-functions and quantification from true statements containing only logical signs.*

But (c) includes (b), since whatever is provable is also, vacuously, deducible from anything else. Also (c) includes (a), since any statement is deducible from itself. Thus (c) by itself might be adopted as an appropriate *definition of logical truth* (outside modal logic). The "deducibility" spoken of in (c) can be expanded into purely syntactical terms by an enumeration of the familiar rules, which are known to be complete; and the reference to "logical signs" can likewise be expanded by enumeration of the familiar primitives. The word 'true' in (c) cannot indeed be expanded; no enumeration of axioms or axiom-schemata would serve the purpose here, because of Gödel's proof that there can be no complete consistent system of higher logic. We could desire otherwise, especially in view of the logical paradoxes which are known to be connected with the general concept of truth.<sup>2</sup> Nevertheless (c) is not without explicative value, as marking out the special notion of logical truth within the general notion (such as it is) of truth.

---

Received July 17, 1946

<sup>1</sup> Ruth C. Barcan, *A functional calculus of first order based on strict implication*, this JOURNAL, vol. 11 (1946), pp. 1-16.

<sup>2</sup> See Alfred Tarski, *The semantic conception of truth and the foundation of semantics, Philosophy and phenomenological research*, vol. 4 (1944), pp. 341-376.

The class of *analytic* statements is broader than that of logical truths, for it contains in addition such statements as 'No bachelor is married.' This example might be assimilated to the logical truths by considering it a definitional abbreviation of 'No man not married is married,' which is indeed a logical truth; but I should prefer not to rest analyticity thus on an unrealistic fiction of there being standard definitions of extra-logical expressions in terms of a standard set of extra-logical primitives. What is rather in point, I think, is a relation of *synonymy*, or sameness of meaning, which holds between expressions of real language though there be no standard hierarchy of definitions. In terms of synonymy and logical truth we could define analyticity: a statement is *analytic* if by putting synonyms for synonyms (e.g. 'man not married' for 'bachelor') it can be turned into a logical truth.

The particular synonymy relation wanted is one of several which have about equal right to the name "synonymy" and are all describable as "sameness of meaning"—in varying senses of "meaning." Synonymy of the kind which renders expressions interchangeable without violence to indirect quotation, for example, would be a narrower relation than the one here concerned; and the sense of synonymy proposed by Lewis<sup>3</sup> is yet a third relation, I believe, intermediate between the two.

The particular synonymy relation which I have in mind can easily be described in terms of analyticity, at least when the expressions related happen to be statements, names, or predicates. Statements are synonymous if the biconditional ('if and only if') which joins them is analytic; names are synonymous if the statement of identity which joins them is analytic; and predicates are synonymous if, when they are applied to like variables and then combined into a universally quantified biconditional, the result is analytic.

But of course we must seek some other definition of synonymy so as to avoid circularity, if synonymy is to be available for use in defining analyticity. Actually the synonymy relation sought is, insofar as applied to propositions, names, and predicates, precisely what Lewis (*loc. cit.*) calls *sameness of intension*; but expansion in turn of his definition of "intension" leaves us with phrases which I do not feel to be appreciably clearer than "synonymy" and "analyticity" themselves.

Synonymy, like other linguistic concepts, involves parameters of times and persons, suppressible in idealized treatment: the expression  $x$  is synonymous with the expression  $y$  for person  $z$  at time  $t$ . A satisfactory definition of this tetradic relation would no doubt be couched, like those of other general concepts of general linguistics, in behavioristic terms. I should like, as a service both to empirical semantics and to philosophy, to offer a satisfactory definition; but I have none. So long, however, as we persist in speaking of expressions as alike or unlike in meaning (and regardless of whether we countenance meanings themselves in any detached sense), we must suppose that there is an eventually formulable criterion of synonymy in some reasonable sense of the term—and probably, in particular, in the sense appropriate to present purposes. Given it,

<sup>3</sup> C. I. Lewis, *The modes of meaning*, *ibid.*, vol. 4 (1943), p. 245.

and given the general notion of truth (in terms of which logical truth has been seen to be definable), we could define analyticity as previously indicated.

The notion of analyticity thus appears, at the present writing, to lack a satisfactory foundation. Even so, the notion is clearer to many of us, and obscurer surely to none, than the notions of modal logic; so we are still well advised to explain the latter notions in terms of it. This can be done, as sketched in the next section, so long as modal logic stops short of quantification theory.<sup>4</sup>

**2. Interpretation of pre-quantificational modal logic.** We need consider only the mode of logical *necessity*, symbolized by '□'; for the other modal ideas (possibility, impossibility, and the strict conditional and biconditional) are expressible in terms of necessity in obvious fashion. Now '□' is not quite interchangeable with 'is analytic,' for this reason: the former attaches to a *statement* (as '~' does) to form a statement *containing* the original statement, whereas 'is analytic' (like 'is true,' 'is false') attaches to the *name* of a statement to form a statement *about* the named statement. Grammatically '□' is an adverb; 'is analytic' is a verb. The formal difference becomes immediately apparent in the case of iteration: '□' can significantly be applied repeatedly (because the result of any application, being a statement, is the kind of expression to which '□' can still significantly be prefixed) whereas 'is analytic' can be applied only once (because the result of the application is a statement, whereas 'is analytic' can be appended only to a name.)<sup>5</sup>

However, '□' can be explained in terms of analyticity as follows:

(i) *The result of prefixing '□' to any statement is true if and only if the statement is analytic.*

Let us speak of a statement as of *first intention* if it is non-modal; and let us speak of a statement as of ( $n + 1$ )st intention if '□' occurs in it in application to statements of  $n$ th intention and no higher. The portion of modal logic which admits statements of  $n$ th intention at most (and includes also all non-modal logic) will be called the logic of  $n$ th intention; non-modal logic itself is the logic of first intention. Then (i) above may be taken as our explanation of all statements of second intention; and the word 'analytic' in it (explained in the preceding section) may be read 'analytic in first intention.' Relative to the logic of second intention, now, which has some new truths of its own in addition to those of the logic of first intention, we may adopt new definitions of logical truth

---

<sup>4</sup> Dr. Nelson Goodman has suggested (in conversation) the dismal possibility that what we think of as synonymy may be wholly a matter of degree, ranging from out-and-out orthographical sameness of expressions on the one hand to mere factual sameness of designatum (as in the case of 'nine' and 'the number of the planets') on the other. In this case analyticity in turn would become a matter of degree—a measure merely of our relative reluctance to give up one statement rather than another from among a set of statements whose conjunction has proved false. But if it does develop that the boundary between analytic and synthetic statements has thus to be rubbed out, no doubt it will be generally agreed that the logical modalities have to be abandoned as well. The explanation of modal logic in terms of analyticity remains of interest so long as there is interest in modal logic itself.

<sup>5</sup> Cf. my *Mathematical logic*, pp. 27–33.

and analyticity paralleling those of the preceding section (but admitting un-iterated use of '□' in addition to truth-functions in (c)). There upon we may repeat (i) above as an explanation now of all statements of third intention. Supposing this process continued *ad infinitum*, we have an explanation of '□' in application to any statements.

I suppose that some such conception underlies the intuition where by axioms are evaluated and adopted for modal logic. The explanation of modal logic thus afforded is adequate so long as modalities are not used inside the scopes of quantifiers; i. e., so long as '□' is applied only to statements and not to matrices.<sup>6</sup> Of course in actual presentations even of this part of modal logic the sign '□' commonly appears before non-statements such as '*p*' or '*p* ⊃ *p*', but there is no harm in this; such expressions are schemata, used diagrammatically and imagined replaced in any given case by actual statements of the depicted forms.

**3. Quantification in modal logic.** It will be convenient now to think of our basic modality not as '□' but as '◇' (possibility); the two are interdefinable, amounting respectively to '∼◇∼' and '∼□∼'. And let us think of our quantifiers as existential, since the universal quantifier '(*x*)' can be explained as '∼(∃*x*)∼'.

When '◇' is applied to a matrix within a quantification, the whole context admits, in certain cases, of reduction to the type of case dealt with in the preceding section. For example, the combined prefix '(∃*x*)◇' is equated by Miss Barcan<sup>7</sup> to '◇(∃*x*)', wherein '◇' applies no longer to the matrix but to the statement formed by quantifying it (supposing there are no further free variables); and accordingly the explanation of the preceding section can thereupon be brought to bear. If every matrix containing '◇' were transformable into a matrix containing '◇' only at the beginning, then repeated conversion of '(∃*α*)◇' to '◇(∃*α*)' would be adequate to explaining the general case; but unfortunately matrices are not generally thus transformable.

Equating '(∃*x*)◇' with '◇(∃*x*)' gives us no clue to the interpretation of such statements, for example, as:

$$(\exists x) (x \text{ is red} \cdot \diamond(x \text{ is round})).$$

No doubt a modal logician would regard this example as true, but only by virtue of some supplementary intuitive criterion—perhaps this:

(ii) *An existential quantification holds if there is a constant whose substitution for the variable of quantification would render the matrix true.*

This is at best a partial criterion (both in modal and non-modal logic), because of unnameable objects; an unnameable object might satisfy the matrix though there be no constant expression whose substitution for the variable would yield a

<sup>6</sup> I use the word 'matrix' (as in *Mathematical logic*) for one of the meanings of the ambiguous phrase 'propositional function.' A matrix is an expression which is like a statement except for containing, at grammatically permissible places, some free occurrences of variables of the kind that are admissible in quantifiers. Briefly, a matrix is a non-statement which can be turned into a statement by applying quantifiers.

<sup>7</sup> *Op. cit.*, theorem 38.

truth. But let us adopt (ii) as a partial criterion (sufficient condition) and see how it fares.

What I shall show is that it has queer ontological consequences. It leads us to hold that there are no concrete objects (men, planets, etc.), but rather that there are only, corresponding to each supposed concrete object, a multitude of distinguishable entities (perhaps "individual concepts," in Church's phrase). It leads us to hold, e. g., that there is no such ball of matter as the so-called planet Venus, but rather at least three distinct entities: Venus, Evening Star, and Morning Star.

To see this, let us use 'C' (for 'congruence') to express the relation which Venus, the Evening Star, and the Morning Star, e. g., bear to themselves and, according to empirical evidence, to one another. (It is the relation of *identity* according to materialistic astronomy, but let us not prejudge this.) Then

Morning Star C Evening Star .  $\square$  (Morning Star C Morning Star).

Therefore, according to (ii),

(1)  $(\exists x) (x \text{ C Evening Star} . \square (x \text{ C Morning Star}))$ .

But also

Evening Star C Evening Star .  $\sim \square$  (Evening Star C Morning Star),

so that, by (ii),

(2)  $(\exists x) (x \text{ C Evening Star} . \sim \square (x \text{ C Morning Star}))$ .

Since the matrix quantified in (1) and the matrix quantified in (2) are mutual contraries, the  $x$  whose existence is affirmed in (1) and the  $x$  whose existence is affirmed in (2) are two objects; so there must be at least two objects  $x$  such that  $x \text{ C Evening Star}$ . If we were to introduce the term 'Venus' we could infer a third such object in similar fashion.

Thus it is that the contemplated version of quantified modal logic is committed to an ontology which repudiates material objects (such as the Evening Star properly so-called) and leaves only multiplicities of distinct objects (perhaps the Evening-Star-concept, the Morning-Star-concept, etc.) in their place. For, the ontology of a logic is nothing other than the range of admissible values of the variables of quantification.<sup>8</sup>

A quite parallel conclusion could be drawn in higher logical types, to show that the contemplated version of quantified modal logic is committed to an ontology which repudiates classes and admits only attributes. But this consequence is likely to be regarded as less objectionable, from the point of view at least of modal logicians, than the consequence concerning individuals.

The modal logician who finds the repudiation of material objects (or, indeed, of classes) uncongenial may have recourse to either of the following alternatives.

(a) He may regard his quantified modal logic as only a fragment of the total logic to which he is prepared to subscribe, so that the undesirably limited ontology

<sup>8</sup> See my *Designation and existence*, *Journal of philosophy*, vol. 36 (1939), pp. 701-709.

of the former comes to be only a fragment of a more inclusive ontology which embraces also material objects (and perhaps even classes). Those variables of the total logic which do admit material objects (or classes) as values would then be withheld from quantified modal contexts, or limited to harmless manners of occurrence in them,<sup>9</sup> by special grammatical rules. The total logic would not be one in which we could meaningfully apply a modal operator to any matrix at will and then meaningfully quantify the result at will with respect to any free variable.

(b) He may insist on the universality of his quantified modal logic, but disavow the criterion (ii) which underlies my argument. But then we have yet to see what might plausibly be put forward in its stead.

HARVARD UNIVERSITY

---

<sup>9</sup> Such is Church's procedure in *A formulation of the logic of sense and denotation* (abstract), this JOURNAL, vol. 11 (1946), p. 31. I am indebted to Professor Church for several helpful letters in this connection.—I am also indebted, along more general lines, to Professor Rudolf Carnap; correspondence with him on modal logic over recent years has been very instrumental in clarifying my general position.