Handbook of Modal Logic – P. Blackburn et al. (Editors)

21

MODAL LOGIC AND PHILOSOPHY

Sten Lindström and Krister Segerberg

Modal logic is one of philosophy's many children. As a mature adult it has moved out of the parental home and is nowadays straying far from its parent. But the ties are still there: philosophy is important for modal logic, modal logic is important for philosophy. Or, at least, this is a thesis we try to defend in this chapter. Limitations of space have ruled out any attempt at writing a survey of all the work going on in our field — a book would be needed for that. Instead, we have tried to select material that is of interest in its own right or exemplifies noteworthy features in interesting ways. Here are some themes which have guided us throughout the writing:

- *The back-and-forth between philosophy and modal logic.* There has been a good deal of give-and-take in the past. Carnap tried to use his modal logic to throw light on old philosophical questions, thereby inspiring others to continue his work and still others to criticise it. He certainly provoked Quine, who in his turn provided — and continues to provide — a healthy challenge to modal logicians. And Kripke's and David Lewis's philosophies are connected, in interesting ways, with their modal logic. Analytic philosophy would have been a lot different without modal logic!

- *The interpretation problem.* The problem of providing a certain modal logic with an intuitive interpretation should not be conflated with the problem of providing a formal system with a model-theoretic semantics. An intuitively appealing model-theoretic semantics may be an important step towards solving the interpretation problem, but only a step. One may compare this situation with that in probability theory, where definitions of concepts like 'outcome space' and 'random variable' are orthogonal to questions about "interpretations" of the concept of probability.

- *The value of formalisation.* Modal logic sets standards of precision, which are a challenge to — and sometimes a model for — philosophy. Classical philosophical questions can be sharpened and seen from a new perspective when formulated in a framework of modal logic. On the other hand, representing old questions in a formal garb has its dangers, such as simplification and distortion.

- *Why modal logic rather than classical (first or higher order) logic?* The idioms of modal logic — today there are many! — seem better to correspond to human ways of thinking than ordinary extensional logic. (Cf. Chomsky's conjecture that the NP + VP pattern is wired into the human brain.)

In his *An Essay in Modal Logic* [107] von Wright distinguished between four kinds of modalities: *alethic* (modes of truth: necessity, possibility and impossibility), *epistemic* (modes of being known: known to be true, known to be false, undecided), *deontic* (modes of obligation: obligatory, permitted, forbidden) and *existential* (modes of existence: universality, existence, emptiness). The existential modalities are not usually counted as modalities, but the other three categories are exemplified in three sections into which this chapter is divided. Section 1 is devoted to alethic modal logic and reviews some main themes at the heart of philosophical modal logic. Sections 2 and 3 deal with topics in epistemic logic and deontic logic, respectively, and are meant to illustrate two different uses that modal logic or indeed any logic can have: it may be applied to already existing (non-logical) theory, or it can be used to develop new theory.

1 ALETHIC MODAL LOGIC

In this part we consider the challenge that Quine posed in 1947 to the advocates of modal logic to provide an account of modal notions that is intuitively clear, allows "quantifying in", and does not presuppose intensional entities. The modal notions that Quine and his contemporaries were primarily concerned with in the 1940's were, broadly speaking, the logical modalities rather than the metaphysical ones that have since come to prevail. In the 1950's modal logicians responded to Quine's challenge by providing quantified modal logic with model-theoretic semantics of various types. In doing so they also, explicitly or implicitly, addressed Quine's interpretation problem. Here we shall consider the approaches developed by Carnap in the late 1940's, and by Kanger, Hintikka, Montague, and Kripke in the 1950's and early 1960's, and discuss to what extent these approaches were successful in meeting Quine's doubts about the intelligibility of quantified modal logic.

It is useful to divide the reactions to Quine's challenge into two periods. During the first period modal logicians provided modal logic with formal semantics as just mentioned. In the second period philosophers — inspired by the success of possible worlds semantics — came to take the notion of a possible world seriously as a tool for philosophical analysis. Philosophical analyses in terms of possible worlds were provided for many concepts of central philosophical importance: propositional attitudes [42, 43, 45], metaphysical necessity, identity, and naming [69, 70], "intensional entities" like propositions, properties and events [84, 61, 102, 103], counterfactual conditionals and causality [77, 78], supervenience [62]. At the same time the notion of a possible world itself came in for philosophical analysis. The problems of giving a satisfactory analysis of this notion indicates that Quine's interpretational challenge is still alive. The basic philosophical questions surrounding the notions of alethic necessity and possibility are as puzzling as ever! We end this section by discussing the relationship between the logical and metaphysical interpretation of the alethic modalities.

1.1 The search for the intended interpretation

Starting with the work of C. I. Lewis, an immense number of formal systems of modal logic have been constructed based on classical propositional or predicate logic. The originators of modern modal logic, however, were not very clear about the intuitive meaning of the symbols $\Box$ and $\Diamond$, except to say that these should stand for some kind of necessity and possibility, respectively. For instance, in *Symbolic Logic* [72], Lewis and Langford write:

> It should be noted that the words "possible", "impossible" and "necessary" are highly ambiguous in ordinary discourse. The meaning here assigned to $\Diamond p$ is a *wide* meaning of "possibility" — namely, logical conceivability or the absence of self-contradiction. (160–61)

This situation led to a search for more rigorous interpretations of modal notions. Gödel [35] suggested interpreting the necessity operator $\Box$ as standing for provability (*informal provability* or, alternatively, *formal provability* in a fixed formal system), a suggestion that subsequently led to the modern *provability interpretations* of Solovay, Boolos and others.[1]

After Tarski [105, 106] had developed rigorous notions of satisfaction, truth and logical consequence for classical extensional languages, the question arose whether the same methods could be applied to the languages of modal logic and related systems. One natural idea, that occurred to Carnap in the 1940's, was to let $\Box\varphi$ be true of precisely those formulæ φ that are *logically valid* (or logically true) according to the standard semantic definition of logical validity. This idea led him to the following semantic clause for the operator of logical necessity:

> $\Box\varphi$ is true in an interpretation $\mathcal{I}$ iff φ is true in every interpretation $\mathcal{I}'$.

This kind of approach, which we may call the *validity interpretation*, was pursued by Carnap, using so-called state descriptions, and subsequently also by Kanger [53, 54] and Montague [83], using Tarski-style model-theoretic interpretations rather than state descriptions. In Hintikka's and Kanger's early work on modal semantics other interpretations of $\Box$ were also considered, especially, epistemic ('It is known that φ') and deontic ones ('It ought to be the case that φ'). In order to study these and other non-logical modalities, the introduction by Hintikka and Kanger of *accessibility relations* between

[1] Cf. [101] and [13].

possible worlds (models, domains) was crucial. Finally, Kripke [66, 67, 68] introduced the kind of model structures that are nowadays the standard formal tool for the model-theoretic study of modal and related non-classical logics: Kripke models. Thus Kripke gave possible worlds semantics its modern and mature form.

In Carnap's, Kanger's and Montague's early theories, the space of possibilities (the "possible worlds") is represented by one comprehensive collection containing *all* state descriptions, domains, or models, respectively. Hence, every state description, domain, or model is thought of as representing a genuine possibility. Hintikka, Kripke and modern possible worlds semantics are instead working with semantic interpretations in which the space of possibilities is represented by an arbitrary non-empty set **K** of model sets (in the case of Hintikka) or "possible worlds" (Kripke). Following Hintikka's [46, 47] terminology, one may say that the early theories of Carnap, Kanger, and Montague were considering *standard interpretations* only, where one quantifies over what is, in some formal sense, *all* the possibilities. In the possible worlds approach, one also considers *non-standard* interpretations, where arbitrary non-empty sets of possibilities are considered.[2] The consideration of interpretations (model structures) that are non-standard in this sense — in combination with the use of accessibility relations between worlds in each interpretation — made it possible for Kripke [64, 67, 68] to prove completeness theorems for various systems of propositional and quantified modal logic (**T**, **B**, **S4**, etc.).

1.2 Carnap's formal semantics for quantified modal logic

The proof theoretic study of quantified modal logic was pioneered by Ruth Barcan Marcus [5, 6, 7] and Rudolf Carnap [16, 17] who were the first to formulate axiomatic systems that combined quantification theory with (**S4**- and **S5**-type) modal logic. The attempts to interpret quantified modal logic by means of formal semantic methods also began with Carnap.

Carnap's project was not only to develop a semantics (in the sense of Tarski) for intensional languages, but also to use metalinguistic notions from formal semantics to throw light on the modal ones. In 'Modalities and quantification' from 1946 he writes:

> It seems to me ... that it is not possible to construct a satisfactory system before the meaning of the modalities are sufficiently clarified. I further believe that this clarification can best be achieved by correlating each of the modal concepts with a corresponding semantical concept (for example, necessity with **L**-truth).

In [16, 17] Carnap presented a formal semantics for logical necessity based on Leibniz's old idea that a proposition is necessarily true if and only if it is true in all possible worlds. Suppose that we are considering a first-order predicate language $\mathcal{L}$ with predicate symbols and individual constants, but no function symbols. In addition to Boolean connectives, quantifiers and the identity symbol $=$ (considered as a logical symbol), the language $\mathcal{L}$ also contains the modal operator $\Box$ for logical necessity. We assume that $\mathcal{L}$ comes with a *domain of individuals* D and that there is a one-to-one correspondence between the individual constants of $\mathcal{L}$ and the individuals in D. Intuitively speaking, each individual in D has exactly one individual constant as its (canonical) name. A *state description* S for $\mathcal{L}$ is simply a set of (closed) atomic sentences of the form $P(a_1, \ldots, a_n)$, where P is

[2]For the standard/non-standard distinction, see also [23].

an n-ary predicate in $\mathcal{L}$ and $a_1, \ldots, a_n$ are individual constants in $\mathcal{L}$.[3] Carnap [17, p. 9] writes "...the state descriptions represent Leibniz's possible worlds or Wittgenstein's possible states of affairs".

In order to interpret quantification, Carnap introduced the notion of an *individual concept* (relative to $\mathcal{L}$): An individual concept is simply a function f that assigns to every state description S an individual constant $f(S)$ (representing an individual in D). Intuitively speaking, individual concepts are functions from possible worlds to individuals. According to Carnap's semantics, individual variables are assigned values *relative to* state descriptions. An *assignment* is a function g that to every state description S and every individual variable x assigns an individual constant $g(x, S)$. Intuitively, $g(x, S)$ represents the individual that is the value of x under the assignment g in the possible world represented by S. We may speak of $g(x, S)$ as the *value extension* of x in S relative to g. Analogously, the individual concept $(\lambda S)g(x, S)$ that assigns to every state description S the value extension of x in S relative to g, we call the *value intension* of x relative to g. Thus, according to Carnap's semantics a variable is assigned both a *value intension* and a *value extension* [17, p. 45]. The value extension assigned to a variable in a state description S is simply the value intension assigned to the variable applied to S.

With these notions in place, we can define what it means for a formula φ of $\mathcal{L}$ to be *true* in a state description relative to an assignment g (in symbols, $S\varphi[g]$).

For atomic formulæ of the form $P(t_1, \ldots, t_n)$, where $t_1, \ldots, t_n$ are individual terms, i.e., variables or individual constants, we have:

(1) $S \vDash P(t_1, \ldots, t_n)[g]$ iff $P(S(t_1, g), \ldots, S(t_n, g)) \in S$.

Here, $S(t_i,\ g)$ is the extension of the term t_i in the state description S relative to the assignment g. Thus, if t_i is an individual constant, then $S(t_i, g)$ is t_i itself; and if t_i is a variable, then $S(t_i, g) = g(t_i, S)$.

The semantic clause for the identity symbol is:

(2) $S \vDash (t_1 = t_2)[g]$ iff $S(t_1, g) = S(t_2, g)$.

That is, the identity statement $t_1 = t_2$ is true in a state description S relative to an assignment g if and only if the terms t_1 and t_2 have the same extension in S relative to g.

The clauses for the Boolean connectives are the usual ones. Carnap's clause for the universal quantifier is:

(3) $S \vDash \forall x\varphi[g]$ iff for every assignment g' such that $g =_x g', S \vDash \varphi[g']$,

where $g =_x g'$ means that the assignments g and g' assign the same value intensions to all the variables that are distinct from x and possibly assign different value intensions to x. Intuitively, then $\forall x\varphi(x)$ may be read: "for every assignment of an individual concept to $x, \varphi(x)$".

Finally, the semantic clause for the necessity operator is the expected one:

(4) $S \vDash \Box\varphi[g]$ iff, for every state description $S', S' \vDash \varphi[g]$.

[3] Actually Carnap's state descriptions are sets of literals (i.e., either atomic sentences or negated atomic sentences) that contain for each atomic sentence either it or its negation. However, for our purposes we may identify a state description with the set of atomic sentences that it contains.

That is, the modal formula 'it is (logically) necessary that φ' is true in a state description S (relative to an assignment g) if and only if φ is true in every state description S' (relative to g).

A formula φ is *true in a state description* S (in symbols, $S \models \varphi$) if it is true in S relative to every assignment. *Logical truth* (logical validity) is defined as truth in all state descriptions. We write $\models \varphi$ for φ being logically true.

Carnap's semantics satisfies the following principles:

(5) All truth-functional tautologies are logically true.

(6) The set of logical truths is closed under modus ponens.

(7) The standard principles of quantification theory (without identity) are valid. In particular,

(US) $\forall x\varphi(x) \to \varphi(x)$ (*Universal Specification*)
(EG) $\varphi(t/x) \to \exists x\varphi$ *(Existential Generalisation)*
(where t is substitutable for x in φ)

hold without restrictions.

It is easy to verify that $\Box$ satisfies the usual laws of the system **S5**, together with the so-called Barcan formula and its converse, and the rule of necessitation:

(K) $\models \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$.
(T) $\models \Box\varphi \to \varphi$
(S4) $\models \Box\varphi \to \Box\Box\varphi$.
(S5) $\models \neg\Box\varphi \to \Box\neg\Box\varphi$
(BF) $\models \forall x\Box\varphi(x) \to \Box\forall x\varphi(x)$. (*The Barcan formula*)
(CBF) $\models \Box\forall x\varphi(x) \to \forall x\Box\varphi(x)$. (*The Converse Barcan formula*)
(Nec) If $\models \varphi$, then $\models \Box\varphi$.

Notice that the Barcan formula (BF) and its converse (CBF) are schemata rather than single formulæ.

The following schemata are also valid in Carnap's semantics:

(8) $\models \Box\varphi$ iff $\models \varphi$.

(9) $\models \neg\Box\varphi$ iff $\not\models \varphi$.

(10) Either $\models \Box\varphi$ or $\models \neg\Box\varphi$.

For identity, we have:

(LI) $\models t = t$. (*Law of Identity*)

However, the unrestricted principle of *indiscernibility of identicals* is not valid in Carnap's semantics. In other words, the following principle does not hold for all formulæ φ:

$(I =)$ $\models \forall x\forall y(x = y \to (\varphi(x/z) \to \varphi(y/z)))$.

Instead, we have a restricted version of $(I =)$:

$(I =_{\text{restr}})$ $\models \forall x\forall y(x = y \to (\varphi(x/z) \to \varphi(y/z)))$, provided that φ does not contain any occurrences of $\Box$.

For the unrestricted case, we only have:

$(I\Box =) \models \forall x \forall y(\Box(x = y) \rightarrow (\varphi(x/z) \rightarrow \varphi(y/z)))$.

The following principle is of course not valid according to Carnap's semantics:

$(\Box =)$ $\quad \forall x \forall y(x = y \rightarrow \Box(x = y))$. $\quad$ (*Necessity of Identity*)

In the presence of the other principles, it is equivalent to the unrestricted principle of indiscernibility of identicals. Nor do we have:

$(\Box \neq)$ $\quad \forall x \forall y(x \neq y \rightarrow \Box(x \neq y))$. $\quad$ (*Necessity of Non-Identity*)

In view of Church's undecidablity theorem for the predicate calculus, it is easy to prove that Carnap's quantified modal logic is not axiomatizable. For every sentence φ of predicate logic φ, either $\Box\varphi$ or $\neg\Box\varphi$ is true in every state description. So, if Carnap's logic were axiomatizable, then we could decide effectively whether φ is provable in predicate logic. But this is contrary to Church's theorem.

THEOREM 1. *The set of all logically true sentences according to Carnap's semantics is not recursively enumerable, so there is no formal axiomatic system with this set as its theorems.*

Carnap introduced the notion of a *meaning postulate* to account for analytic connections between the non-logical symbols of a predicate language. Thus, suppose that *MP* is the set of all the meaning postulates of a given language $\mathcal{L}$. *MP* is then a set of sentences in the non-modal fragment of $\mathcal{L}$. We say that a state description S is *admissible* if *MP* $\cup S$ is consistent. Then, we can interpret $\Box$ as 'analytic necessity' by modifying clause (4) above to:

(4′) $S \models \Box\varphi$ iff, for every admissible state description $S', S' \models \varphi$.

We also say that φ is *analytically true* iff φ is true in all admissible state descriptions. In the modified semantics, we have:

$S \models \Box\varphi$ iff φ is analytically true.
$S \models \neg\Box\varphi$ iff φ is not analytically true.

Carnap's semantics for the quantifiers can be understood in two ways. The most straightforward interpretation is to say that the quantifiers simply range over individual concepts. Sometimes Carnap himself characterises his interpretation of the quantifiers in this way, and this is how Quine describes it. There is, however, another more subtle interpretation according to which every individual term, including the (free) variables, has a double semantic role given by its extension and its intension, respectively. Each variable has a value extension as well as a value intension. According to this interpretation — which presumably is the one that Carnap really had in mind — it is simply wrong to ask for *the* range of the individual variables. In ordinary extensional contexts the variables can be thought of as ranging over ordinary individuals. However, in intensional contexts the intensions associated with the variables come into play. This is what explains why the principle $(\Box =)$ fails.

Carnap's interpretation of the quantifiers can still be criticised for being unintuitive. The problem is that he lacks a way of discriminating between those individual concepts that, intuitively speaking, pick out one and the same individual in all possible worlds and those that don't. Suppose that we have assigned to the variable x as its value intension the individual concept: *the number of planets*. Relative to this assignment it is true that:

(1) $x = 9 \wedge \neg\Box(x = 9)$.

However, there is no *object* that has the property of being identical with 9 but doesn't have this property necessarily. So from (1) it should not follow that:

(2) $\exists x(x = 9 \wedge \neg\Box(x = 9))$.

But of, course, on Carnap's interpretation of the quantifiers, (2) is a logical consequence of (1). Intuitively, one should be able to make the inference from (1) to (2) only if the concept assigned to x in (1) is, what might be called, a *logically rigid concept*, i.e. a concept that picks out the same individual relative to every state description.[4]

1.3 Quine's interpretational challenge

Quine's criticism of quantified modal logic comes in different strands. First, there is the simple observation that classical quantification theory with identity cannot be applied to a language in which substitutivity of identicals for singular terms fails. It seems, from the so-called Morning Star Paradox, that either universal specification (US) (and its mirror image: existential generalisation (EG)) or indiscernibility of identicals, (I=), has to be given up. This observation gives rise to the following weak, and apparently uncontroversial, Quinean claim: Classical quantification theory (with identity and individual constants) cannot be combined with non-extensional operators (i.e., operators for which substitutivity of identicals for singular terms fail) without being modified in some way. This weak claim already gives rise to the challenge of extending quantification theory in a consistent way to languages with non-extensional operators.

In addition to the weak claim, there is the much stronger claim that one sometimes can find in Quine's early works, that objectual quantification into non-extensional (so called "opaque") constructions simply does not make sense [91, 93, 94]. The argument for this claim is based on the idea that occurrences of variables inside of opaque constructions do not have purely referential occurrences, i.e., they do not serve simply to refer to their objects, and cannot therefore be bound by quantifiers outside of the opaque construction. Thus quantifying into contexts governed by non-extensional operators would be like trying to quantify into quotations. This claim is hardly credible in the face of the multitude of quantified intensional logics that have been developed since it was first made, and we take it to be refuted by the work of among others, David Kaplan [59, 61] and Kit Fine [26, 27].[5]

Then, there is Quine's claim that quantified modal logic is committed to *Aristotelian essentialism*, i.e., the view that it makes sense to say of an object, quite independently of how it is described, that it has certain of its traits necessarily, and others only contingently. Aristotelian essentialism, however, comes in stronger and weaker forms. Kripke's "metaphysical necessity" of *Naming and Necessity* represents a strong form of essentialism, while there are weaker forms according to which only logical properties that are shared by all individuals are essential. A quantified modal logic needs only be committed to this weak relatively benign form of essentialism.

---

[4]The notion of a logically rigid concept is closely related Carnap's [17, Part II] notion of an L-determinate intension. Intuitively, an L-determinate intension picks out the same extension in every state description. Thus, Carnap's notion of L-determinacy may be viewed as a precursor of Kripke's notion of rigidity.

[5]See also Burgess [14] and Neale [86] for recent evaluations of Quine's criticism of quantified modal logic.

Here we shall only consider the specific criticism that Quine directed in 1947 toward quantification into contexts of logical or analytical necessity. In his paper 'The problem of interpreting modal logic' from 1947, Quine formulates what one might call *Quine's challenge* to the advocates of quantified modal logic:

> 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 ... to provide somewhat the type of explanation required. When modal logic is extended (as by Miss Barcan) 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.

What Quine demands of the modal logicians is nothing less than an explanation of the notions of quantified modal logic in non-modal terms. Such an explanation should satisfy the following requirements:

(i) It should be expressed in an extensional language. Hence, it cannot use any non-extensional constructions.

(ii) The explanation should be allowed to use concepts from the 'theory of meaning' like analyticity and synonymy applied to expressions of the metalanguage. Quine is, of course, quite sceptical about the intelligibility of these notions as well. But he considers it to be progress of a kind, if modal notions could be explained in these terms.

(iii) The explanation should make sense of sentences like:

$$\exists x(x \text{ is red } \wedge \Diamond(x \text{ is round})),$$

in which a quantifier outside a modal operator binds a variable within the scope of the operator and the quantifier ranges over ordinary physical objects (in distinction from Frege's "Sinne" or Carnap's "individual concepts"). In other words, the explanation should make sense of 'quantifying in' in modal contexts.

Quine [92] — like Carnap before him — starts out from a metalinguistic interpretation of the necessity operator $\Box$ in terms of the predicate '... is analytically true'. Disregarding possible complications in connection with the interpretation of iterated modalities, we have for sentences φ of the object language:

$$\text{`}\Box\varphi\text{' is true iff } \varphi \text{ is analytically true.}$$

Now Quine argues for the thesis that it is impossible to combine analytical necessity with a standard theory of quantification (over physical objects). The argument (a variation of "the Morning Star Paradox") is based on the premises:

(1) $\Box$(Hesperus = Hesperus)

(2) Phosphorus = Hesperus

(3) $\neg\Box(\text{Phosphorus} = \text{Hesperus})$,

where 'Phosphorus' and 'Hesperus' are two proper names (individual constants) and $\Box$ is to be read 'It is analytically necessary that'. We assume that 'Phosphorus' is used by the language community as a name for a certain bright heavenly object sometimes visible in the morning and that 'Hesperus' is used for some bright heavenly object sometimes visible in the evening. Unbeknownst to the community, however, these objects are one and the same, namely, the planet Venus. 'Hesperus = Hesperus' being an instance of the Law of Identity is clearly an analytic truth. It follows that the premise (1) is true. (2) is true, as a matter of fact. 'Phosphorus = Hesperus' is obviously not an analytic truth, 'Phosphorus' and 'Hesperus' being two different names with quite distinct uses. So, (3) is true.

From (1), (2), (3) and the Law of Identity, we infer by sentential logic:

(4) $\text{Phosphorus} = \text{Hesperus} \wedge \neg\Box(\text{Phosphorus} = \text{Hesperus})$,

(5) $\text{Hesperus} = \text{Hesperus} \wedge \Box(\text{Hesperus} = \text{Hesperus})$.

Applying (EG) to (4) and (5), we get:

(6) $\exists x(x = \text{Hesperus} \wedge\neg\Box(x = \text{Hesperus}))$,

(7) $\exists x(x = \text{Hesperus} \wedge \Box(x = \text{Hesperus}))$.

As Quine [92] points out, however, (6) and (7) are incompatible with interpreting $\forall x$ and $\exists x$ as objectual quantifiers meaning "for all objects x (in the domain D)" and "for at least one object x (in D)" and letting the identity sign stand for genuine identity between objects (in D). Because, under this interpretation, (6) and (7) imply that one and the same object, Hesperus, both is and is not necessarily identical with Hesperus, which seems absurd.

The following are classical proposals for solving Quine's interpretational challenge:

(i) *Russell–Smullyan* (Smullyan [99]). According to this proposal, all singular terms except variables are treated as *Russellian terms*, i.e., as "abbreviations" of definite descriptions that are eliminated from the language by means of contextual definition à la Russell. If we let 'Hesperus' and 'Phosphorus' be Russellian terms having minimal scope everywhere — which clearly corresponds to the intended reading — then the inference will not go through (i.e., once the Russell terms have been contextually eliminated): the (EG)-steps above will not correspond to valid steps in primitive notation. With this treatment of singular terms, the paradox is avoided. One has the feeling, however, that the problem has been circumvented rather than solved.

(ii) *Carnap* (at least the way Quine reads him): The individual variables are not taken to range over physical objects, but instead over individual concepts. According to this reading, the names 'Phosphorus' and 'Hesperus' stand for different but coextensive individual concepts. The identity sign is interpreted not as a genuine identity between physical objects but as coextensionality between individual concepts. That is, an identity statement '$u = v$' is true if and only if the terms 'u' and 'v' stand for coextensive individual concepts. According to this interpretation, (6) and (7) mean:

(6′) There is an individual concept x which actually coincides with the individual concept Hesperus but does not do so by analytical necessity.

(7′) There is an individual concept x which not only happens to coincide with the individual concept Hesperus but does so by analytic necessity.

No contradiction ensues from these two statements. The price for this interpretation, however, seems to be as Quine expresses it: "a curiously idealistic ontology which repudiates material objects".

1.4 The advent of possible worlds semantics

1.4.1 Semantics for quantified modal logic in 1957: Hintikka and Kanger

1957 was a pivotal year in the history of modal logic.[6] In that year Stig Kanger published his dissertation *Provability in Logic* and a number of other papers where he outlined a new model-theoretic semantics for quantified modal logic. In the same year, Jaakko Hintikka published two papers on the semantics of quantified modal logic: 'Modality as referential multiplicity' and 'Quantifiers in deontic logic' (Hintikka [39, 40]). There are some striking parallels between these works by Hintikka and Kanger, but there are also notable differences.

Hintikka and Kanger had both done important and closely similar work in non-modal predicate logic. Using so-called model sets (nowadays often called "Hintikka sets" or "downward saturated sets") for predicate logic, Hintikka [38] had developed a new complete and effective proof procedure for predicate logic.

Let $\mathcal{L}$ be a language of predicate logic with identity and let U be a non-empty set of individual constants that do not belong to $\mathcal{L}$. A *model set* (over U) is a set m of sentences of the expanded language $\mathcal{L}_U$ satisfying the following conditions:[7]

$(C.\neg)$	if $\neg\varphi \in m$, then $\varphi \notin m$,
$(C.\neg\neg)$	if $\neg\neg\varphi \in m$, then $\varphi \in m$,
$(C.\wedge)$	if $\varphi \wedge \psi \in m$, then $\varphi \in m$ and $\psi \in m$,
$(C.\neg\wedge)$	if $\neg(\varphi \wedge \psi) \in m$, then $\neg\varphi \in m$ or $\neg\psi \in m$,
$(C.\forall)$	if $\forall x\varphi \in m$, then for every constant a in U, $\varphi(a/x) \in m$,
$(C.\neg\forall)$	if $\neg\forall x\varphi \in m$, then for some constant a in U, $\neg\varphi(a/x) \in m$,
$(C.=)$	for no individual constant a in $\mathcal{L}_U, a \neq a \in m$,
$(C.\mathrm{Ind})$	if $\varphi(a/x) \in m$, where φ is atomic, and $a = b \in m$, then $\varphi(b/x) \in m$.

Hintikka showed, what nowadays goes under the name *Hintikka's lemma*, namely, that a set Γ of sentences is satisfiable (true in some Tarski-style model) iff it can be imbedded in a model set over some non-empty set U of (new) individual constants. Furthermore, he provided an effective proof procedure for classical predicate logic. The method is very similar to the nowadays more familiar semantic tableaux method of Beth [11].

Hintikka [38, p. 47] points out that there is a close connection between his proof procedure and proofs in Gentzen's sequent calculus. The systematic search for a counterexample of a formula φ corresponds to the backward application of the rules of Gentzen's

[6] See [24] for a comprehensive historical account of the development of possible worlds semantics. For a mathematical exposition of the development of modal logic, see [36]).

[7] Here we have assumed that $\neg, \wedge$ and $\forall$ are primitive and that $\vee, \rightarrow$ and $\exists$ are introduced as abbreviations in the usual way. For other choices of primitive logical constants, the definition of a model set has to be adjusted accordingly.

cut-free calculus for predicate logic. As a matter of fact, Kanger in *Provability in Logic* [53] provided an elegant effective proof procedure for classical predicate logic based on a sequent calculus that is equivalent to Hintikka's.

Hintikka's formal semantics for modal logic. When studying classical predicate logic, Hintikka and Kanger used strikingly similar techniques and obtained similar results. However, their approaches to modal logic were different. Kanger started out from the work of Tarski and set himself the task of extending the method of Tarski-style truth-definitions to predicate languages with modal operators. Hintikka, on the other hand, generalised his method of model sets to the case of modal logic. In doing so he invented the notion of a *model system*. Roughly speaking, a model system consists of a set Ω of model sets and a binary relation R defined between the members of Ω. Different versions of Hintikka's semantics impose different conditions on model sets, but in order simplify the exposition, we can say that a model system is an ordered pair $\mathcal{S} = \langle \Omega, R \rangle$, such that:

(a) Ω is a non-empty set of model sets for $\mathcal{L}$,

(b) R is a binary relation between the members of Ω (the alternativeness relation),

(c) for all $m \in \Omega$, if $\Box\varphi \in m$, then for all $n \in \Omega$ such that $mRn, \varphi \in n$,

(d) for all $m \in \Omega$, if $\neg\Box\varphi \in m$, then $\neg\varphi \in n$, for some $n \in \Omega$ such that mRn.

Hintikka thought of the members of Ω as partial descriptions of possible worlds. A set Γ of sentences is *satisfiable* (in the sense of Hintikka) iff there exists a model system $\mathcal{S} = \langle \Omega, R \rangle$ and a model set $m \in \Omega$ such that $\Gamma \subseteq m$. A sentence φ is valid iff the set $\{\neg\varphi\}$ is not satisfiable.

Hintikka [40] sketched a tableaux-style method of proving completeness theorems in modal logic. The idea is a generalisation of his proof procedure for first order logic. Hintikka [41] states (without formal proofs) that the systems **T**, **B**, **S4**, **S5** for sentential logic are sound and complete with respect to the Hintikka-style semantics where R is assumed to be reflexive, symmetric, reflexive and transitive and an equivalence relation, respectively. Rigorous completeness proofs using the tableaux method were published by Kripke, [64], for the case of quantified S5, and for numerous systems of propositional modal logic in [67, 68].[8]

An important difference between Hintikka's semantics for modal logic, on the one hand, and the ones developed by Carnap, Kanger and Montague [83], on the other, is that Hintikka allows the space of possibilities Ω to vary from one system to another. The only requirement is that Ω is a non-empty set satisfying the constraints (b), (c) and (d) above. In the formal semantics of Carnap, Kanger and Montague, on the other hand, the space of possibilities is fixed once and for all to be the set of all state descriptions (Carnap), the class of all systems (or alternatively, domains) (Kanger), or all first-order models over a given domain (Montague). One could say that Carnap, Kanger and Montague only allow interpretations of modalities that are in a sense *standard* and disallow *non-standard interpretations*. Thus, the relationship between Hintikka's semantics (and the one later developed by Kripke) and the ones developed by Carnap, Kanger and Montague is analogous to that between *standard* and *non-standard* semantics for higher-order

[8] In [65], Kripke announces a great number of completeness results in modal propositional logic. He also notes "For systems based on **S4**, **S5**, and **M**, similar work has been done independently and at an earlier date by K. J. J. Hintikka".

predicate logic. This distinction between the various approaches has been emphasised by Cocchiarella [23] and Hintikka [46]. Allowing non-standard interpretations for modal logics, of course, facilitated the proofs of completeness results, since the logics for logical or analytical necessity corresponding to the standard semantics are in general not recursively enumerable.

Kanger's Tarski-style semantics for quantified modal logic. Kanger's ambition was to provide a language of quantified modal logic with a model-theoretic semantics à la Tarski.[9]

A Tarski-style interpretation for a first-order predicate language $\mathcal{L}$ consists of a non-empty domain D and an assignment of appropriate extensions in D to every non-logical symbol and variable of $\mathcal{L}$. Kanger's basic idea was to relativise the notion of extension to various possible domains. In other words, he thought of an interpretation for a given language $\mathcal{L}$ as a *function* that *simultaneously* assigns extensions to the non-logical symbols and variables of $\mathcal{L}$ for *every* possible domain. Such a function Kanger called a (*primary*) *valuation*. Formally, a valuation for a language L of quantified modal logic is a function v which for *every* non-empty domain D assigns an appropriate extension in D to every individual constant, individual variable, and predicate constant in $\mathcal{L}$. Kanger also introduced the notion of a *system* $\mathcal{S} = \langle D, v\rangle$ consisting of a designated domain D and a valuation v. Notice that v does not only assign extensions to symbols relative to the designated domain D, but relative to *all* domains simultaneously.

Kanger then defined the notion of a formula φ being *true in a system* $\mathcal{S} = \langle D, v\rangle$ (in symbols, $\mathcal{S} \vDash \varphi$):

(1) $\mathcal{S} \vDash (t_1 = t_2)$ iff $v(D, t_1) = v(D, t_2)$,

(2) $\mathcal{S} \vDash P(t_1, \ldots, t_n)$ iff $\langle v(D, t_1), \ldots, v(D, t_n)\rangle \in v(D, P)$,

(3) $\mathcal{S} \nvDash \bot$,

(4) $\mathcal{S} \vDash (\varphi \to \psi)$ iff $\mathcal{S} \nvDash \varphi$ or $\mathcal{S} \vDash \psi$

(5) $\langle D, v\rangle \vDash \forall x\varphi$ iff $\langle D, v'\rangle \vDash \varphi$, for each v' such that $v' =_x v$,

(6) for every operator $\Box, \mathcal{S} \vDash \Box\varphi$ iff $\forall \mathcal{S}'$, if $\mathcal{S} R_\Box \mathcal{S}'$, then $\mathcal{S}' \vDash \varphi$.

Explanation: v' is like v except possibly at x (also written, $v' =_x v$) if and only if, for every domain U and every variable y other than $x, v'(U, y) = v(U, y)$. In the above definition, $R_\Box$ is a binary relation between systems that is associated with the modal operator $\Box$. $R_\Box$ is what is nowadays called the *accessibility relation* associated with the operator $\Box$. Kanger points out that by imposing certain formal requirements on the accessibility relation, like reflexivity, symmetry, transitivity, etc., one can make the operator satisfy corresponding well-known axioms of modal logic.

One source of inspiration for Kanger's use of accessibility relations in modal logic was no doubt the work of Jónsson and Tarski [52] on representation theorems for Boolean algebras with operators.[10] Jónsson and Tarski define operators $\Diamond$ on arbitrary subsets X of a set U in terms of binary relations $R \subseteq U \times U$ in the following way:

$$\Diamond X = \{x \in U : \exists y \in X(yRx)\},$$

[9] Cf. Kanger [53, 54, 55, 56, 57]). See also Lindström [81] for a more extensive discussion of Kanger's approach to quantified modal logic.

[10] On [53, p. 39] Kanger makes an explicit reference to Jónsson and Tarski [52].

that is $\Diamond X$ is the image of X under R. They also point to correspondences between properties of $\Diamond$ and properties of R. Among other things, they prove a representation theorem for so-called closure algebras that, via the Tarski-Lindenbaum construction, yields the completeness theorem for propositional **S4** with respect to Kripke models with a reflexive and transitive accessibility relation. However, Jónsson and Tarski do not say anything about the relevance of their work to modal logic.

Among the modal operators in $\mathcal{L}$, Kanger introduced two designated ones, **N** ("analytic necessity") and **L** ("logical necessity"), with the following semantic clauses:

$$\langle D, v\rangle \vDash \mathbf{N}\varphi \text{ iff for every domain } D', \langle D', v\rangle \vDash \varphi$$
$$\langle D, v\rangle \vDash \mathbf{L}\varphi \text{ iff for every system } \mathcal{S}, \mathcal{S} \vDash \varphi.$$

A formula φ is *true* in a system $\langle D, v\rangle$ iff $\langle D, v\rangle \vDash \varphi$. A formula φ is said to be *valid (logically true)* if it is true in every system $\langle D, v\rangle$. A formula φ is a *logical consequence* of a set Γ of formulæ (in symbols, $\Gamma \vDash \varphi$) if φ is true in every system in which all the formulæ in Γ are true.

In order to get a clearer understanding of Kanger's treatment of quantification, we shall speak of selection functions that pick out from each domain an element of that domain as *individual concepts.* We can think of a system $\mathcal{S} = \langle D, v\rangle$ as assigning to each individual constant c the individual concept $\{\langle D, v(D, c)\rangle : D \text{ is a domain}\}$ and to each variable x the individual concept $\{\langle D, v(D, x)\rangle : D \text{ is a domain}\}$. The formula $P(t_1, \ldots, t_n)$ is true in $\mathcal{S} = \langle D, v\rangle$ if and only if the individual concepts designated by $t_1, \ldots, t_n$ pick out objects in the domain D that stand in the relation $v(D, P)$ to each other. The identity symbol designates the relation of *coincidence* between individual concepts (at the "actual" domain D). That is, $t_1 = t_2$ is true in a system $\mathcal{S} = \langle D, v\rangle$ if and only if the individual concepts designated by t_1 and t_2, respectively, pick out one and the same object in the domain D of $\mathcal{S}$.

The universal quantifier $\forall x$ can now be thought of as an objectual quantifier that ranges not over the "individuals" in the "actual" domain D, but over the (constant) domain of all individual concepts. That is, $\forall x\varphi$ is true in a system $\langle D, v\rangle$ if and only if φ is true in every system that is exactly like $\langle D, v\rangle$ except, possibly, for the individual concept that it assigns to the variable x.

Kanger's solution to Quine's paradox of identity is essentially the same as Carnap's. Quine's objection to Kanger would therefore be the same as to Carnap: Kanger's quantifiers do not range over ordinary individuals but over individual concepts instead. Moreover, Kanger's treatment of quantification in modal contexts does not provide any means of *identifying* individuals from one domain to another. Hence there is no way of saying in Kanger's modal language that *one and the same* individual has a property P and possibly could have lacked P. That is, neither Carnap's nor Kanger's semantics can account for modality *de re.*

1.4.2 Hintikka's response to Quine's challenge

Quine's interpretational challenge seemed to place the advocates of quantified modal logic in a dilemma. They would either have to accept standard quantification theory (with the usual laws of universal instantiation, existential generalisation and indiscernibility of identicals) and reject quantified modal logic, or accept a quantified modal logic, where the quantifiers were interpreted in a non-standard way à la Carnap as ranging over intensional entities (individual concepts), rather than over robust extensional entities as Quine would demand.

Hintikka [39, 40], however, rejected the terms in which Quine's interpretational challenge was stated. First of all he broadened the discussion by not only considering the logical modalities and Quine's metalinguistic interpretation of these, but also epistemic modalities ('It is known that φ') and deontic ones ('It is obligatory that φ'). He then introduced the idea of *referential multiplicity*. In answer to Quine's question whether a certain occurrence of a singular term in a modal context is purely referential, and thus open to substitution and existential generalisation, or non-referential, in which case substitution and existential generalisation would fail according to Quine, Hintikka [39] pointed to a third possibility. According to the classical Fregean approach [32] singular terms would in non-extensional contexts not have their standard reference but instead refer to intensional entities, their ordinary senses. Hintikka saw no need to postulate special intensional entities for the singular terms to refer to in non-extensional contexts. The failure of substitutivity was instead explained by the referential multiplicity of the singular terms and by the fact that in intensional contexts the reference of the terms in various alternative courses of events ("possible worlds") is considered simultaneously.

Informally Hintikka [39] expressed the basic ideas behind the possible worlds interpretation of modal logic in the following words:

> ...we often find it extremely useful to try to chart the different courses the events may take even if we don't know which one of the different charts we are ultimately going to make use of. ...This analogy is worth elaborating. The concern of a general staff is not limited to what there will actually be. Its business is not just to predict the course of a planned campaign, but rather to be prepared for all the contingencies that may crop up during it. ...Most of the maps prepared by the general staff represent situations that will never take place. ...There are for the most parts some actual units for which the marks on the map stand, and the mutual positions of the units are such that the situation could conceivably arise. ...But the location of the units on the maps may be different from the locations the units have or ever will have. Some of the marks may stand for units which have not yet been formed; other maps may be prepared for situations in which some of the existing units have been destroyed. All these features have their analogues in modal logic.

In this example Hintikka informally speaks of the same units as occurring in different situations ("cross-world identification of individuals") and of individuals coming into existence or disappearing as one goes from one situation to another ("varying domains").

Hintikka goes on to explain the bearing of the above example on referential opacity.

> We may perhaps say that when we are doing modal logic, we are doing more than one thing at one and the same time. We use certain symbols — constants and variables — to refer to the actually existing objects of our domain of discourse. But we are also using them to refer to the elements of certain other states of affairs that need not be realized. Or, which amounts to the same, we are employing these symbols to build up 'maps' or models for the purpose of sketching certain situations that will perhaps never take place. If we could confine our attention to one of these possible states of affairs at a time, the occurrences of our symbols would be purely referential. The

> interconnections between the different models interfere with this. But since the symbols are purely referential within each particular model, the deviation from pure referentiality is not strong enough to destroy the possibility of employing quantifiers with pretty much the same rules as in the ordinary quantification theory. If I had to characterize the situation briefly, I should say that the occurrences of our terms in modal contexts are not usually *purely referential*, but rather that they are *multiply referential*.

This idea of referential multiplicity is perhaps the basic intuitive idea behind the possible worlds interpretation of modal notions and of indexical semantics in general. It seems that Hintikka here gives one of the earliest, or perhaps the earliest, clear expression of the idea.

Hintikka's semantics for quantified modal logic is informally interpreted in such a way that the quantifiers range over genuine individuals. Thus, Hintikka has a notion of cross-world identification: one and the same individual may occur in different worlds. However, the semantics allows individuals to *split* from one world to another, i.e., the individuals a and b may be identical in one world w_0 but they may fail to be identical in some alternative world to w_0. Thus, the principle:

$$(\Box =) \quad \forall x \forall y(x = y \rightarrow \Box(x = y)), \quad (\textit{Necessity of Identity})$$

is not valid in Hintikka's semantics. As a consequence, the unrestricted principle of indiscernibility of identicals does not hold in modal contexts according to Hintikka (cf., Hintikka [41] and later writings).

Hintikka's solution to Quine's paradox of identity. There are two cases to consider:

(1) One or the other of the singular terms under consideration ('Hesperus' or 'Phosphorus') is not a "rigid designator", that is it does not designate the same individual in every possible world (or "scenario") under consideration. Then, existential generalisation fails and Quine's paradoxical argument does not go through.

(2) Each of the two names picks out "the same" individual in every world under consideration. However, some scenario w under consideration is such that the individual Hesperus in w is distinct from the individual Phosphorus in w. In this case, Quine's argument goes through, but Hintikka has to argue that the conclusion:

    (6) $\exists x(x = \text{Hesperus} \wedge \neg\Box(x = \text{Hesperus}))$

    (7) $\exists x(x = \text{Hesperus} \wedge \Box(x = \text{Hesperus}))$,

    contrary to appearance, is not absurd, since an individual can "split" when we go from one possible scenario to one of its alternatives. Consider for example:

    > Superman and Clark Kent are in fact identical, but Lois Lane doesn't believe that they are identical.

Hintikka may explain the apparent truth (according to the story) of this sentence by the fact that some scenarios (possible worlds) in which Superman and Clark Kent are different individuals are among Lois Lane's doxastic alternatives in the actual world (where they are identical).

1.4.3 Montague's early semantics for quantified modal logic

A semantic approach to first-order modal predicate logic that has a certain resemblance to Kanger's was developed by Montague [83].[11] Like Kanger, Montague starts out from the standard model-theoretic semantics for non-modal first-order languages and extends it to languages with modal operators. He defines an *interpretation* for an ordinary first-order predicate language $\mathcal{L}$ to be a triple $\mathcal{I} = \langle D, I, g\rangle$, where (i) D is a non-empty set (the *domain*); (ii) I is a function that assigns appropriate denotations in D to the non-logical constants (predicate symbols and individual constants) of $\mathcal{L}$; and (iii) a function g (an *assignment* in D) that assigns values in D to the individual variables of $\mathcal{L}$. For each non-logical constant or variable X, let $\mathcal{I}(X)$ be the *semantic value* (i.e., *denotation* for non-logical constants and *value* for variables) of X in the interpretation $\mathcal{I}$. Then the notion of *truth* relative $\mathcal{I}$ is defined as follows:

(1) $\mathcal{I} \vDash P(t_1, \ldots, t_n)$ iff $\langle \mathcal{I}(t_1), \ldots, \mathcal{I}(t_n)\rangle \in \mathcal{I}(P)$,

(2) $\mathcal{I} \vDash (t_1 = t_2)$ iff $I(t_1) = I(t_2)$,

(3) $\mathcal{I} \vDash \neg\varphi$ iff $\mathcal{I} \nvDash \varphi$,

(4) $\mathcal{I} \vDash (\varphi \rightarrow \psi)$ iff $\mathcal{I} \nvDash \varphi$ or $\mathcal{I} \vDash \psi$,

(5) $\mathcal{I} \vDash \forall x\varphi$ iff for every object $a \in D, \mathcal{I}(a/x) \vDash \varphi$.

Here, $\mathcal{I}(a/x)$ is the interpretation that is exactly like $\mathcal{I}$, except for assigning the object a to the variable x as its value.

Montague now asks the same question as Kanger: How can this definition of the truth-relation be generalised to first-order languages with modal operators? As we recall, Kanger solved the problem by modifying the notion of an interpretation: a Kanger-type interpretation (what he called 'a system') assigns denotations to the non-logical constants and values to the variables not only for one single domain (the 'actual' one) but for all domains in one fell swoop. Montague's approach is simpler than Kanger's: he keeps the notion of an interpretation $\mathcal{I}$ of first-order logic intact, and just adds semantic evaluation clauses for the modal operators. As in the Kanger semantics, each modal operator $\Box$ is associated with an accessibility relation $R_\Box$. Now, however accessibility relations are relations between interpretations $\mathcal{I} = \langle D, I, g\rangle$ of the underlying non-modal first-order language. The semantic clause corresponding to the operator $\Box$, with associated accessibility relation $R_\Box$, is:

(6) $\mathcal{I} \vDash \Box\varphi$ iff for every interpretation $\mathcal{I}'$ such that $\mathcal{I}R_\Box\mathcal{I}', \mathcal{I}' \vDash \varphi$.

Montague associates with the operator **L** of *logical necessity* the accessibility relation R_L defined by:

$$\langle D, I, g\rangle R_L \langle D', I', g'\rangle \text{ iff } D = D' \text{ and } g = g'.$$

Thus, his semantic clause for **L** becomes:

[11] Montague [83] writes: "The present paper was delivered before the Annual Spring Conference in Philosophy at the University of California, Los Angeles, in May, 1955. It contains no results of any great technical interest; I therefore did not initially plan to publish it. But some closely analogous, though not identical, ideas have recently been announced by Kanger [54, 55] and by Kripke in [64]. In view of this fact, together with the possibility of stimulating further research, it now seems not wholly inappropriate to publish my early contribution."

(7) $\langle D, I, g\rangle \models \mathbf{L}\varphi$ iff for every I' defined over D, $\langle D, I', g\rangle \models \varphi$.

That is, $\mathbf{L}\varphi$ is true in an interpretation $\mathcal{I}$ iff φ is true in every interpretation $\mathcal{I}'$ that is like $\mathcal{I}$ except for, possibly, assigning different semantic values to the non-logical constants of $\mathcal{L}$.

Stated in contemporary terms, Montague's semantic clause for the logical necessity operator becomes:

(8) $\mathbf{L}\varphi$ is true in a model $\mathcal{M} = \langle D, I\rangle$ relative to an assignment g iff for every model $\mathcal{M}'$ with domain D, φ is true in $\mathcal{M}'$ relative to g.

Let us say that a formula φ of $\mathcal{L}$ is *D-valid relative to g* iff for every model $\mathcal{M}$ with domain D, φ is true in $\mathcal{M}$ relative to g. We say that φ is *D-valid* iff it is D-valid relative to every assignment g in D. Then, from Montague's semantic clause for $\mathbf{L}$, we can conclude:

(9) $\mathbf{L}\varphi$ is true in $\mathcal{M} = \langle D, I\rangle$ relative to g iff φ is D-valid relative to g.

and

(10) $\mathbf{L}\varphi$ is true in $\mathcal{M} = \langle D, I\rangle$ iff φ is D-valid.

We say that a formula φ of $\mathcal{L}$ is *logically true* iff it is D-valid in every non-empty domain D.

Montague's [83] semantics for $\mathbf{L}$ is exactly what Cocchiarella [23] refers to as the "primary semantics" for logical necessity. Hence, we can reformulate Cocchiarella's [23] *incompleteness theorem* for that semantics as follows:

THEOREM 2. *Suppose that $\mathcal{L}$ contains at least one binary predicate symbol. Then, the set of logically true sentences in Montague's [83] semantics for logical necessity is not recursively enumerable. Thus, Montague's [83] logic for logical necessity is not axiomatizable.*

Montague's solution to Quine's paradox of identity. According to Montague's interpretation, $\mathbf{L}\varphi$ is logically equivalent with a *formula of second-order predicate logic* $()\varphi$, where $()$ stands for a string of universal quantifiers that bind all non-logical symbols in φ. In other words, Montague's semantics induces a translation from first-order modal logic to extensional second-order predicate logic. According to Montague's semantics from [83], the quantifier $\forall x$ is interpreted as a genuine quantifier over individuals. Free variables are "directly referential", i.e., a free variable is interpreted uniformly inside a formula as standing for one and the same individual regardless of where in the formula it occurs. Individual constants, on the other hand, are reinterpreted freely from one interpretation to another.

Montague's semantics validates the following principles without restrictions:

(LI) $\forall x(x = x)$, *(Law of Identity)*
(I=) $\forall x\forall y(x = y \to (\varphi(x/z) \to \varphi(y/z)))$. *(Indiscernibility of Identicals)*

In addition, we have: $\forall x\mathbf{L}(x = x)$. Therefore, the following principle is valid:

($\Box$I) $\forall x\forall y(x = y \to \mathbf{L}(x = y))$. *(Necessity of Identity)*

But the following is not valid:

$$\text{Phosphorus} = \text{Hesperus} \to \mathbf{L}(\text{Phosphorus} = \text{Hesperus}).$$

It follows that the principles of *Universal Specification* (US) and *Existential Generalisation* (EG) are not valid. Thus, Quine's paradoxical argument (Section 1.3, (1)–(7)) cannot be carried through within Montague's logic. Although (US) and (EG) cannot be applied to individual constants, they do hold for variables.

It appears that Montague's semantical interpretation satisfies all requirements imposed by Quine [92] on an interpretation of quantified modal logic for the logical modalities. However, Montague's semantics still has counterintuitive consequences. Consider, for instance, the following proof of the thesis that *everything there is exists necessarily*:

(1) $\forall x \exists y (x = y)$ predicate logic

(2) $\mathbf{L}\forall x \exists y (x = y)$ from (1) by necessitation

(3) $\forall x \exists y (x = y) \to \exists y (x = y)$ universal specification (US) (for variables)

(4) $\mathbf{L}(\forall x \exists y (x = y) \to \exists y (x = y))$ from (3) by necessitation

(5) $\mathbf{L}\exists y (x = y)$ from (2) and (4) by modal logic

(6) $\forall x \mathbf{L}\exists y (x = y)$ from (5) by universal generalization (UG)

This proof is valid according to Montague's semantics: line (1) is logically true and the steps in the proof preserve logical truth. It is also easy to see directly that the conclusion (6) of the argument is logically true according to Montague's definition. This conclusion, however, is extremely counterintuitive (provided we read the quantifiers in the normal way as ranging over ordinary objects). Intuitively, it is simply false that everything there is exists necessarily. Hence, there are still problems with Montague's semantics. We shall return to the above problematic argument in connection with Kripke's [66] possible worlds semantics.

It should also be noted that Montague's semantics validates the schema:

(I) $\exists x \mathbf{L}\varphi(x) \leftrightarrow \forall x \mathbf{L}\varphi(x)$.

i.e., φ holds necessarily of one thing just in case φ holds necessarily of everything. Moreover, the semantics validates the Barcan schema and its converse:

(BF) $\forall x \mathbf{L}\varphi(x) \to \mathbf{L}\forall x \varphi(x)$

(CBF) $\mathbf{L}\forall x \varphi(x) \to \forall x \mathbf{L}\varphi(x)$.

From (1), (BF) and (CBF) we infer:

(II) $\exists x \mathbf{L}\varphi(x) \leftrightarrow \mathbf{L}\forall x \varphi(x)$.

That is, a property holds necessarily of one thing just in case it is necessary that it holds of everything.

According to Montague's semantics the logically necessary properties are the same for everything; namely, just those properties that by logical necessity hold of everything. That is, Montague's semantics is *essentialist* in the weak Quinean sense of distinguishing between properties that hold necessarily of a thing and properties that hold only contingently of it. But it rejects the *strong essentialist thesis* that there are properties that some objects have necessarily and others do not have at all, or have only contingently (cf. [8, 89]).[12] Hence, condition (I) seems to be correct, as long as we speak of logical necessity. Logic does not discriminate between individuals, so if F is a logically necessary property of one thing, it is a logically necessary property of everything there is.[13]

The Barcan formula and its converse, however, are dubious. Consider first (BF). Suppose that a is the only thing that exists. Then, $\forall x\mathbf{L}(x = a)$. However, it does not seem intuitively correct to infer: $\mathbf{L}\forall x(x = a)$. Next, consider (CBF). Clearly, $\mathbf{L}\forall x\exists y(x = y)$. If (CBF) were valid, we could infer $\forall x\mathbf{L}\exists y(x = y)$, which — as we have already pointed out — is counterintuitive. We will return to the semantic significance of (BF) and (CBF) in Section 1.4.4. Finally, condition (II) is clearly counterintuitive. Burgess [14] says of (II) that it "could silence any critic who claimed the notion of *de re* modality to be more obscure than that of *de dicto* modality, but would do so only at the cost of making *de re* notation pointless".

1.4.4 Kripke's semantics for quantified modal logic

Kripke 1959. The possible worlds semantics introduced by Kripke [64] may be cast in the following form (which differs from Kripke's original formulation in terminology as well as in some minor details). We consider a language $\mathcal{L}$ of modal predicate logic with identity containing for each $n \geq 1$, a denumerably infinite list of n-ary predicate symbols, but no function symbols or individual constants. Let D be a non-empty set. We define a *valuation* for $\mathcal{L}$ over D to be a function V which to every n-ary predicate symbol $P(n \geq 1)$ in $\mathcal{L}$ assigns a value $V(P) \subseteq D^n$. An *assignment* in D is a function g which to every individual variable x assigns a value $g(x) \in D$. A model over D is an ordered pair $\mathcal{M} = \langle \mathbf{K}, V_0\rangle$ such that (i) $\mathbf{K}$ is a set of valuations for $\mathcal{L}$ over D, and (ii) $V_0 \in \mathbf{K}$.

Given a model $\mathcal{M} = \langle \mathbf{K}, V_0\rangle$ over D, an evaluation V in $\mathbf{K}$, assignment g in D, and formula φ we define recursively what it means for φ to be *true in V relative to $\mathcal{M}$ and g* (in symbols: $V \vDash_M \varphi[g]$):

(1) $V \vDash_{\mathcal{M}} P(x_1, \ldots, x_n)[g]$ iff $\langle g(x_1), \ldots, g(x_n)\rangle \in V(P)$,

(2) $V \vDash_{\mathcal{M}} (x = y)[g]$ iff $g(x) = g(y)$,

(3) $V \vDash_{\mathcal{M}} \neg\varphi[g]$ iff $V \nvDash_{\mathcal{M}} \varphi[g]$,

(4) $V \vDash_{\mathcal{M}} (\varphi \rightarrow \psi)[g]$ iff $V \nvDash_{\mathcal{M}} \varphi[g]$ or $V \vDash_{\mathcal{M}} \psi[g]$,

(5) $V \vDash_{\mathcal{M}} \forall x\varphi[g]$ iff for every object $a \in D, V \vDash_{\mathcal{M}} \varphi[g(a/x)]$,

(6) $V \vDash_M \Box\varphi$ iff for every valuation V' in $\mathbf{K}$, $V' \vDash_{\mathcal{M}} \varphi$.

As usual, $g(a/x)$ is the assignment that is exactly like g except for assigning a to the variable x.

[12] See also Kaplan's [61] penetrating analysis of the distinction between logical and metaphysical necessity. According to Kaplan, logical necessity is committed to a *benign* form of Aristotelian essentialism that "makes a specification of an individual essential only if it is logically true of that individual". Metaphysical necessity, on the other hand, is *invidious*, since it allows for distinct individuals to have different essential properties.

[13] On the other hand, (I) is clearly counterintuitive for metaphysical necessity. Let, for example, $\varphi(x)$ be the formula '$(\exists y(y = x) \rightarrow x \in \{\text{Socrates}\})$' and let $\Box$ stand for metaphysical necessity. Then, $\Box\varphi(\text{Socrates})$ is true. Socrates is a member of {Socrates}, in every possible world where Socrates exists. But, of course, $\Box\varphi(\text{Plato})$ is false. Thus (I) fails for metaphysical necessity.

We say that φ is *true in $\mathcal{M}$ relative to g* if $V_0 \models_{\mathcal{M}} \varphi[g]$. φ is *true in $\mathcal{M}$* if $V_0 \models_M \varphi[g]$ for every assignment g in D. φ is *valid in the domain D* if φ is true in all models $\mathcal{M}$ over D. φ is *universally valid* if φ is valid in every non-empty domain D (i.e., just in case φ is true in every model $\mathcal{M}$).

Kripke gives the following intuitive motivation for this semantics: The valuations in **K** are thought of as representing the set of all "possible" (or "conceivable" or "imaginable") worlds. The valuation V_0 represents the "real" world. It is assumed that the set D of individuals is the same for all possible worlds. Necessity is defined as truth in all possible worlds.

Kripke's [64] semantics validates all the classically valid schemata of first-order predicate logic with identity, the characteristic axioms of **S5**, as well as the Barcan formula (BF) and its converse (CBF). The set of valid sentences is closed under modus ponens, uniform substitution, necessitation, and universal generalization. In [64], Kripke defines a formal system $\mathbf{S5}^{*=}$ for quantified modal logic and proves using semantic tableaux methods that it is sound and complete for the given semantics.

Let us now compare Kripke's [64] semantics with Montague's semantics [83] for logical necessity. Let us say that a Kripke [64] model $\mathcal{M} = \langle \mathbf{K}, V_0 \rangle$ over a non-empty domain D is *maximal* if **K** contains all valuations for $\mathcal{L}$ over D.[14]

Montague's semantics for logical necessity differs from Kripke's [64] semantics in considering maximal models only. We obtain Montague's semantics for logical necessity by imposing the requirement on Kripke's [64] models that the set **K** should contain all valuations V for $\mathcal{L}$ over D. Hence, a sentence φ of $\mathcal{L}$ is logically true in Montague's [83] semantics for logical necessity iff it is true in all maximal Kripke [64] models. By restricting our attention to maximal models, we get what Cocchiarella [23] calls the "primary semantics" for logical necessity.

At this point it is natural to ask what intended interpretation Kripke had in mind for the necessity operator in 1959. Was it logical necessity, analytical necessity, or perhaps some kind of metaphysical necessity? One reason for thinking that Kripke's notion of necessity in 1959 was not logical necessity is his use of models that are non-maximal (or "non-standard" in the terminology of Hintikka [46]). Instead of working with all models or valuations over D, like Montague, or with all possible systems as Kanger, Kripke is considering an arbitrary non-empty subset of all possible valuations. This feature of his models may suggest that Kripke's intended interpretation of the necessity operator is not strict logical necessity, but perhaps instead some kind of metaphysical necessity. This conclusion is however, not unavoidable: Kripke's intended interpretation of the necessity operator could still have been logical necessity and his *intended interpretations* could still be some or all of the *maximal models*. Kripke's reason for allowing non-maximal models, in addition to maximal ones, when defining validity, could have been logical rather then philosophical.[15] If Kripke, like Kanger and Montague, had chosen to work only with maximal models, the set of valid sentences would not have been recursively enumerable and there would be no completeness theorem to be proved. Kripke's intended model could, for instance, be a maximal model over some infinite set. A modal sentence of an interpreted language of modal predicate logic would then be *true* if it was true in the

[14] The term "maximal model" was introduced by Parsons [89] in connection with Kripke's [66] semantics for quantified logic. It is less tendentious than Hintikka's term "standard model".

[15] Ballarin [4] argues that Kripke's development of his possible worlds semantics was driven entirely "by formal considerations, not interpretive concerns".

intended model. Interpreted in this way, Kripke's 1959 approach would be very close to Montague's of 1960. The only essential difference would be Kripke's use of non-standard models in addition to the standard ones for the purpose of defining a notion of universal validity that is recursively enumerable.

On the other hand, in [64, p. 3], Kripke speaks of **K** as representing the set of all "conceivable" worlds. He writes "... a proposition $\Box B$ is evaluated as true when and only when B holds in all conceivable worlds". This seems to indicate that Kripke's operator $\Box$ of [1959] should not be interpreted as strict logical necessity. It is very likely that the set of valuations representing all "conceivable" worlds is a proper subset of the set of absolutely all valuations. Thus Kripke may have had philosophical reasons, in addition to formal ones, for favouring a "non-standard" semantics allowing non-maximal models to a "standard" one.[16]

Kripke 1963. We present a version of Kripke's [66] semantics for modal predicate logic with identity, where the notion of a possible world is an explicit ingredient of the semantic theory. We differ from Kripke [66] in letting the language $\mathcal{L}$ contain individual constants.

A (*Kripke*) *frame* (or to use Kripke's own terminology, a *model structure*) for a language $\mathcal{L}$ of first-order modal predicate logic (with identity and individual constants, but no function symbols) is a quintuple $\mathcal{F} = \langle W, D, R, E, w_0 \rangle$ where, (i) W is a non-empty set; (ii) D is a non-empty set; (iii) $R \subseteq W \times W$; (iv) E is a function which to each $w \in W$ assigns a subset E_w of D; and (v) w_0 is a designated element of W. Intuitively we think of matters thus: W is the set of all (*possible*) *worlds* (possible states of affairs, possible ways the world could have been), D is the set of all (*possible*) *individuals*, R is the *accessibility relation* between worlds, for each world w, E_w is the set of *individuals that exist in* w; and w_0 is the *actual world*. It is required that $D = \bigcup_{w \in W} E_w$, i. e., that every possible individual exists in at least one world.

Next, let us say that I is an *interpretation* (in D with respect to W) if it is a family of functions I_w, where w ranges over W, such that I_w assigns a subset $I_w(P)$ of D^n to each n-ary predicate constant P of $\mathcal{L}$ and an element $I_w(c) \in D$ to each individual constant c of $\mathcal{L}$. A *Kripke model* (for $\mathcal{L}$) is an ordered pair $\mathcal{M} = \langle \mathcal{F}, I \rangle$, where $\mathcal{F} = \langle W, D, R, E, w_0 \rangle$ is a frame and I is an interpretation in D with respect to W. A model $\mathcal{M}$ of the form $\langle \mathcal{F}, I \rangle$ is said to be *based on* the frame $\mathcal{F}$.

Observe that $I_w(P)$ is not necessarily a subset of $(E_w)^n$, i. e., the extension of P in w may contain individuals that do not exist in w. Nor do we require that $I_w(c) \in E_w$. An *assignment* in $\mathcal{M}$ is a function g which assigns to each variable x an element $g(x)$ in D. For any term t in $\mathcal{L}$, we define $\mathcal{M}_w(t, g)$ to be $g(t)$ if t is a variable; and $I_w(t)$ if t is an individual constant. We speak of $\mathcal{M}_w(t, g)$ as the *denotation of the term* t *at the world* w *relative to the model* $\mathcal{M}$ *and the assignment* g.

With these notions in place, we can define what it means for a formula φ to be *true at a world* w *with respect to the model* $\mathcal{M}$ *and the assignment* g (in symbols, $w \vDash_{\mathcal{M}} \varphi[g]$):

(1) $w \vDash_{\mathcal{M}} P(t_1, \ldots, t_n)[g]$ iff $\langle \mathcal{M}_w(t_1, g), \ldots, \mathcal{M}_w(t_n, g) \rangle \in I_w(P)$.

(2) $w \vDash_{\mathcal{M}} (t_1 = t_2)[g]$ iff $\mathcal{M}_w(t_1, g) = \mathcal{M}_w(t_2, g)$.

(3) $w \vDash_{\mathcal{M}} \neg\varphi[g]$ iff $w \nvDash_{\mathcal{M}} \varphi[g]$.

[16] Cf., however, Almog [1, p. 217], who writes about Kripke [64]: "... Kripke had at the time nothing more than "complete assignments," and the modality he worked with was definitely *logical* possibility".

(4) $w \models_{\mathcal{M}} (\varphi \rightarrow \psi)[g]$ iff $w \not\models_{\mathcal{M}} \varphi[g]$ or $w \models_{\mathcal{M}} \psi[g]$.

(5) $w \models_{\mathcal{M}} \forall x\varphi[g]$ iff, for every $a \in E_w$, $w \models_{\mathcal{M}} \varphi[g(a/x)]$.

(6) $w \models_{\mathcal{M}} \Box\varphi[g]$ iff, for every $u \in W$ such that $wRu, u \models_{\mathcal{M}} \varphi[g]$.

We say that φ is *true with respect to the model* $\mathcal{M}$ *and the assignment* g (in symbols $\models_{\mathcal{M}} \varphi[g]$), iff φ is true at the actual world w_0 with respect to $\mathcal{M}$ and g. φ is *true in the model* $\mathcal{M}$ (in symbols, $\models_{\mathcal{M}} \varphi$), if for every assignment $g, \models_{\mathcal{M}} \varphi[g]$. φ is *true in a frame* $\mathcal{F}$ (in symbols, $\models_{\mathcal{F}} \varphi$) if φ is true in every model based on $\mathcal{F}$. Let $\mathbf{K}$ be a class of frames. We say that φ is $\mathbf{K}$*-valid* if φ is true in every $\mathcal{F} \in \mathbf{K}$.

Observe that there are two notions of validity that are naturally defined on classes of Kripke frames. With respect to the notion that we have just defined — we may call it *real-world validity* — the actual world plays a special role: a sentence φ is real-world valid in a class $\mathbf{K}$ of frames if it is true at the actual world in every frame in $\mathbf{K}$. Then, there is another notion of validity that we may call *general validity*: A sentence φ is generally valid in a class $\mathbf{K}$ just in case it is true at each world w in each frame in $\mathbf{K}$.[17] In the definition of general validity, the designated point of a Kripke model does not play any role. Thus, if we are only interested in general validity, there is no need to provide Kripke frames with designated worlds. Let us write $\models_{\mathbf{K}}$ and $\models^*_{\mathbf{K}}$ for real-world validity in $\mathbf{K}$ and general validity in $\mathbf{K}$, respectively. Then we have, for any sentence φ of $\mathcal{L}$

(1) $\models^*_{\mathbf{K}} \varphi$ iff $\models_{\mathbf{K}} \Box\varphi$

Let us say that a class $\mathbf{K}$ of Kripke frames is *normal* iff it satisfies the condition:

> Whenever $\mathcal{F}$ is in $\mathbf{K}$ and $\mathcal{F}'$ is a frame that differs from $\mathcal{F}$ only with respect to which world is the actual one, then $\mathcal{F}'$ is also in $\mathbf{K}$.

For normal classes of frames, real-world validity coincides with the general validity. Thus, for any sentence φ of $\mathcal{L}$,

(1) if $\mathbf{K}$ is normal, then $\models_{\mathbf{K}} \varphi$ iff $\models^*_{\mathbf{K}} \varphi$

The semantic import of the Barcan formula and its converse. Notice that Kripke frames in general have *varying domains*, i.e., the domains of quantification E_w are allowed to vary from one possible world to another. We say that a frame $\mathcal{F} = \langle W, D, R, E, w_0\rangle$ has *increasing domains* iff for all $u, v \in W$, if uRv, then $E_u \subseteq E_v$. $\mathcal{F}$ has *decreasing domains* iff for all $u, v \in W$, if uRv, then $E_v \subseteq E_u$. $\mathcal{F}$ has *locally constant domains* iff for all $u, v \in W$, if uRv, then $E_u = E_v$. $\mathcal{F}$ has *globally constant domains* iff for all $u \in W, E_u = D$. We also say that F is a *constant domain frame* iff $\mathcal{F}$ has globally constant domains.

Consider now the following conditions on frames $\mathcal{F}$:

(ID) $\mathcal{F}$ has increasing domains.

(DD) $\mathcal{F}$ has decreasing domains.

(LCD) $\mathcal{F}$ has locally constant domains.

[17] Cf. [51, 22–24], for a comparison between the two concepts of logical truth (validity) and for the history of the distinction between the two.

(CBF) Every instance of the converse Barcan formula: $\Box\forall x\varphi(x) \to \forall x\Box\varphi(x)$, is generally valid in every model based on $\mathcal{F}$.

(BF) Every instance of the Barcan formula: $\forall x\Box\varphi(x) \to \Box\forall x\varphi(x)$, is generally valid in every model based on $\mathcal{F}$.

(CBF + BF) Every instance of the Barcan formula and its converse is generally valid in every model based on $\mathcal{F}$.

There is an exact correspondence between the conditions (ID), (DD), (LCD) and (CBF), (BF) and (CBF + BF), respectively (cf. [30]). That is:

(i) $\mathcal{F}$ has increasing domains iff it satisfies (CBF).

(ii) $\mathcal{F}$ has decreasing domains iff it satisfies (BF).

(iii) $\mathcal{F}$ has locally constant domains iff it satisfies (CBF + BF).

Moreover,

(iv) A sentence is generally valid in the class of all constant domain frames iff it is generally valid in all locally constant domain frames.

We may introduce an *existence predicate* **E** as a new logical constant and give it the semantic clause:

$$w \vDash_{\mathcal{M}} \mathbf{E}(t)[g] \text{ iff } \mathcal{M}_w(t,g) \in E_w.$$

However, this is unnecessary as long as we have identity in the language, since the predicate **E** is definable in terms of the existential quantifier and identity:

$$w \vDash_{\mathcal{M}} \mathbf{E}(t)[g] \text{ iff } w \vDash_{\mathcal{M}} \exists y(y = t)[g], \text{ where } y \text{ is a variable that is distinct from } t.$$

Hence, we may take $\mathbf{E}(t)$ as an abbreviation of $\exists y(y = t)$.

In terms of **E** we can express the requirements of increasing and decreasing domains in a simple way:

(v) $\mathcal{F}$ has increasing domains iff the sentence $\Box\forall x\Box\mathbf{E}(x)$ is valid in $\mathcal{F}$.

(vi) $\mathcal{F}$ has decreasing domains iff the formula $\Box(\Diamond\mathbf{E}(x) \to \mathbf{E}(x))$ is valid in $\mathcal{F}$.

We are especially interested in frames where R is the *universal relation* in W, i.e., in which:

$$w \vDash_{\mathcal{M}} \Box\varphi[g] \text{ iff, for every } u \in W, u \vDash_{\mathcal{M}} \varphi[g].$$

Let **QS5=** be the class of all such frames. It follows from what we have stated above, that neither the Barcan formula nor its converse is (**QS5**=)-valid.

In order to illustrate the difference between Kripke's [66] semantics and his earlier semantics from 1959, consider again the purported proof that *everything there is exists necessarily* (Section 1.4.3). The proof is valid in the semantics of Montague [83] as well as in Kripke [64]. However, according to Kripke [66], the argument fails. It is easy to see that the conclusion is not valid according to Kripke [66]. When we look at the purported proof, we see that it is line (3) that fails:

(3) $\forall x \exists y(x = y) \rightarrow \exists y(x = y)$ universal specification (US) (for variables)

That is, (US) is not valid according to Kripke [66] (not even for variables): The universal quantifier in the antecedent of (3) ranges over the domain of actually existing objects, while the free variable x in the succedent may take possible objects as values that lie outside the domain of actually existing objects. The failure of this intuitively invalid argument in Kripke's [66] semantics speaks in favour of this semantics in comparison with Montague [83] and Kripke [64].

Rigid designators. Kripke's [66] semantics validates the *Law of Identity*,

(L=) $\forall x(x = x)$,

as well as the principle of *Indiscernibility of Identicals*,

(I=) $\forall x \forall y[x = y \rightarrow (\varphi(x/z) \rightarrow \varphi(y/z))]$,

applicable without restrictions also to modal contexts $\varphi(z)$. From these principles, together with the rule of Necessitation it is easy to infer:

($\Box =$) $\forall x \forall y(x = y \rightarrow \Box(x = y))$ *(Necessity of Identity)*

($\Box \neq$) $\forall x \forall y(x \neq y \rightarrow \Box(x \neq y))$. *(Necessity of Distinctness)*

However, neither

(1) $c = d \rightarrow \Box(c = d)$

nor

(2) $c \neq d \rightarrow \Box(c \neq d)$,

is valid, for arbitrary individual constants c, d. This reflects an important difference between how individual variables and individual constants are treated in our modelling: in spite of their name, the denotation of individual constants may vary from one possible world to another, whereas the denotation of variables — in spite of their name — remains fixed throughout the universe of possible worlds. Here is obviously a niche to be filled! Suppose we introduce a new syntactic category of *names* and require that the interpretation of a name $\mathbf{n}$ be constant over the set of all possible worlds in any model $\mathcal{M}$; formally,

$$I_u(\mathbf{n}) = I_v(\mathbf{n}),$$

for all $u, v \in W$. Then, if $\mathbf{n}$ and $\mathbf{m}$ are any names, then:

(3) $\mathbf{n} = \mathbf{m} \rightarrow \Box(\mathbf{n} = \mathbf{m})$

(4) $\mathbf{n} \neq \mathbf{m} \rightarrow \Box(\mathbf{n} \neq \mathbf{m})$.

are both valid. The proposed modification amounts to treating the elements of the new category of names as what is now known, after Kripke [71], as *rigid* designators. In [71] Kripke made the claim that ordinary "proper names" in natural language are rigid designators.

Maximal models and maximal validity. Next, we introduce a special kind of Kripke models that we refer to as *maximal models*. We say that an ordered triple $\langle D, A, V\rangle$ is a *first-order model* for $\mathcal{L}$ with *outer domain* D and *inner domain* A iff (i) $D \neq \varnothing, A \subseteq D$; and (ii) for each n-ary predicate constant $P, V(P) \subseteq D^n$; (iii) for each individual constant $c, V(c) \in D$.

A Kripke model $\mathcal{M} = \langle W, D, R, E, w_0, I\rangle$ is *maximal* if (i) $R = W \times W$; (ii) for every subset A of D and every first-order model $\langle D, A, V\rangle$ with outer domain D and inner domain A, there exists a $w \in W$ such that $E_w = A$ and $I_w = V$; and (iii) if $u, v \in W$ and $E_u = E_v$ and $I_u = I_v$, then $u = v$. Thus, in a maximal Kripke model with individual domain D, the possible worlds can be identified with all first-order models with outer domain D. Thus, for each non-empty set D, there is a unique maximal Kripke model with individual domain D.

The notion a maximal Kripke model is due to Terence Parsons [89]. Montague's [83] models correspond to the maximal Kripke models with a constant domain, i.e. where each $E_w = D$. If $\mathcal{M}$ is the maximal Kripke model with domain D, then for every formula φ of $\mathcal{L}$:

> $\Box\varphi$ is true at a world w in $\mathcal{M}$ relative an assignment g iff φ is true in every first-order model with outer domain D relative to g.

Thus, it is natural to interpret $\Box$ as a kind of logical (or combinatorial) necessity with respect to maximal Kripke models: $\Box\varphi$ is true in a maximal model with domain D iff φ is true in every first-order model with outer domain D.

Let us say that a formula φ is *maximally valid* iff for every maximal Kripke model $\mathcal{M}$ and every assignment g in $\mathcal{M}, \models_M \varphi[g]$. Observe that the set of maximally valid sentences is not closed under uniform substitution of arbitrary sentences for atomic sentences: for an atomic formula $Pc, \Diamond Pc$ is maximally valid, but, of course, $\Diamond\varphi$ is not in general maximally valid. Moreover, if φ is a formula that does not contain $\Box$ or $\Diamond$ which is not a theorem of first-order logic, then $\neg\Box\varphi$ is maximally valid. Of course, neither the Barcan schema nor its converse is maximally valid.

Suppose now that the intended model of $\mathcal{L}$ is some maximal Kripke model $\mathcal{M}_0$ with an infinite domain D_0. Then, all sentences of the form:

$$(n)\ \ \Diamond\exists x_1 \ldots \exists x_n(x_1 \neq x_2 \wedge \ldots \wedge x_1 \neq x_n \wedge x_2 \neq x_3 \wedge \ldots \wedge x_2 \neq x_n \wedge \ldots \wedge x_{n-1} \neq x_n),$$

where $x_1, \ldots, x_n$ are $n(n > 1)$ distinct variables, are *true* in (the intended model for) $\mathcal{L}$. This appears to be as it should be, given the interpretation of $\Diamond$ as (a kind of) logical possibility. With this notion of truth in $\mathcal{L}$, we can associate various notions of *logical truth*. One alternative is to say that a sentence in $\mathcal{L}$ is logically true iff it is true in every maximal model with the given outer domain D. With this notion all the sentences (n) come out as logically true. Another alternative is to say that a sentence is logically true if it is maximally valid, i.e., true in all maximal Kripke models. Then the sentences (n) are no longer logically true. Finally, we may identify logical truth in $\mathcal{L}$ with truth in all **QS5=**-Kripke models. Of these choices, only the last one satisfies the standard requirement on a logic of being closed under uniform substitution. Thus, if we insist that a logic should be closed under uniform substitution, it is reasonable to identify logical truth in $\mathcal{L}$ with Kripke's notion of universal validity. Hence, regardless of whether the intended model is a maximal model or not, we may reasonably conclude that the logic of alethic necessity is the set of all **QS5**=-valid sentences. By this line of reasoning, we come to the conclusion that regardless of whether we interpret $\Box$ as standing for logical or metaphysical necessity, the *logic* of $\Box$ will be the same.[18]

Kripke versus Quine. In 1959 Kripke wrote:

> It is noteworthy that the theorems of this paper can be formalized in a metalanguage (such as Zermelo set theory) which is "extensional," both in the sense of possessing set-theoretic axioms of extensionality *and* in the sense of postulating no sentential connectives other than the truth-functions. Thus it is seen that at least a certain non-trivial portion of the semantics of modality is available to an extensionalist logician.

Perhaps, Kripke meant that he had refuted Quine's scepticism about quantified modal logic. Had he not after all done for quantified modal logic what Tarski and others had done for non-modal predicate logic: provided it with an extensional set-theoretic semantics? In addition he had axiomatised the logic and proved it complete for the given semantics. What else could one require of the interpretation of a logic?

Quine, however, was not satisfied. In 1972 he writes in a review of Kripke's paper 'Identity and Necessity' [96]:

> The notion of possible world did indeed contribute to the semantics of modal logic, and it behoves us to recognize the nature of its contribution: it led to Kripke's precocious and significant theory of models of modal logic. Models afford consistency proofs; also they have heuristic value; but they do not constitute explication. Models, however clear in themselves, may leave us still at a loss for the primary, intended interpretation.

Whatever was his aim in 1959 or 1963, in his later work Kripke's project is not to give an explanation of modal concepts in non-modal terms. In the Preface to *Naming and Necessity*, 1980 he writes:

> I do not think of 'possible worlds' as providing a *reductive* analysis in any philosophically significant sense, that is, as uncovering the ultimate nature, from either an epistemological or a metaphysical point of view, of modal operators, propositions, etc., or as 'explicating' them.

Clearly, Kripke's essentialist concept of necessity ("metaphysical necessity") simply cannot be reductively explained in non-modal terms.

Among other modellings for predicate modal logic, David Lewis's counterpart theory should be mentioned.[19] According to the Kripke paradigm, an individual may exist in more than one possible world (with respect to our formal modelling, it is possible that E_u and E_v should overlap, even if $u \neq v$). For Lewis, however, each individual inhabits only its own possible world; but it may have counterparts in other possible worlds. This approach has also been influential, both in philosophical and in mathematical quarters.

1.5 General intensional logic

1.5.1 Carnap-Montague's Intensional Logic

Frege's theory of *Sinn* (*sense*) and *Bedeutung* (*denotation*, *reference*), which was outlined in the article 'Über Sinn und Bedeutung' [32] has great intuitive appeal. In particular,

[18] Cf. [15].
[19] Cf. [75, 37].

it seems to provide elegant and intuitively appealing solutions to the familiar difficulties concerning:

(i) the cognitive significance of identity statements: how can '$a = b$' if true, be an informative statement differing in cognitive significance from '$a = a$'?

(ii) the problem of oblique or non-extensional contexts: how can two meaningful expressions with the same denotation (extension) ever fail to be interchangeable *salva veritate*?

(iii) the problem of providing an adequate truth-conditional semantics for propositional attitude reports.

Fregean solutions to these problems essentially involve the distinction between sense and denotation. The appearance of oblique contexts in natural languages was interpreted by Frege as indicating a certain kind of systematic ambiguity rather than a failure of extensionality. According to Frege's doctrine of indirect denotation, expressions denote in (unembedded) oblique contexts what is ordinarily their sense. Frege's extensional point of view has been advocated and developed in the 20th century by Alonzo Church [19, 20, 21] in his *Logic of Sense and Denotation*.[20]

Carnap [17], although still working within the Fregean tradition, saw the occurrence of oblique contexts in natural languages as genuine counterexamples to the *principle of extensionality*, according to which the denotation of a meaningful expression is always a function of the denotations of its semantically relevant parts.

According to Carnap [17], each well-formed expression of a language has both an *extension* (corresponding to Frege's denotation) and an *intension* (roughly corresponding to Frege's sense). Intuitively, the intension of a sentence is the proposition that the sentence expresses and the extension is the truth-value (true or false) of the sentence. A proposition partitions the set of all possible worlds in two cells: (i) the set of all worlds in which the proposition is true; and (ii) the set of all worlds in which the proposition is false. Carnap, therefore, proposed to identify a proposition p with the function f_p from the set W of all possible worlds to truth-values which for every possible world w has the value $f_p(w) =$ the truth-value of p in the world w. Thus, propositions are identified with functions from possible worlds (or in Carnap's case, from state descriptions, or set-theoretical models, that are taken to represent possible worlds) to truth-values. The *intension* of a sentence is the proposition it expresses and its *extension* in a possible world w is the truth-value in w of the proposition it expresses.

The intension of a predicate expression is intuitively the property (or relation-in-intension) that the predicate expresses. A property of individuals determines for every possible world w, the set of individuals that has the property in that world. Hence, a *property* P, can according to Carnap and Montague be identified with a function f_P from the set W of all possible worlds to sets of individuals, which for every possible world w

[20] As emphasised by Church [22] and Kaplan [60], the Fregean tradition in intensional logic should be distinguished from the quite different tradition stemming from Russell where the sense/denotation distinction is avoided. Russellian semantics, in contrast to Fregean semantics, assigns only one kind of semantic value, most naturally thought of as a kind of denotation, to the well-formed expressions of a language. In Russellian semantics, (logically) proper names refer (directly) to objects, sentences designate Russellian propositions, i.e. complexes of properties and objects, and predicates stand for propositional functions. Modern so-called theories of direct reference belong to the Russellian tradition (cf., for instance, [98]).

has the value $f_P(w)$ = the set of all entities that in the world w has the property P. For instance, the property of being red, is identified with the function form possible worlds to individuals that associates with every possible world the set of red objects in that world. Similarly, an n-ary *relation-in-intension* R is identified with a function from possible worlds to sets of ordered n-tuples. The intension of a predicate expression is the property or relation-in-intension it expresses and its extension in a possible world w is the set or relation-in-extension that is the value of that intension in the world w.

Finally, singular terms have individuals as their extensions and their intensions are what Carnap calls *individual concepts*, i.e., functions from possible worlds to individuals. The singular term 'the Greek philosopher that taught Alexander the Great' has in the actual world Aristotle as its extension. In another possible world, the extension may be Plato. In possible worlds where there is no unique Greek philosopher that taught Alexander, the singular term might be assigned an arbitrary conventional extension, the *null extension.* Since proper names, presumably, are *rigid designators* (cf. [71]) they have the same extension in every possible world (or at least in every possible world where the bearer of the name exists). Hence, the intension of a proper name is a constant function picking out the same object in every possible world (or at least this is the case for rigid designators of objects that exist necessarily, for instance, the numerals designating the natural numbers). On Kripke's view, co-referring proper names have the same intension. As a result, if a and b are co-referring proper names, then '$a = a$' and '$a = b$' have the same intension. Thus, it seems that difference in cognitive significance cannot be explained by difference in intension.

Kripke's [66, 67, 68] major innovation was his use — within each model structure — of a set of abstract points (indices, "possible worlds") to represent the space of possibilities. This innovation made it possible for Montague [84] — building on ideas from Carnap [17] — to represent intensional entities (senses, intensions) by set-theoretic functions from points (representing possible worlds) to extensions. Every kind of meaningful expression has according to Carnap-Montague semantics a suitable *intension*, i.e., a function from possible worlds to appropriate extensions. If E is an expression with intension $Int(E)$, and w is a possible world, then $Int(E)(w)$, i.e., the result of applying the intension of E to the possible world w, is the *extension of E in the world w* (in symbols $Ext_w(E)$). The *extension of E*, $Ext(E)$, is the extension of E in the actual world.

Following Carnap [17] we distinguish between different kind of constructions (or contexts) Φ:

(i) Φ is *extensional* iff there exists a function f_Φ such that for every possible world w, and all (appropriate) expressions $E_1, \ldots, E_n$, $Ext_w(\Phi(E_1, \ldots, E_n)) = f_\Phi(Ext_w(E_1), \ldots, Ext_w(E_n))$. An *extensional language* is a language where every grammatical construction is extensional. An extensional language satisfies the *principle of extensionality*, i.e., the principle that the extension of a complex expression is always a function of the extensions of its semantically meaningful constituents.

(ii) Φ is *intensional* iff there exists a function F_Φ such that for all (appropriate) expressions $E_1, \ldots, E_n$, $Int(\Phi(E_1, \ldots, E_n)) = F_\Phi(Int(E_1), \ldots, Int(E_n))$. An intensional language is a language in which every grammatical construction is intensional. Intensional languages satisfy the *principle if intensionality*, i.e., the principle that the intension of a complex expression is always a function of the intensions of its semantically meaningful constituents.

The principles of extensionality and intensionality are special cases of the *principle of compositionality*, i.e., the principle that the meaning of a complex expression is determined by its structure and the meaning of its constituents (cf., [104]).

The classical Boolean connectives are, of course, paradigm examples of extensional constructions. By modifying the above definitions slightly, in order to take variable binding operators into account, the classical quantifiers $\forall$ and $\exists$ are naturally construed as extensional operators as well. The modal operators $\Box$ and $\Diamond$, on the other hand, are examples of constructions that are intensional but not extensional. Carnap also considered propositional attitude constructions like 'John believes that ... ', that in his opinion were not even intensional. Such constructions for which the principle of intensionality fails, may be called *ultraintensional*.

In order to give a semantic analysis of belief contexts, Carnap introduced the notion of *intensional isomorphism* [17, §14]. Roughly speaking, two expressions are intensionally isomorphic iff they are built up from atomic expressions with the same intensions in the same way. Intensionally isomorphic expressions were said to have the same *intensional structure*. The intensional structure of an expression can thus be identified with the equivalence class of all expressions of the given language that are intensionally isomorphic with it. Intensional isomorphism and intensional structure was Carnap's explications of the intuitive notions of synonymy and meaning, respectively.[21] The intensional structures that correspond to sentences may be viewed as *structured propositions* in contrast to Carnapian propositions (functions from possible worlds to truth-values) that lack syntactical structure.[22] Carnap suggested that belief and other propositional attitudes be operators on such structured propositions rather than on intensions. If so, then intensionally isomorphic expressions are substitutable *salva veritate* in propositional attitude contexts. This seems fairly reasonable since one might argue that synonymous expressions are substitutable in such contexts.

Montague's intensional logic **IL** is a typed λ-calculus.[23] There are two basic types e and t of (possible) *individuals* and *truth-values* (true and false), respectively. Then, there is for every two types α and β, a type $(\alpha\beta)$ of *functions* from entities of type α to entities of type β. Finally, for every type α, there is a type $(s\alpha)$ of *senses* appropriate for entities of type α. Montague identifies the senses with Carnapian intensions, i.e., the members of $(s\alpha)$ are functions from possible worlds to entities of type α. All the domains of the various types are constant from one world to another. In particular, there is one domain of individuals that is common to all possible worlds. Thus, the domain of individuals is best thought of as the domain of all *possible* individuals.

For every type α, the language of **IL** contains variables and non-logical constants of type α. It also contains the logical constants: $=$ (identity), λ (lambda-abstraction), $\hat{}$ (intensional abstraction), $\check{}$ (intensional application), and brackets [,]. The sentential connectives, quantifiers $\forall$, $\exists$, and modal operators $\Box, \Diamond$, are definable in terms of $=$, λ, $\hat{}$, and $\check{}$ (Gallin [33, 15-16]). For each type α, one can quantify in IL over all the entities of type α. In particular, one can quantify over the collection of all *possible individuals.*

[21] This theme is developed further in Lewis [76].

[22] See King [63] for an overview of more recent work on structured propositions and references to the relevant literature (including work by David Kaplan, Nathan Salmon, Scott Soames, Jeff King, and others within the "direct reference"-tradition on so-called "Russellian propositions").

[23] See Montague [84], and especially Gallin [33] for a thorough model-theoretic study of Montague's intensional logic. In particular, Gallin presents an axiomatisation of Montague's intensional logic and proves that it is strongly complete with respect to general Henkin-type models.

In other words, **IL** is committed to an ontology of possible individuals.

Complex terms of **IL** are built up from atomic terms (variables and constants as follows): (i) If A is a term of type $(\alpha\beta)$ and B is a term of type α, then $[AB]$ is a term of type β; (ii) If A is a term of type β and x is a variable of type α then $\lambda x A$ is a term of type $(\alpha\beta)$: (iii) If A, B are terms of the same type, then $[A = B]$ is a term of type t; (iv) If A is a term of type α, then $\hat{}A$ is a term of type $(s\alpha)$; (v) If A is a term of type $(s\alpha)$, then $\check{}A$ is a term of type α. Terms of type t are called *formulæ*.

In the semantics, every (closed) term A of type α is assigned an extension $Ext_w(A)$ of type α relative to w, for each possible world w. The intension $Int(A)$ of A is then the function from worlds to extensions such that for each w, $Int(A)(w) = Ext_w(A)$. For each term A, $\hat{}A$ is a name of the intension of A. And, for each term A denoting an intension F, $\check{}A$ is a term which at every world w, refers to the value of F at w. Hence, $(A = \check{}\hat{}A)$ will always hold. The semantics of **IL** satisfies the principle of intensionality and $\hat{}$ is the only primitive non-extensional construction of **IL**. The modal operator $\Box$ is defined in **IL** as follows:

$$\Box\varphi =_{df.} [\hat{}\varphi = \hat{}T],$$

that is, φ is necessarily true iff the intension of φ equals the intension of any tautology T. $\Box$ is an **S5**-operator and the Barcan formulæ and their converses are valid in the semantics.

Montague's intensional logic admits quantifying into intensional constructions. According to Montague's intended interpretation, the individual quantifiers range over *possible* individuals. Quantification over actual individuals can be analysed by means of the introduction of an existence predicate. However, Montague's use of quantifiers ranging over possibilia is of course an abomination in the eyes of Quine and likeminded philosophers who favour an actualist metaphysics.

1.5.2 Church's logic of sense and denotation

The expressions of natural language are according to the Fregean view *systematically ambiguous*: both the sense and the denotation of an expression vary with the linguistic context in which it occurs. This systematic ambiguity is the basis for Church's program [19, 20, 21] of representing natural language discourse involving oblique contexts within a formal language the logic of which is completely *extensional*, that is, in which the ordinary principles of substitutivity of classical logic are valid. His fundamental idea is to let each expression A of the natural language be represented by different expressions $A_0, A_1, A_2, \ldots$ of the formal language depending on the context in which A occurs. Suppose, for instance, that the sentence "Tom is married", when it occurs in a non-oblique context, is translated as **Married(Tom)**. Then, the sentence (1), where the verb phrase "suspects that" gives rise to an oblique context, may be represented as:

(2) $\mathbf{Suspects}(\mathbf{Mary}, \mathbf{Married}_1(\mathbf{Tom}_1))$,

where $\mathbf{Married}_1$ and $\mathbf{Tom}_1$ are atomic expressions that denote the (ordinary) senses of **Married** and **Tom**, respectively. Analogously,

(3) George knows that Mary suspects that Tom is married

may be represented as

(4) $\mathbf{Knows}(\mathbf{George}, \mathbf{Suspects}_1(\mathbf{Mary}_1, \mathbf{Married}_2(\mathbf{Tom}_2)))$.

Using Church's terminology, we may say that $\mathbf{Tom}_1$ and $\mathbf{Tom}_2$ denote *the concept of being Tom* and *the concept of being the concept of being Tom,* respectively. In this way ambiguity is avoided in the representing language and the classical principles of substitutivity as well as all other principles of classical logic are preserved.

Church's logic of sense and denotation is a simple type theory that has much in common with Montague's intensional logic **IL** but which differs from **IL** in not violating the principle of extensionality. In Montague's language there is, as we recall, only one non-extensional operator $\hat{}$ which transforms a term A into a term $\hat{}A$ that denotes the intension of A. Since A occurs in $\hat{}A$, $\hat{}$ is non-extensional. Church's logic of sense and denotation, on the other hand, is fully extensional. For each denoting expression A, there is in Church's language another expression $\langle A \rangle$, denoting the sense of A. Since $\langle A \rangle$ does not contain A as a syntactic part, the occurrence of A in the language does not violate extensionality. $\langle A \rangle$ replaces A in oblique contexts. For instance, the indirect discourse construction: 'John believes that φ' is replaced by the direct discourse version: 'John believes $\langle \varphi \rangle$', where $\langle \varphi \rangle$ is a name of the proposition expressed by the sentence φ. The construction 'John believes $\langle \varphi \rangle$' differs from 'John believes $\hat{}\varphi$' in being fully extensional.

In Church [18] and [19], three alternative principles of individuation for senses were proposed referred to as Alternatives (0), (1) and (2). The alternative that individuates senses most coarsely is Alternative (2), according to which two expressions have the same sense iff they are logically equivalent. Roughly speaking, Alternative (2) amounts to identifying Fregean senses with Carnapian intensions, i.e., with functions from possible worlds (or models or state descriptions representing possible worlds) to denotations (or extensions). Thus, Alternative (2) is the alternative which is closest to modern possible worlds semantics.

The alternative that is closest to Frege's own conception of sense is probably Alternative (0), according to which two terms A and B have the same sense, if and only if they are *intensionally isomorphic* in the sense of Carnap [17]. In addition to alternatives (0) and (2), Church also considered an intermediate alternative called Alternative (1), which is fairly close to Alternative (0) but seems to have less intuitive motivation. According to Alternative (1) expressions that are lambda-convertible to each other have the same sense.

Church's logic of sense and denotation is not directly concerned with linguistic expressions and their senses and denotations, but rather with the language-independent so-called *concept relation* that holds between senses and the entities that they are senses of. As Church points out in [21], the more finely senses are individuated, the more closely will the abstract theory of senses and their objects resemble the more concrete theory of names and their denotations, with the concept relation playing a role similar to the one played by the *denotation predicate* of semantics. Consequently, antinomies analogous to the semantic antinomies may arise for formulations of the logic of sense and denotation along the lines of Alternative (0) or (1). Indeed, Myhill [85] points out that Church's Alternative (0) is threatened by the antinomy described by Russell in *The Principles of Mathematics*, Appendix B, p. 527, the so-called *Russell-Myhill paradox* (cf. Anderson [2]).

The development of a logic of sense and denotation along the lines of Alternative (0) — taking Carnap's intensional isomorphism, Church's synonymous isomorphism, or some related notion as a criterion for two expressions having the same sense — is of great theoretical interest. First of all, the fundamental principle of Alternative (0):

$$sense(FA) = sense(FB) \rightarrow sense(A) = sense(B),$$

seems to be involved whenever a difference in sense between FA and FB is *explained* in terms of a difference in sense between A and B. Secondly, any principle of individuation for senses that is substantially less strict than Alternative (0) seems to be inadequate for a Fregean treatment of the logic of propositional attitudes.

Unfortunately, however, the attempts so far to develop a logic of sense and denotation along the lines of Alternative (0) have led either to inconsistency or to great complications, for instance, in the form of an infinite hierarchy of concept relations of different orders. Furthermore, no entirely satisfactory explanation has so far been given of the notion of sense involved in Alternative (0). Related to this is the lack of an intuitive semantic theory for Alternative (0) and a corresponding notion of logical validity.

However, the pursuit of Church's Alternative (2) has made considerable progress. Thus, David Kaplan [58, 60] and Charles Parsons [88] have provided versions of Church's logic of sense and denotation with a possible worlds semantics of Carnap-Montague type. Parsons [88] even shows that his version of Church's logic of sense and denotation is exactly equivalent to (intertranslatable with) Montague's intensional logic. Moreover he provides an axiomatisation of Church's Alternative 2 that is equivalent to Gallin's axiomatisation of Montague's intensional logic.

1.6 Logical and metaphysical necessity

We make a rough distinction between two types of intuitive interpretations of the operators $\Diamond$ and $\Box$ of alethic modal logic. First there is the *metaphysical* or *counterfactual* interpretation:

$\Diamond\varphi$: either φ, or it could have been the case that φ.
$\Box\varphi$: φ, and it could not have been the case that $\neg\varphi$.

Then, there is the *logical* or *metalogical* interpretation:

$\Diamond\varphi$: it is not self-contradictory to assume that φ is the case.
$\Box\varphi$: it is self-contradictory to assume that $\neg\varphi$ is the case.

From now on, we shall use $\mathbf{L}\varphi$ and $\mathbf{M}\varphi$ for the logical modalities and reserve $\Box$ and $\Diamond$ for the metaphysical ones.

According to the *possible worlds analysis* of metaphysical necessity:

$\Box\varphi$ is true at a possible world w iff φ is true at every possible world.

There is an extensive and fast growing philosophical literature on the proper analysis of the notion of a possible world (cf. [25, 87]). Roughly speaking, we are distinguishing between *the world* as the (concrete) totality of everything there is and *possible worlds* as total alternative ways the world could have been (cf. [71, pp. 15–20]). Characterised in this way, possible worlds are abstract entities: *total possible states of the world.* This notion of possible world should be contrasted with David Lewis's notion of a possible world as a concrete alternative universe (cf. [80]). Regardless of our ultimate understanding of possible worlds, to say that a statement φ is *true at* a possible world w means, intuitively, that φ, with its actual meaning, would have been true (simpliciter) had w obtained.

A delicate question that now arises is how metaphysical necessity relates to logical necessity. The answer, of course, depends on how precisely we characterise the notion of logical necessity. Different semantic characterisations give rise to different answers. Suppose that we define logical necessity in terms of a class K of (admissible) models (or interpretations). Each model $\mathcal{M}$ is associated with a set $U_{\mathcal{M}}$ of points (representing "possible worlds") of which one is the designated point $@_{\mathcal{M}}$ (representing "the actual world"). We write $u \vDash_{\mathcal{M}} \varphi$ for the sentence φ being *true at the point u in the model $\mathcal{M}$*. *Truth in a model $\mathcal{M}$* is defined as truth at the designated point $@_{\mathcal{M}}$ of the model $\mathcal{M}$. *Logical truth*, or *validity*, is defined as truth in every model in K. We assume that:

(i) $u \vDash_{\mathcal{M}} \neg\varphi$ iff not: $u \vDash_{\mathcal{M}} \varphi$

(ii) $u \vDash_{\mathcal{M}} (\varphi \rightarrow \psi)$ iff either $u \nvDash_{\mathcal{M}} \varphi$ or $u \vDash_{\mathcal{M}} \psi$.

(iii) $u \vDash_{\mathcal{M}} \mathbf{L}\varphi$ iff for every model $\mathcal{N}$ in $K, @_{\mathcal{N}} \vDash_{\mathcal{N}} \varphi$.

(iv) $u \vDash_{\mathcal{M}} \Box\varphi$ iff for every point $v \in U_{\mathcal{M}}, v \vDash_{\mathcal{M}} \varphi$.

Given this type of semantics, there is no guarantee that logical necessity implies metaphysical necessity. Suppose, for example, that the language contains a logical constant **actually** with the semantic clause:

(v) $u \vDash_{\mathcal{M}}$ **actually** (φ) iff $@_{\mathcal{M}} \vDash_{\mathcal{M}} \varphi$,

i.e., **actually** (φ) is true at a point in a model iff φ is true at the designated point in the model. Then, every instance of the following schema is valid:

(1) $\mathbf{L}(\varphi \leftrightarrow$ **actually** $(\varphi))$,

although, the following schema fails (in both directions):

(2) $\Box(\varphi \leftrightarrow$ **actually** $(\varphi))$.

We can easily construct models $\mathcal{M}$ for a sentential language of the indicated kind for which (2) fails.

Thus it appears, as Zalta [108] has argued, that logical necessity does not imply metaphysical necessity. There are logical truths that are metaphysically contingent. However, this claim is highly counterintuitive. There are various ways of avoiding the conclusion that logical truth does not imply metaphysical necessity. One may, for one reason or another, refuse constructions like **actually**, that make reference to special worlds, the status of logical constants.

Another option is to modify the notion of logical truth. The notion of logical truth that we have employed is the one we have called *real-world validity*. It is the notion according to which a statement φ is logically true (valid) iff it is true at the actual world in each model. As we have seen, however, there is an alternative notion, *general validity*, according to which a statement is logically true iff it is true at each world in each model.

Let us write $\vDash$ and $\vDash^*$ for real-world validity and general validity, respectively. The two notions are related as follows: For any statement φ,

(1) $\vDash \varphi$ iff $\vDash^*$ **actually** (φ).

(2) $\vDash^* \varphi$ iff $\vDash \Box\varphi$.

The operator **L** was introduced by "reflecting" the meta-linguistic notion of real-world validity into the object language. We can also introduce an operator $\mathbf{L}^*$ corresponding to the notion of general validity. The semantic clauses for **L** (*real-world logical necessity*) and $\mathbf{L}^*$ (*general logical necessity*) are:

(vi) $u \vDash_{\mathcal{M}} \mathbf{L}\varphi$ iff for every model $\mathcal{N}$ in $K, @_{\mathcal{N}} \vDash_{\mathcal{N}} \varphi$.

(vii) $u \vDash_{\mathcal{M}} \mathbf{L}^*\varphi$ iff for every model $\mathcal{N}$ in K and every point v in $\mathcal{N}, v \vDash_{\mathcal{N}} \varphi$.

That is, **L** corresponds to truth at the actual world in each model and $\mathbf{L}^*$ corresponds to truth at every world in each model. The two notions of logical necessity are interdefinable:

(1) $\vDash^* \mathbf{L}\varphi \leftrightarrow \mathbf{L}^*\mathbf{actually}(\varphi)$.

(2) $\vDash^* \mathbf{L}^*\varphi \leftrightarrow \mathbf{L}\Box\varphi$.

Moreover, we have:

(3) $\vDash^* \mathbf{L}^*\varphi \to \Box\varphi$,

although, as we have seen, the corresponding implication does not hold for real-world logical necessity, i.e., for **L**.

Metaphysical necessity does not imply logical necessity. It does not appear self-contradictory to think, as the Greeks did, that water is an element. But since water, as it turned out, is a compound of oxygen and hydrogen, it could not have been an element. There is, so to speak, no counterfactual situation, or possible world, where *water* is not a compound. So even if it is not logically necessary, it is metaphysically necessary that water is a compound. Hence, the statement:

(1) Water is a compound

is metaphysically necessary (assuming that "water", is a rigid designator), but it is not logically necessary. In conclusion, we can say that real-world logical necessity (**L**) neither implies nor is implied by metaphysical necessity ($\Box$). General logical necessity ($\mathbf{L}^*$) on the other hand, implies metaphysical necessity, but is not implied by it.

The (epistemological) distinction between *a priori* and *a posteriori* also comes in here. In Kripke's theory, (1) exemplifies a statement that, although metaphysically necessary, is nevertheless *a posteriori*. On the other hand, given certain assumptions, "The Paris meter is one meter long" may be an example of a statement that is true *a priori* but is not metaphysically necessary [71].

2 THE MODAL LOGIC OF BELIEF CHANGE

In this section, modal logic is brought to bear on an area which has already reached a degree of maturity (although still in need of further development) and which has been formulated with little or no regard to modal logic. By re-formulating the theory in terms of modal logic, a degree of systematisation is gained, and — it is hoped! — the theoretical understanding of the theory is enhanced.