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

(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, $S \vDash \varphi$) if it is true in S relative to every assignment. *Logical truth* (logical validity) is defined as truth in all state descriptions. We write φ 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) $\vDash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$.
(T) $\vDash \Box\varphi \to \varphi$
(S4) $\vDash \Box\varphi \to \Box\Box\varphi$.
(S5) $\vDash \neg\Box\varphi \to \Box\neg\Box\varphi$
(BF) $\vDash \forall x\Box\varphi(x) \to \Box\forall x\varphi(x)$. (*The Barcan formula*)
(CBF) $\vDash \Box\forall x\varphi(x) \to \forall x\Box\varphi(x)$. (*The Converse Barcan formula*)
(Nec) If $\vDash \varphi$, then $\vDash \Box\varphi$.

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) $\vDash \Box\varphi$ iff $\vDash \varphi$.

(9) $\vDash \neg\Box\varphi$ iff $\nvDash \varphi$.

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

For identity, we have:

(LI) $\vDash t = t$. (*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:

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

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

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

For the unrestricted case, we only have:

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

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

$(\Box =)$ $\forall x \forall y(x = y \rightarrow \Box(x = y))$. (*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:

$(\Box \neq)$ $\forall x \forall y(x \neq y \rightarrow \Box(x \neq y))$. (*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′) $S \vDash \Box\varphi$ iff, for every admissible state description S', $S' \vDash \varphi$.

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

$S \vDash \Box\varphi$ iff φ is analytically true.
$S \vDash \neg\Box\varphi$ 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:

'$\Box\varphi$' is true iff φ 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 $\mathcal{L}_U$, $a \neq a \in m$,
$(C.\text{Ind})$ 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 $I = \langle D, I, g\rangle$, 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 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) $\langle D, I, g\rangle \vDash \mathbf{L}\varphi$ iff for every I' defined over D, $\langle D, I', g\rangle \vDash \varphi$.

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=) $\forall x\forall y(x = y \rightarrow (\varphi(x/z) \rightarrow \varphi(y/z)))$. *(Indiscernibility of Identicals)*

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

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

But the following is not valid:

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

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

(4) $\mathbf{L}(\forall x \exists y(x = y) \rightarrow \exists y(x = y))$ 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) $\forall x \mathbf{L}\varphi(x) \rightarrow \mathbf{L}\forall x \varphi(x)$
(CBF) $\mathbf{L}\forall x \varphi(x) \rightarrow \forall x \mathbf{L}\varphi(x)$.

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) $V \vDash_{\mathcal{M}} (\varphi \to \psi)[g]$ 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 '$(\exists y(y = x) \to x \in \{\text{Socrates}\})$' and let $\Box$ stand for metaphysical necessity. Then, $\Box\varphi$(Socrates) is true. Socrates is a member of {Socrates}, in every possible world where Socrates exists. But, of course, $\Box\varphi$(Plato) is false. Thus (I) fails for metaphysical necessity.

We say that φ is *true in* $\mathcal{M}$ *relative to* g if $V_0 \vDash_{\mathcal{M}} \varphi[g]$. φ is *true in* $\mathcal{M}$ if $V_0 \vDash_M \varphi[g]$ 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 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 (to appear) 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) $w \vDash_{\mathcal{M}} (\varphi \to \psi)[g]$ iff $w \nvDash_{\mathcal{M}} \varphi[g]$ or $w \vDash_{\mathcal{M}} \psi[g]$.

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

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

We say that φ is *true with respect to the model* $\mathcal{M}$ *and the assignment* g (in symbols $\vDash_{\mathcal{M}} \varphi[g]$), 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, $\vDash_{\mathcal{M}} \varphi$), if for every assignment $g, \vDash_{\mathcal{M}} \varphi[g]$. φ is *true in a frame* $\mathcal{F}$ (in symbols, $\vDash_{\mathcal{F}} \varphi$) 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 $\vDash_{\mathbf{K}}$ and $\vDash^*_{\mathbf{K}}$ 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) $\vDash^*_{\mathbf{K}} \varphi$ iff $\vDash_{\mathbf{K}} \Box\varphi$

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 $\vDash_{\mathbf{K}} \varphi$ iff $\vDash^*_{\mathbf{K}} \varphi$

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: $\Box\forall x\varphi(x) \rightarrow \forall x\Box\varphi(x)$, is generally valid in every model based on $\mathcal{F}$.

(BF) Every instance of the Barcan formula: $\forall x\Box\varphi(x) \rightarrow \Box\forall x\varphi(x)$, 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* $\mathbf{E}$ 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 $\mathbf{E}$ is definable in terms of the existential quantifier and identity:

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

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

In terms of $\mathbf{E}$ 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 $\Box(\Diamond\mathbf{E}(x) \rightarrow \mathbf{E}(x))$ is valid in $\mathcal{F}$.

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

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

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 $\mathcal{M}, \vDash_M \varphi[g]$. 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 veritat*e 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 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 λxA 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) **Knows**(**George**, **Suspects**$_1$(**Mary**$_1$, **Married**$_2$(**Tom**$_2$))).

Using Church's terminology, we may say that **Tom**$_1$ and **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 φs 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) $u \vDash_M (\varphi \rightarrow \psi)$ iff either $u \nvDash_{\mathcal{M}} \varphi$ or $u \vDash_M \psi$.

(iii) $u \vDash_M \mathbf{L}\varphi$ 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) $u \vDash_M$ **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 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) $\vDash^* \mathbf{L}^*\varphi \rightarrow \Box\varphi$,

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.

2.1 Introduction

2.1.1 Two paradigms

The theory of belief change is a fairly sprawling phenomenon. In the tradition examined here, one is interested in how new information is handled by a "rational" agent. By assumption, the agent is situated in an environment, often referred to as "the world". The world is always in some world state or other, and the agent is always in some belief state or other. From time to time, the agent is presented with new information about the world. The problem is to describe how the new information affects the current belief state. In the two cases studied here, two further assumptions are made: that the new information is always accepted, and that acceptance always leads to a uniquely defined (usually, but not necessarily, different) belief state.

We distinguish between two cases: the case when the world is static (the world does not change) and the case when the world is dynamic (the world might change); belief change is called *belief revision* in the former case, *belief update* in the latter. We also distinguish between two attitudes which an agent may have and which are called *conditionalisation* and *imaging*, respectively; these terms, which have an origin in probability theory, will not be explained (see Lewis [128], Gärdenfors [112]). The two paradigms selected for study here, AGM and KM, exemplify those two attitudes: AGM is a conditionalising and KM is an imaging theory. It is commonly accepted that a conditionalising attitude is appropriate for belief revision and an imaging attitude for update. Thus AGM is said to be a theory of belief revision and KM a theory of update.

In this section we shall offer explications within modal logic of both the AGM paradigm and the KM paradigm. They are not meant to be exact counterparts of AGM and KM as they were historically defined; they are rather meant to bring out what we take to be essential to those conceptions. (Our use of the terms "AGM" and "KM" is ambiguous: they stand both for certain people (Alchourrón, Gärdenfors and Makinson in the former case, Katsuno and Mendelzon in the latter) and for the theories developed by those authors.)

2.1.2 Revision

There is a strong connection between the theory of belief change and the logic of conditionals. That this should be so is not so surprising if it is remembered that the following much quoted passage in a paper of Frank Ramsey, published posthumously, inspired both fields:

> If two people are arguing 'If p will q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q; so that in a sense 'If p, q' and 'If $p, \sim q$' are contradictories. We can say that they are fixing their degrees of belief in q given p. If p turns out false, these degrees of belief are rendered *void*. If either party believes p for certain, the question ceases to mean anything for him except as a question about what follows from certain laws or hypothesis. [133, p. 149].

Both Robert Stalnaker and David Lewis cite this passage as a point of departure for their respective theories of conditional logic (Stalnaker [140], Lewis [127]). But it was read also by Peter Gärdenfors, who was looking for a different kind of theory of conditionals, one

with a semantics not formulated in terms of possible worlds. Given a formal language of some familiar sort, what an agent believes on a certain occasion may be represented by what Gärdenfors called a *belief set*, namely, the set of propositions believed by the agent. It is assumed that belief sets are theories in Tarski's sense, that is, that they contains all tautologies and are closed under modus ponens. Fundamental for Gärdenfors's theory of belief change is the existence of an operation $*$ such that, for any T, if T is an agent's belief set on a certain occasion and φ is a proposition, then $T * \varphi$ is the agent's belief set if and after he has revised his beliefs by φ. Given $*$, Ramsey may be read as suggesting that two people, who share a belief set T and argue 'If φ will ψ?', can be represented as arguing whether ψ is an element of $T * \varphi$. In this way Gärdenfors was led to look for *assertability conditions* for conditionals rather than *truth conditions*. In particular, he had hoped to find a conditional $\Rrightarrow$ satisfying the following form of the so-called Ramsey Test:

$$(RT) \qquad \varphi \Rrightarrow \psi \in T \text{ iff } \psi \in T * \varphi.$$

Everything now hangs on the properties of the revision operation $*$. Proceeding in the same way as C. I. Lewis when the latter was trying to characterise his modal operators, Gärdenfors laid down a number of postulates in order to characterise $*$. Let K be a certain *background theory*, that is, a special belief set that is taken for granted and subject to revision. A *K-theory* is a theory that includes K. We say that a formula φ is *K-consistent* if the set $K \cup \{\varphi\}$ is consistent, that a formula φ is *K-consistent with* a belief set T if $T \cup \{\varphi\}$ is consistent, and that two formulæ φ and ψ are *K-equivalent* if $\varphi \leftrightarrow \psi \in K$. We write $Cn(T)$ for the set $\{\theta : \exists\varphi(\varphi \in T \text{ and } \varphi \rightarrow \theta \in T)\}$.

(AGM1) For any K-theory T and formula $\varphi, T * \varphi$ is a K-theory.

(AGM2) $\varphi \in T * \varphi$

(AGM3) $T * \varphi \subseteq Cn(T \cup \{\varphi\})$.

(AGM4) φ is K-consistent with T, then $Cn(T \cup \{\varphi\}) \subseteq T * \varphi$

(AGM5) If φ is K-consistent, then $T * \varphi$ is K-consistent.

(AGM6) If φ and ψ are K-equivalent, then $T * \varphi = T * \psi$.

(AGM7) $T * (\varphi \wedge \psi) \subseteq Cn(T * \varphi \cup \{\psi\})$.

(AGM8) If ψ is K-consistent with $T * \varphi$, then $Cn(T * \varphi \cup \{\psi\}) \subseteq T * (\varphi \wedge \psi)$.

Some of these postulates have received their own names in the literature: (AGM2) is the Success Postulate, (AGM4) the Preservation Postulate and (AGM5) the Consistency Postulate.

This, in a nutshell, is the syntactic side of AGM, the famous paradigm created by Gärdenfors in collaboration with Carlos Alchourrón and David Makinson [109, 113]. Now it turns out, as Gärdenfors himself was the first to observe, that if the condition (RT) is added to the AGM-postulates (after the new operator $\Rrightarrow$ has been added to the object language), then the result is, not inconsistency, but triviality. This is yet another interesting example of how intuitions, which on the face of it seem quite reasonable, turn out jointly to be incompatible. But it is also a wonderful example of the old saying that they who seek will find, but not always what they are looking for. For even though Gärdenfors did not find his conditional, he did find, together with Alchourrón and Makinson, a seminal theory of belief revision.

2.1.3 Update

A different theory of belief change is given in Katsuno and Mendelzon [122] (see also Grahne [115]). They emphasise a distinction, which they attribute to Keller and Winslett [123], between *knowledge-adding changes* (revisions) and *change-recording updates*. According to Katsuno and Mendelzon, we believe that the real world is one of a certain set of possible worlds, which one we may not know. When we are informed that the real world has changed in a certain respect, we examine each of the old possible worlds and ask how our beliefs would have changed if that particular one had been the real world (notice the counterfactual!). "The fact the real world has changed gives us no grounds to conclude that some of the old worlds were actually not possible." (This is the feature that made us classify Katsuno and Mendelzon's theory as imaging.)

Where AGM have belief sets, KM have *knowledge bases* (abbreviated KBs), each knowledge base consisting of just one formula ("since we need a finite fixed representation of a KB to store it in a computer"). Like AGM, KM also introduce a new operator: if φ is a KB and ψ is a formula (intuitively, the new information) then $\varphi\diamondsuit\psi$ is the KB that results from updating φ with ψ. Assuming a propositional language with only finitely many letters, they propose the following postulates:

(KM1) $\varphi\diamondsuit\psi$ implies ψ.

(KM2) If φ implies ψ, then $\varphi\diamondsuit\psi$ is equivalent to φ.

(KM3) If both φ and ψ are satisfiable, then $\varphi\diamondsuit\psi$ is also satisfiable.

(KM4) If φ is equivalent to φ' and ψ is equivalent to ψ' then $\varphi\diamondsuit\psi$ is equivalent to $\varphi'\diamondsuit\psi'$.

(KM5) $(\varphi\diamondsuit\psi)\wedge\theta$ implies $\varphi\diamondsuit(\psi\wedge\theta)$.

(KM6) If $\varphi\diamondsuit\psi$ implies ψ' and $\varphi\diamondsuit\psi'$ implies ψ, then $\varphi\diamondsuit\psi$ is equivalent to $\varphi\diamondsuit\psi'$.

(KM7) If φ is such that, for all χ, φ implies χ or φ implies $\neg\chi$, then $(\varphi\diamondsuit\psi)\wedge(\varphi\diamondsuit\psi')$ implies $\varphi\diamondsuit(\psi\vee\psi')$.

(KM8) $(\varphi\vee\varphi')\diamondsuit\psi$ is equivalent to $(\varphi\diamondsuit\psi)\vee(\varphi'\diamondsuit\psi)$.

An important difference between the two paradigms is that, while the AGM operator $*$ is not part of the language in which the agent's beliefs are expressed, the KM operator $\diamondsuit$ is. For this reason, AGM is not a logic, in the usual sense of the word, but KM is.

2.1.4 Translations

To a modal logician, it is obvious that AGM can be re-formulated as a modal logic. A Rosetta stone with inscriptions in ordinary language, AGM language and the language of modal logic might contain the following text:

ordinary language	AGM	modal logic
the agent believes that φ	$\varphi\in T$	$\mathbf{B}\varphi$
after revising his beliefs by φ, the agent believes that χ	$\chi\in T*\varphi$	$[*\varphi]\mathbf{B}\chi$

What is called modal logic here is of course doxastic logic enriched with change operators loaned from dynamic logic. In particular, if φ is a formula, then the expression $*\varphi$ functions like a term that denotes 'the agent's acceptance of φ' — an event, possibly an action. In this way the nonformal theory of AGM can be translated, more or less faithfully, into a formal theory within what we shall call dynamic doxastic logic (DDL); details are provided below. (We prefer "doxastic" to "epistemic" since belief need not be veridical.)

To give a direct translation of KM is more difficult. The KM-operator $\diamondsuit$ (a binary operator not to be confused with the homonymous unary higher-order operator appearing in DDL-operators of type $[\diamondsuit\varphi]$) is a propositional connective but not one of classical logic. A KM-formula

$(*)\ (\varphi\diamondsuit\psi) \rightarrow \chi$

might at first sight seem to represent the claim that if an agent, who believes that φ, updates his beliefs by ψ, then he will believe, after the update, that χ. If so, then the DDL-formula $\mathbf{B}\varphi \rightarrow [\diamondsuit\psi]\mathbf{B}\chi$ would be a natural translation of $(*)$. However, there is a great difference between the total of an agent's beliefs — a knowledge base, to use KM-terminology — and a single belief of the agent. Therefore a faithful translation into DDL requires a strengthening of our current language. One possibility would be to adopt an operator $\mathbf{E}$ of a kind first considered by Hector Levesque, $\mathbf{E}\varphi$ carrying the intuitive meaning "the agent believes exactly that φ (and what follows logically from φ)" or "all that the agent believes is that φ (and what follows logically from φ)". In this more expressive language

$$\mathbf{E}\varphi \rightarrow [\diamondsuit\psi]\mathbf{B}\chi$$

would be an adequate translation of $(*)$. Unfortunately, Levesque's operator is not easy to axiomatise (see [124]).

2.1.5 Some object languages

A number of object languages will figure in the sequel, and it is a good idea to give careful definitions of them at this point. Let lett be a denumerable set of letters. We assume a truth-functionally complete supply of *boolean* operators, *conditional* operators $\sqsupset$ and $>$, *doxastic* operators $\mathbf{B}$, $\mathbf{b}$, $\mathbf{K}$, and $\mathbf{k}$, as well as a *star operator* $*$, a *rhombus operator* $\diamondsuit$, and *change operators* $[\ ]$ and $\langle\ \rangle$. The operators $\mathbf{B}$, $\mathbf{K}$ and $[\ ]$ are so-called box operators, while the operators $\mathbf{b}$, $\mathbf{k}$ and $\langle\ \rangle$ are dual so-called diamond operators; for simplicity, in what follows we shall treat the latter as abbreviatory devices: that is, for all appropriate formulæ $\chi, \mathbf{b}\chi =_{\mathrm{df}} \neg\mathbf{B}\neg\chi, \mathbf{k}\chi =_{\mathrm{df}} \neg\mathbf{K}\neg\chi, \mathbf{b}\chi =_{\mathrm{df}} \neg\mathbf{B}\neg\chi, \langle*\varphi\rangle\chi =_{\mathrm{df}} \neg[*\varphi]\neg\chi$ and $\langle\diamondsuit\varphi\rangle\chi =_{\mathrm{df}} \neg[\diamondsuit\varphi]\neg\chi$. In the same way, we also stipulate that $\varphi > \psi =_{\mathrm{df}} \neg(\varphi \sqsupset \neg\psi)$.

Informally, formulæ of type $\varphi \sqsupset \psi$ and $\varphi > \psi$ may be read as "if φ then ψ" or, if a distinction between them is called for, as "if φ then certainly ψ" and "if φ then conceivably ψ". (But we are not committed to any particular reading, be it "ontic", "epistemic", "dynamic", or whatever.) The operators $\mathbf{B}$ and $\mathbf{K}$ are for belief and for doxastic commitment, respectively. For many purposes a reading of laziness of $\mathbf{K}\varphi$ as "the agent knows that φ" is all right, but "the agent is doxastically committed to φ" is better, for implicit in the theories we are considering below is the assumption that what is referred to is implied by a certain, usually not specified, sometimes possibly variable, background theory. The change operators are the after-operators of dynamic logic.

Classical language
LETT $\subseteq$ BOOLE,
$\varphi, \psi \in$ BOOLE $\Rightarrow (\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi), \neg\varphi \in$ BOOLE.

Basic doxastic language
BOOLE $\subseteq$ BASIC·DOX,
$\varphi, \psi \in$ BASIC·DOX$\Rightarrow (\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi), \neg\varphi \in$ BASIC·DOX,
$\varphi \in$ BOOLE $\Rightarrow \mathbf{B}\varphi, \mathbf{b}\varphi, \mathbf{K}\varphi, \mathbf{k}\varphi \in$ BASIC·DOX.

Full doxastic language
BOOLE $\subseteq$ FULL·DOX,
$\varphi, \psi \in$ FULL·DOX $\Rightarrow$
$(\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi), \neg\varphi, \mathbf{B}\varphi, \mathbf{b}\varphi, \mathbf{K}\varphi, \mathbf{k}\varphi \in$ FULL·DOX.

Basic revision language
BASIC·DOX $\subseteq$ BASIC·REV,
$\varphi, \psi \in$ BASIC·REV $\Rightarrow (\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi), \neg\varphi \in$ BASIC·REV,
$(\varphi \in$ BOOLE $\ \&\ \chi \in$ BASIC·REV$) \Rightarrow [*\varphi]\chi, \langle *\varphi\rangle\chi \in$ BASIC·REV.

Full revision language
FULL·DOX $\subseteq$ FULL·REV,
$\varphi, \psi \in$ FULL·REV $\Rightarrow (\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi), \neg\varphi \in$ FULL·REV,
$(\varphi \in$ FULL·DOX $\ \&\ \chi \in$ FULL·REV$) \Rightarrow [*\varphi]\chi, \langle *\varphi\rangle\chi \in$ FULL·REV.

Unlimited revision language
FULL·DOX $\subseteq$ UNLIM·REV,
$\varphi, \psi \in$ UNLIM·REV $\Rightarrow$
$(\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi), \neg\varphi, \mathbf{B}\varphi, \mathbf{b}\varphi, \mathbf{K}\varphi, \mathbf{k}\varphi \in$ UNLIM·REV,
$(\varphi \in$ UNLIM·REV $\ \&\ \chi \in$ UNLIM·REV$) \Rightarrow [*\varphi]\chi, \langle *\varphi\rangle\chi \in$ UNLIM·REV.

Conditional language
BOOLE $\subseteq$ COND,
$\varphi, \psi \in$ COND $\Rightarrow$
$(\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi), \neg\varphi, (\varphi \sqsupset \psi), (\varphi > \psi) \in$ COND.

Update language
BASIC·DOX $\cup$ COND $\subseteq$ UPDATE,
$\varphi, \psi \in$ UPDATE $\Rightarrow (\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi), \neg\varphi \in$ UPDATE,
$\varphi \in$ COND $\Rightarrow \mathbf{B}\varphi \in$ UPDATE,
$(\varphi \in$ COND $\ \&\ \chi \in$ UPDATE$) \Rightarrow [\Diamond\varphi]\chi, \langle\Diamond\varphi\rangle\chi \in$ UPDATE.

We say that a formula is an *agent formula*, relative to BASIC·REV, FULL·REV, UNLIM·REV, or UPDATE, if it can occur both within the scope of a doxastic operator and within the scope of the star operator or rhombus operator, whichever is appropriate.

2.2 Conditional logic

As remarked above, there is a close connection between the theory of belief change and the theory of conditionals. For this reason, we devote an entire section to this topic, which is also, by our lights, a chapter of modal logic.

2.2.1 Topology

Let U be any nonempty set. A *topology* in U is a set T of subsets of U satisfying two conditions: for all $S \subseteq T$, (i) $\bigcup S \in T$, and (ii) if S is finite and nonempty, then $\bigcap S \in T$. A topology always has at least two elements: U is the union of all subsets of U and is therefore a member, and $\varnothing$ is the union of the empty set of subsets of U and is therefore also a member. The structure (U, T) is called a *topological space*, but when it is clear what the intended topology is one usually refers to U itself as the topological space. The elements of T are said to be *open* sets; a *closed* set is one that is the complement of an open set. In general, a set need not be either open or closed, but on the other hand some sets are both; we will use the term *clopen* (adjective or noun) for the latter. Notice that the complement of a clopen set is clopen and that U and $\varnothing$ are clopen in any topology. The (topological) *closure* of a set X, the smallest closed set that includes X, is defined as the intersection of all closed sets that X.

A *cover* of a set $X \subseteq U$ is a family C of subsets of U such that $X \subseteq \bigcup C$. A cover, every element of which is an open set, is an *open cover*. If C is a cover of X and there is a family $D \subseteq C$ such that $X \subseteq \bigcup D$, then D is a *subcover* of C of X. A topology T is *compact* if every open cover of the whole space has a finite subcover; a logically equivalent condition is that every family of closed subsets of U whose intersection is empty has a finite subfamily whose intersection is empty. A topology T is *totally separated* if, for any pair of distinct elements of U, one is an element of a clopen set of which the other is not. A *Stone topology* is a topology that is compact and totally separated.

A family B of subsets of T is a *base* for the topology if, for every $X \in T$, there is some family $C \subseteq B$ such that $X = \bigcup C$. In other words, B is a base if every open element is the union of some elements of B. It is not difficult to prove that in a Stone topology, the family of clopen sets forms a base.

Let U be a space with a Stone topology. A *sphere system* or, more colloquially, an *onion* (*in* U) is a nonempty family of closed subsets (*spheres*) of U that satisfies two conditions: it is closed under arbitrary nonempty intersection, and it is linearly ordered by set inclusion. An onion is *trivial* if it contains only one sphere and that sphere is the empty set; hence there is a unique trivial onion, namely $\{\varnothing\}$. The *centre* of an onion O is the set $\bigcap O$, and we say that O is centred on $\bigcap O$; thus the trivial onion is centred on the empty set. We say that an onion O *overlaps with* a set X (we assume that X is a subset of U) if $\bigcup O \cap X \neq \varnothing$. The family of spheres of an onion O that intersect with a set X is denoted by $O \bullet X$. If S is a family of sets and X is the smallest element of S, then we may express this by the notation $X\ \mu\ S$ (thus "mu" is a special case of "epsilon"). It is not difficult to prove that if O is an onion that overlaps with a clopen set X, then there is a smallest sphere in O that intersects with X; in symbols, $\bigcup O \cap X \neq \varnothing \Rightarrow \exists Z(Z\ \mu\ (O \bullet X))$; this important condition is called *the limit condition*.

Sphere systems were introduced by David Lewis (who never called them onions) [127]. Ours differ from his in one notable respect: his spheres, but not ours, are closed under unrestricted union. One particular consequence of Lewis's condition is that the empty set is an element of every sphere system of his, while our onions may, but need not, contain the empty set as an element.

The reader may find it helpful to think of the clopen sets as the agent's propositions of the frame, and closed sets as possible agent's theories (or theory sets, to use a term of Bengt Hansson). The topological setting may surprise some readers, but it provides an elegant way for keeping tabs on the limit condition.

2.2.2 Semantics

Let us say a quadruple (U, P, Q, D) is a *Lewis frame* if (i) U is a Stone space, (ii) P is the set of clopen sets, (iii) Q is a quantity (that is, set) of onions in U, (iv) D is a function (the *onion determiner*) assigning to each element u of U an onion Q, and (v) whenever X and Y are clopen subsets of U, then so are

$$\{u \in U : \forall Z(Z\ \mu\ (D(u) \bullet X) \Rightarrow Z \cap X \subseteq Y)\}$$

and

$$\{u \in U : \exists Z(Z\ \mu\ (D(u) \bullet X)\ \&\ Z \cap X \cap Y \neq \varnothing)\}.$$

We consider a language cond for conditional logic. A *valuation* in a Lewis frame (U, P, Q, D) is a function from the set lett of propositional letters to P. A *Lewis model* (U, P, Q, D, V) is a Lewis frame (U, P, Q, D) cum valuation V. We define the truth of a formula in a given Lewis model $\mathcal{M} = (U, P, Q, D, V)$ as follows (we suppress reference to $\mathcal{M}$ in the notation). The definition, which proceeds by induction in the usual way, is relative to a point u of U. We use the notation $[\![\varphi]\!]$ for the set called the *truth set* of φ. If $u \in [\![\varphi]\!]$ we say that φ is *true* at u and may write $u \vDash \varphi$.

$[\![\varphi]\!] = V(P)$, if P is a propositional letter,
$[\![\varphi \wedge \psi]\!] = [\![\varphi]\!] \cap [\![\psi]\!]$,
$[\![\varphi \vee \psi]\!] = [\![\varphi]\!] \cup [\![\psi]\!]$,
$[\![\neg\varphi]\!] = U = [\![\varphi]\!]$, etc.,
$[\![\varphi \sqsupset \psi]\!] = \{u \in U : \forall Z(Z\ \mu\ D(u) \bullet [\![\varphi]\!] \Rightarrow Z \cap [\![\varphi]\!] \subseteq [\![\psi]\!])\}$,
$[\![\varphi > \psi]\!] = \{u \in U : \exists Z(Z\ \mu\ D(u) \bullet [\![\varphi]\!]\ \&\ Z \cap [\![\varphi]\!] \cap [\![\psi]\!] \neq \varnothing)\}$.

(Note that, thanks to the closure rules on P, including (v), $[\![\varphi]\!]$ is a clopen set, for every formula φ.) We say that a formula is *valid* if it is true at all points in all models. A schema is *valid* if all its instances are valid.

2.2.3 Postulates for David Lewis's VC and VCU

First, assume as postulates all tautologies and the rule of modus ponens. Then add the rules

(REA) $\varphi \leftrightarrow \varphi' / (\varphi \sqsupset \theta) \leftrightarrow (\varphi' \sqsupset \theta)$,

(REC) $\theta \leftrightarrow \theta' / (\varphi \sqsupset \theta) \leftrightarrow (\varphi \sqsupset \theta')$

and, as axioms, all instances of the following schemata:

(ML1) $(\varphi \sqsupset (\psi \wedge \theta)) \leftrightarrow ((\varphi \sqsupset \psi) \wedge (\varphi \sqsupset \theta))$,

(ML2) $\varphi \sqsupset \top$,

(DF>) $(\varphi > \psi) \leftrightarrow \neg(\varphi \sqsupset \neg\psi)$,

(CL1) $\varphi \sqsupset \varphi$,

(CL2) $(\varphi > \psi) \rightarrow (\psi > \top)$,

(CL3) $\varphi \rightarrow (\psi \rightarrow (\varphi \sqsupset \psi))$,

(CL4) $\varphi \to ((\varphi \sqsupset \psi) \to \psi)$,

(CL5) $((\varphi \wedge \psi) \sqsupset \theta) \to (\varphi \sqsupset (\psi \to \theta))$,

(CL6) $(\varphi > \psi) \to ((\varphi \sqsupset (\psi \to \theta)) \to ((\varphi \wedge \psi) \sqsupset \theta))$.

Brian Chellas suggested a different notation which highlights the connection with modal logic: writing $[\varphi]\psi$ for $\varphi \sqsupset \psi$ and $\langle\varphi\rangle\psi$ for $\varphi > \psi$. If we rewrite the preceding conditions in Chellas's notation, we get the following result:

(REA′) $\varphi \leftrightarrow \varphi'/[\varphi]\theta \leftrightarrow [\varphi']\theta$,

(REC′) $\theta \leftrightarrow \theta'/[\varphi]\theta \leftrightarrow [\varphi]\theta'$

and, as axioms, all instances of the following schemata:

(ML1′) $[\varphi](\psi \wedge \theta) \leftrightarrow ([\varphi]\psi) \wedge [\varphi]\theta)$,

(ML2′) $[\varphi]\top$,

(DF$\langle\ldots\rangle$) $\langle\varphi\rangle\psi \leftrightarrow \neg[\varphi]\neg\psi)$,

(CL1′) $[\varphi]\varphi$,

(CL2′) $\langle\varphi\rangle\psi \to \langle\psi\rangle\top$,

(CL3′) $\varphi \to (\psi \to [\varphi]\psi)$,

(CL4′) $\varphi \to ([\varphi]\psi \to \psi)$,

(CL5′) $[\varphi \wedge \psi]\theta \to [\varphi](\psi \to \theta)$,

(CL6′) $\langle\varphi\rangle\psi \to ([\varphi](\psi \to \theta) \to [\varphi \wedge \psi]\theta)$.

The set of theorems of this axiom system coincides with David Lewis's logic VC. To get his VCU, add the schemata:

(4) $\Box\varphi \to \Box\Box\varphi$

(5) $\neg\Box\varphi \to \Box\neg\Box\varphi$,

where $\Box\varphi$ is an abbreviation of $\neg\varphi \sqsupset \bot$ (or $[\neg\varphi]\bot$, in Chellas's notation). (Note that the schema $\Box\varphi \to \varphi$ is derivable in VC).

We say after Lewis that a frame (U, T, Q, D) is *centred* if the onion $D(u)$ is centred on $\{u\}$ for each $u \in U$, and we say that it is *uniform* if $\bigcup D(u) = \bigcup D(v)$, for all $u, v \in U$.

THEOREM 3 ([127]). *A formula of conditional logic is derivable in VC [alternatively: in VCU] if and only if it is valid in all centred [alternatively: uniform centred] Lewis frames.*

There are of course many more completeness results of this kind.

2.3 Update and the logic of conditionals

2.3.1 Semantics

Consider an update language update. Let $\mathcal{M} = (U, P, Q, D, V)$ be a given Lewis model. The definition of truth in M (we will suppress reference to $\mathcal{M}$ in the notation) will be with respect to a *situation*, defined as an ordered pair (B, u) where B is a subset of U and x is a point in U. (Intuitively, B represents the current beliefs of the agent, while x represents the point that is currently actual.) The definition proceeds in two steps. First, we define the *truth set* $[\![\theta]\!]$ of formulæ $\theta \in$ COND in the usual way; thus $[\![\theta]\!]$ is a subset of U. Second, we define *truth in a situation* (B, x) of formulæ φ:

$$
\begin{array}{l}
(B,x) \vDash \varphi \text{ iff } x \in [\![\varphi]\!], \text{ if } \varphi \in \text{ COND},\\
(B,x) \vDash \varphi \wedge \psi \text{ iff } (B,x) \vDash \varphi \text{ and } (B,x) \vDash \psi\\
(B,x) \vDash \varphi \vee \psi \text{ iff } (B,x) \vDash \varphi \text{ or } (B,x) \vDash \psi,\\
(B,x) \vDash \neg\varphi \text{ iff } (B,x) \nvDash \varphi,\\
\qquad \text{etc.},\\
(B,x) \vDash \mathbf{B}\varphi \text{ iff } B \subseteq [\![\varphi]\!],\\
(B,x) \vDash \mathbf{K}\varphi \text{ iff } \bigcup\{\bigcup D(u) : u \in B\} \subseteq [\![\varphi]\!],\\
(B,x) \vDash [\Diamond\varphi]\chi \text{ iff } (B',x) \vDash \chi, \text{ where}\\
\qquad B' = \bigcup\{Z \cap [\![\varphi]\!] : \exists u(u \in B \;\&\; Z \;\mu\; (D(u) \bullet [\![\varphi]\!]))\}.
\end{array}
$$

We say that a formula is *valid* if it is true in all situations in all models. A schema is *valid* if all its instances are valid.

2.3.2 Postulates

We build an axiom system in stages, one block at the time. First block: all postulates (axioms and rules) of Lewis's VC. Second block: normal modal logic for all modal and dynamic operators. Third block:

($\Diamond$0) $\quad \theta \leftrightarrow [\Diamond\varphi]\theta$, if $\theta \in$ COND,

($\Diamond$1) $\quad \langle\Diamond\varphi\rangle\chi \leftrightarrow [\Diamond\varphi]\chi$,

($\Diamond$RT) $\quad \mathbf{B}(\varphi \sqsupset \psi) \leftrightarrow [\Diamond\varphi]\mathbf{B}\psi$

($\Diamond$K) $\quad \mathbf{K}\varphi \leftrightarrow \ \mathbf{B}\Box\varphi$

($\Diamond$RC) $\quad$ if $\varphi \leftrightarrow \psi$ is derivable, then so is $[\Diamond\varphi]\chi \leftrightarrow [\Diamond\psi]\chi$, for every χ, for all formulæ $\varphi, \psi \in$ COND and $\chi \in$ UPDATE.

(Here, as above, $\Box\varphi$ abbreviates $\neg\varphi \sqsupset \bot$.) The Ramsey condition is essentially a condition of operator shift where both operators and positions change; this fact is especially striking if ($\Diamond$RT) is rewritten in Chellas's notation:

($\Diamond$RT′) $\quad \mathbf{B}[\varphi]\psi \leftrightarrow [\Diamond\varphi]\mathbf{B}\psi$.

Call this system U. (Warning: Lewis's U (for "uniform") must not be be confused with our U (for "update").) Let U45 be the system obtained by adding the schemata (4) and (5) mentioned above.

THEOREM 4. *A formula of the update language is derivable in U [alternatively: in U45] if and only if it is valid in all centred [alternatively: uniform centred] Lewis frames.*

Our axiomatisation, in which ($\Diamond$RT) is the only postulate that is really novel, has thus issued in yet another confirmation of the observation of several authors about the close connection between conditional logic and update logic. It is worth remarking that the class of uniform centred frames, which in our object language determines UT45, also, in another object language which does not include the operator **B**, determines the logic Gösta Grahne calls VCU^2, a logic based directly on Lewis's VCU (see [115]).

2.4 Revision and basic DDL

2.4.1 Semantics

In this section we assume a language basic-rev for basic revision logic. Two intuitions underlying our presentation are that belief change consists in the transition from belief state to belief state, and that belief states can be modelled by sphere systems (onions). Let us define a *basic revision frame* as a structure (U, P, Q, R) where U is a Stone space, P is the set of clopen sets, Q is a quantity of onions, and R is a function assigning to each clopen set X a binary relation RX over Q. The intuition is this: if the agent is in belief state O, then after accepting (the information carried by) a proposition X, his new belief state is a belief state O' such that $(O, O') \in RX$.

Valuations and models are defined as usual. We define the *truth* of a formula in a given model $\mathcal{M} = (U, P, Q, R, V)$ as follows (as usual, we suppress reference to $\mathcal{M}$ in the notation). The definition, which proceeds by induction, is relative to a *situation*, which is an ordered pair (O, x), where O is an onion and x a point of U. (Intuitively, O is the current belief state of the agent, and x is the current state of the world.) We use the notation $[\![\varphi]\!]$ for the truth set of φ if φ is a Boolean formula.

$$
\begin{array}{l}
(O, x) \vDash P \text{ iff } x \in V(P), \text{ if } P \text{ is a propositional letter,} \\
(O, x) \vDash \varphi \wedge \psi \text{ iff } (O, x) \vDash \varphi \text{ and } (O, x) \vDash \psi, \\
(O, x) \vDash \varphi \vee \psi \text{ iff } (O, x) \vDash \varphi \text{ or } (O, x) \vDash \psi, \\
(O, x) \vDash \neg\varphi \text{ iff } (O, x) \nvDash \varphi, \text{ etc.,} \\
(O, x) \vDash \mathbf{B}\varphi \text{ iff } \bigcap O \subseteq [\![\varphi]\!], \\
(O, x) \vDash \mathbf{K}\varphi \text{ iff } \bigcup O \subseteq [\![\varphi]\!], \\
(O, x) \vDash [*\varphi]\chi \text{ iff } \forall O'((O, O') \in R[\![\varphi]\!] \Rightarrow (O', x) \vDash \chi).
\end{array}
$$

Notice that the truth-conditions for the dynamic operators make sense as long as the star operator applies only to Boolean formulæ. *Validity* in a frame [in a class of frames] is then defined as truth relative to all situations in all models on the frame [in all frames].

To try to capture the ideas behind the historical AGM, further conditions are in order. One is that the belief set of a new belief state resulting from some new piece of information equal the overlap between the old onion and the clopen set representing that information:

(i) $(O, O') \in RX$ only if O overlaps with X and $\bigcap O' = Z \cap X$, where $Z \; \mu \; O \bullet X$.

Two other conditions are that every relation RX be serial and functional:

(ii) $\exists O' \in Q(O, O') \in RX$,

(iii) $(O, O') \in RX \;\&\; (O, O'') \in RX \Rightarrow O' = O''$.

Yet another condition to be considered is that what AGM called the "background" theory (in our jargon, the agent's doxastic commitments) not change when beliefs are revised:

(iv) $O, O' \in Q \Rightarrow \bigcup O = \bigcup O'$.

2.4.2 Translations of the AGM postulates

A direct translation of the AGM postulates formulated in Section 2.1.2 gives the following result:

(∗2) $[*\varphi]\mathbf{B}\varphi$,

(∗3) $[*\varphi]\mathbf{B}\chi \rightarrow \mathbf{B}(\varphi \rightarrow \chi)$,

(∗4) $\mathbf{b}\varphi \rightarrow (\mathbf{B}\chi \rightarrow [*\varphi]\mathbf{B}\chi)$,

(∗5) $\mathbf{k}\varphi \rightarrow \langle *\varphi\rangle \mathbf{b}\top$,

(∗6) $\mathbf{K}(\varphi \leftrightarrow \psi) \rightarrow ([*\varphi]\mathbf{B}\chi \leftrightarrow [*\psi]\mathbf{B}\chi)$,

(∗7) $[*(\varphi \wedge \psi)]\mathbf{B}\chi \rightarrow [*\varphi]\mathbf{B}(\psi \rightarrow \chi)$,

(∗8) $\langle *\varphi\rangle \mathbf{b}\psi \rightarrow ([*\varphi]\mathbf{B}(\psi \rightarrow \chi) \rightarrow [(\varphi \wedge \psi)]\mathbf{B}\chi$.

All instances of these schemata are valid.

2.4.3 Postulates for the basic-DDL version of AGM

We build an axiom system in stages, one block at the time. First block: tautologies and modus ponens. Second block: normal modal logic for **B** and **K** and $[*\varphi]$. Third block: the postulates (∗2)–(∗8) mentioned in the preceding section plus the following additional postulates:

(∗0) $\chi \leftrightarrow [*\varphi]\chi$, if χ is Boolean,

(∗1) $\langle *\varphi\rangle\chi \leftrightarrow [*\varphi]\chi$,

(∗KB) $\mathbf{K}\varphi \rightarrow \mathbf{B}\varphi$,

(∗K) $\mathbf{K}\chi \leftrightarrow [*\varphi]\mathbf{K}\chi$.

(RC) if $\varphi \leftrightarrow \psi$ is derivable, then so is $[*\varphi]\chi \leftrightarrow [*\psi]\chi$, for every χ.

THEOREM 5 ([167]). *A formula of basic revision language is provable in the given axiom system if and only if it is valid in all basic revision frames.*

2.4.4 Comparison with KM

The two paradigms AGM and KM seem rather different. It is interesting, therefore, that the differences between the corresponding logics is not greater than it is. Of the AGM postulates, all are valid according to KM as well, except for the most controversial ones: (∗4) and (∗8). But even in these cases the two paradigms come close:

DDL version of KM	DDL version of AGM
$\mathbf{B}\varphi \rightarrow (\mathbf{B}\chi \rightarrow [*\varphi]\mathbf{B}\chi))$	$\mathbf{b}\varphi \rightarrow (\mathbf{B}\chi \rightarrow [*\varphi]\mathbf{B}\chi)$
$\langle *\varphi\rangle\mathbf{B}\psi \rightarrow$	$\langle *\varphi\rangle\mathbf{b}\psi \rightarrow$
$([*\varphi]\mathbf{B}(\psi \rightarrow \chi) \rightarrow [*(\varphi \wedge \psi)]\mathbf{B}\chi)$	$([*\varphi]\mathbf{B}(\psi \rightarrow \chi) \rightarrow [*(\varphi \wedge \psi)]\mathbf{B}\chi)$

Among schemata valid in KM but not in AGM are $\mathbf{B}\bot \rightarrow [*\varphi]\mathbf{B}\bot$ and $\mathbf{B}\bot \rightarrow \mathbf{K}\bot$.

One difference between revision and update, remarked upon by Katsuno and Mendelzon, is what they call the "global" behaviour of revision versus the "local" behaviour of update. What they have in mind can be explicated within our framework as follows. In AGM, the belief states of an agent are represented by onions; the belief set of an agent is not enough to determine the entire onion. By contrast, in KM the belief set is enough to determine the belief state of the agent. The reason for this is of course the centred onions assigned to each point ("possible world") in the universe of a frame. In the latter case, the beliefs of the agent come in two steps: beliefs about the world (represented by centred onions), and beliefs about which possible world is the actual one (represented by a belief set). In AGM, belief change is a progression from onion to onion. In KM, belief change is from belief set to belief set, but against the background of an underlying, constant web of beliefs about how the world can change.

2.5 Revision and full or unlimited DDL

2.5.1 Semantics

Basic DDL tries to explicate AGM as originally formulated. This is why in basic DDL the agent's beliefs are all about the world, and the agent's beliefs are not part of the world. As we saw above, the language of AGM and the language of DDL are intertranslatable. Nevertheless, there is one sense in which DDL offers more flexibility: where AGM has $\chi \in T$, DDL offers $\mathbf{B}\chi$, but AGM has no counterpart to $\mathbf{BB}\chi$ — $(\chi \in T) \in T$ is not a well-formed expression. There is no reason why one could not extend the language of AGM to include the **B**-operator, but no-one seems to have done so. And rather than doing so, it seems easier to study the resulting theory in a DDL context.

Therefore, let us move to the language FULL·REV of full revision, in which the agent formulæ are the formulæ of the full doxastic language FULL·DOX. Define a *full revision frame* as a structure (U, P, Q, R, D) *where* (i) U is a Stone space, (ii) P is the set of clopen sets, (iii) Q is a quantity of onions (not necessarily centred), (iv) R is a function assigning to each clopen set X a binary relation RX over Q, (v) D is a function from $U$to Q and finally (vi) whenever X is a clopen subset of U, then both $\{u \in U : \bigcap D(u) \subseteq X\}$ and $\{u \in U : \bigcup D(u) \subseteq X\}$ are clopen. *Truth* at a point u in a model (U, P, Q, R, D, V) is defined along usual lines:

$u \vDash P$ iff $u \in V(P)$, if P is a propositional letter,
$u \vDash \varphi \wedge \psi$ iff $u \vDash \varphi$ and $u \vDash \psi$
$u \vDash \varphi \vee \psi$ iff $u \vDash \varphi$ or $u \vDash \psi$,
$u \vDash \neg\varphi$ iff $u \nvDash \varphi$, etc.,
$u \vDash \mathbf{B}\varphi$ iff $\bigcap O \subseteq [\![\varphi]\!]$, where $O = D(u)$,
$u \vDash \mathbf{K}\varphi$ iff $\bigcup O \subseteq [\![\varphi]\!]$, where $O = D(u)$,
$u \vDash [*\varphi]\chi$ iff $\forall v((u, v) \in R[\![\varphi]\!] \Rightarrow v \vDash \chi)$.

As usual, a formula is *valid* in a frame if true at all points in all models.

One difference between basic and full DDL is that in the former case the points of a frame represent world states whereas in the latter case they simultaneously represent both world state and belief state. This is why in basic DDL formulæ have to be evaluated at pairs (O,x) where O represents a belief state and x a world state, while in full DDL formulæ are evaluated at points representing total or combined states. In DDL there is thus an ambiguity in the informal term "world state": in a narrow sense, which excludes the agent's beliefs, the points of basic, but not full, DDL represent world states; but in a wide sense, which includes the agent's beliefs, the points of full, but not basic, DDL represent world states. In any case, the intuition in full DDL is this: if the current total state is u, then if the agent accepts (the information carried by) a proposition X, there will be, immediately afterwards, a new current total state v such that $(u, v) \in RX$.

The semantics of unlimited DDL is the same as that of full DDL, with two exceptions: the language whose formulæ are given a meaning is UNLIM·REV, and the definition of an *unlimited revision frame* is obtained from the definition of a full revision frame by adding condition (vii) if X and Y are clopen subsets of U, then both $\{u \in U : \forall v((u, v) \in RX \Rightarrow v \in Y\}$ and $\{u \in U : \exists v((u, v) \in RX \ \&\ v \in Y\}$ are clopen. Intuitively, the points of an unlimited revision frame represents not only the state of the world and the agent's beliefs about the state of the world but also the agent's beliefs about how the world may change.

2.5.2 Redefining revision?

It is interesting that all the old postulates of basic DDL (but now over the formulæ of FULL·REV or UNLIM·REV) are satisfied. However, there are problems. Suppose, for example, that $\mathbf{b}\varphi$ and $\mathbf{B}\neg B\varphi$ are true in a certain situation. Then $[*\varphi]\mathbf{B}\neg\mathbf{B}\varphi$ follows by preservation. By the Success Postulate, we always have $[*\varphi]\mathbf{B}\varphi$. Hence $[*\varphi]\mathbf{B}(\varphi \wedge \neg\mathbf{B}\varphi)$, by modal logic. Or suppose that $\mathbf{b}\varphi$ and $\mathbf{BB}\neg\varphi$ are true in a certain situation. By the same kind of reasoning $[*\varphi]\mathbf{B}(\varphi \wedge \mathbf{B}\neg\varphi)$ follows. Or, even more problematic, note that both $[*(\varphi \wedge \neg\mathbf{B}\varphi)]\mathbf{B}(\varphi \wedge \neg\mathbf{B}\varphi)$ and $[*(\varphi \wedge \mathbf{B}\neg\varphi)]\mathbf{B}(\varphi \wedge \mathbf{B}\neg\varphi)$ are valid. But if it is true, on a certain occasion, that it is raining in Umeå and Sten, who happens to be visiting at Ojmundsbod far from Umeå, does not believe that it is raining in Umeå, or even believes that it is not raining in Umeå, then surely it should be possible for him to accept this information without incurring doxastic inconsistency?

This problem was first noted and left unresolved in van Linder, van der Hoek and Meyer [129]. Two strategies have later been suggested for dealing with it. One is given in Lindström and Rabinowicz [131] in which it is recommended that one perform a certain contraction before revising one's beliefs. Roughly speaking, before accepting new

information, the agent should give up enough currently held beliefs to make sure that new information does not create doxastic inconsistency (assuming the new information is itself consistent). A different though related strategy is proposed in Segerberg [138] where it is recommended that the notion of revision be redefined. Two possibilities are outlined. In both cases, the star operator $*$ is kept but a new revision operator R is introduced. One suggestion is to define

$$[\mathrm{R}\varphi]\chi =_{\mathrm{df}} [*(\varphi \wedge \mathbf{B}\varphi \wedge \ldots \wedge \mathbf{B}^n\varphi)]\chi$$

and require that the logic of $\mathbf{B}$ be of a certain strength, for example, contain at least the schemata $\mathbf{BB}\varphi \rightarrow \mathbf{B}\varphi$ and $\mathbf{B}^n\varphi \rightarrow \mathbf{B}^{n+1}\varphi$. The other suggestion is to introduce yet a new doxastic operator $\mathbf{C}$ (for "complete" belief) with the semantics

$$u \vDash \mathbf{C}\varphi \text{ iff } \forall n > 0(u \vDash \mathbf{B}^n\varphi),$$

define

$$[\mathrm{R}\varphi]\chi =_{df} [*(\varphi \wedge \mathbf{C}\varphi)]\chi$$

and then require the logic of $\mathbf{B}$ to validate the schema $\mathbf{BB}\varphi \rightarrow \mathbf{B}\varphi$. Evidently, the operator $\mathbf{C}$ in effect represents common belief when only one agent is involved. One may note the validity of the following schemata:

$$\begin{array}{l}\mathbf{C}\varphi \rightarrow \mathbf{B}\varphi,\\ \mathbf{C}\varphi \rightarrow \mathbf{CC}\varphi,\\ \mathbf{BC}\varphi \leftrightarrow \mathbf{CB}\varphi,\\ (\mathbf{B}\varphi \wedge \mathbf{C}(\varphi \rightarrow \mathbf{B}\varphi)) \rightarrow \mathbf{C}\varphi.\end{array}$$

In this modelling, which assumes that the doxastic commitments of agents are not open to revision, it is impossible for an agent who values the consistency of his beliefs to accept either the information that it-is-raining-and-he-does-not-believe-that-it-is-raining or the information that it-is-raining-and-he-believes-that-it-is-not-raining. In the terminology of Roy Sorensen [139], $\varphi \wedge \neg\mathbf{B}\varphi$ and $\varphi \wedge \mathbf{B}\neg\varphi$ represent *blindspots*. In general, φ represents a blindspot at u if $\mathbf{k}\varphi \wedge [\mathrm{R}\varphi]\mathbf{B}\bot$ is true at u — if φ is consistent with what the agent knows but revision by φ leads to an absurd belief set.

3 LOGIC OF ACTION AND DEONTIC LOGIC

For natural reasons, deontic logic has been in the hands of deontic logicians from the beginning. As is the case with all modal logicians who do not concentrate on the purely formal aspect of their discipline, they have been acting as philosophers and as logicians at the same time, and so conceptual issues and technical treatment have been intermingled. It is remarkable that, even though deontic logic has been around for a long time, there is as yet not an accepted body of work that extends very far. What is needed to improve the situation in deontic logic, it seems, is to identify and philosophically discuss basic concepts in greater depth than has been done before. Not least must we develop better logics of action. Modal logic should be in a privileged position to inform such work, or so we argue in this essay.

3.1 Logic of action

3.1.1 Logic of action without actions

In the history of the logic of action, there is a line from Anselm in the eleventh century, restarted in modern times by authors like Kanger, Fitch, and A. R. Anderson, and continued by Chellas, which has recently received its most mature expression yet by Belnap, Perloff and Xu. We quickly sketch a version of a theory in this tradition.

We say that a structure (U, A, T, H, E) is a *frame* if the following conditions are met. U is a set of *points* (informally, representing possible (total) states of the world, just as in dynamic logic). A is a finite set of *agents*. T is a linearly ordered set (we refer to the elements of T as *times*), and a *T-history* is a function from T into U. Let H be a certain set of T-histories (from now on, just histories). If $h \in H$ and $t \in T$, then (h, t) is said to be a *moment*. We define two families of equivalence relations over H. First we define two histories h and g as *coinciding up through t*, in symbols, $h \sim_t g$, if $ht' = gt'$, for all t' earlier than t (that is, all t' such that $t' < t$). For each moment (h, t), define $H_{h,t} = \{g \in H : h \sim_t g\}$. E is a function assigning to each moment (h, t) and agent $a \in A$ a partitioning $E_{h,t,a}$ of $H_{h,t}$. We say that two histories h and g are *action-equivalent for a at t* if h and g are equivalent under $E_{h,t,a}$ (or, equivalently, if h and g are equivalent under $E_{g,t,a}$); we write $h \approx_{t,a} g$ if this is the case. It is clear that both $\sim_t$ and $\approx_{t,a}$ are equivalence relations.

Consider a classical propositional language (for example, BOOLE — see above) to which is added, for each natural number i, a propositional operator D_i with the informal reading "the agent denoted by i brings it about that". A *valuation* in a frame (U, A, T, H, E) is a function assigning to each natural number an element in A and to each propositional letter a set of moments. A *model* on a frame is the frame together with a valuation. The truth-value (*truth* or *falsity)* of a formula at a moment in a model is defined with respect to moments along traditional lines. The formal condition for the action operator D_i is

$$(h, t)D_i\varphi \text{ iff } V(i) = a \;\&\; \exists t_0 < t(\forall g(h \approx_{t_0,a} g \Rightarrow (g, t)\varphi) \;\&\; \exists g(h \sim_{t_0} g \;\&\; (g, t) \nVdash \varphi)).$$

The idea is that, if a is the agent denoted by i, then *a brings it about, with respect to h and t, that φ*, in symbols $D_i\varphi$, iff there exists a time t_0 earlier that t such that two conditions are satisfied, (i) (the positive condition) that φ be true with respect to g and t, where g is *any* history that is action-equivalent with h for a at t_0, and (ii) (the negative condition) that φ be false with respect to g and t, where g is *some* history coinciding with h up through t_0. In other words, speaking somewhat freely, we might say that the positive condition guarantees that φ is true in a certain important respect, while the negative condition shows the necessity of that guarantee.

From a formal point of view (but not philosophically: see [144, p. 197 f.]) the system sketched may be said to present, more or less in the spirit of Chellas [150], the theory of Belnap and Perloff of an operator called by them the "achievement stit" ("stit" for "**s**ees **t**o **i**t **t**hat"). If the negative condition of the truth definition is omitted and we require T to be the set of all (negative and nonnegative) integers, we get a definition which is essentially that of Chellas's do-operator. (In the latter case, the element t_0 mentioned in the truth condition of D_i should be identified as $t - 1$.)

As an operator of agency, the operator D_i differs from many of its competitors in the literature. For example, in the Chellas version, D_i is a normal modal operator; in

particular, $D_i\top$ and $D_i(\varphi \to \psi) \to (D_i\varphi \to D_i\psi)$ and $D_i(\varphi \wedge \psi) \leftrightarrow (D_i\varphi \wedge D_i\psi)$ are all valid (that is, true at all moments in all models); in the Belnap version, none of them is. In particular, even though $(D_i\varphi \wedge D_i\psi) \to D_i(\varphi \wedge \psi)$ is valid in the Belnap version, $D_i(\varphi \wedge \psi) \to (D_i\varphi \wedge D_i\psi)$ is not; only the weaker $D_i(\varphi \wedge \psi) \to (D_i\varphi \vee D_i\psi)$ is valid.

Sometimes, "i is causally responsible for the fact that φ" would seem to be a better translation of $D_i\varphi$ than "i sees to it that φ". For let φ be "the door is closed" and $D_i\varphi$ thus "the agent sees to it that the door is closed" (presumably equivalent to "the agent closes the door"). The validity of $D_i\varphi \to \varphi$ (in both the Chellas version and the Belnap version) implies that (at a certain moment in a certain model) $D_i\varphi$ is true only if φ is; which seems to mean that the door is closed when the agent closes it. But why close a door that is already closed? On a somewhat related point, notice that, with the truth-condition of D_i as defined, there may be a model such that both $(h, t_1) \vDash D_i\varphi$ and $(h, t_2) \vDash D_i\varphi$, where t_1 and t_2 are times such that $t_1 < t_2$ and h is a history such that $(h, t) \vDash \varphi$ for all times t between t_1 and t_2. On the official stit-reading, the agent closes the door at (h, t_1) as well as at (h, t_2), never mind that the door is closed in h throughout the interval $[t_1, t_2]$. The awkwardness disappears on the alternative reading, according to which the agent is causally responsible at (h, t_1) and (h, t_2) for the door being closed. (But other naïve questions remain: Where is the action? When did the door closing take place? Or are such questions not appropriate?)

The theory presented by Belnap and his collaborators is the culmination of a long development in modal logic; it surpasses all earlier efforts by its sophistication, power and comprehensiveness. One reservation one might have is perhaps the one hinted at in the preceding paragraph: it is a logic of action without actions. No author in the Anselm-Kanger-Chellas line up through Belnap — Davidson belongs to a different tradition — has countenanced the existence of actions in his logic: action talk, yes; ontology of actions, no.[24] For those who would like a representation not only of *action language* but of *action* there is therefore a reason to continue the search for a logic of action worth the name (without, of course, any guarantee that such a thing will ever be found).

3.1.2 What dynamic logic can offer

One liberating effect of the introduction of dynamic logic was that it finally permitted modal logicians to talk about actions and events (without necessarily knowing exactly what they were talking about). The novelty was the introduction of a syntactic category of terms and a corresponding semantic category of — well, what? Formally, the meaning of a term is what modal logicians know as an accessibility relation. But since terms were introduced to formalise the action of programs, it was natural for dynamic logicians to think of these accessibility relations — which relate points-before to points-after — as event types or action types. In fact, Vaughan Pratt himself, the originator of the modal logic of programs, as dynamic logic was called in the early days, remarked that he saw his theory as a logic of computer action.

It would seem that if it is reasonable to represent real propositions (propositions about

[24] One author to whom this remark does not apply is J. F. Horty. Horty, who works within the stit-tradition, explicitly refers to choice cells as actions and, in his book [159], actually refers to them as "actions". In correspondence he has made the following comment: "These actions are only action tokens, however — individual concrete actions. There is no such thing as the action type of "opening a window", for example. There are individual, concrete openings of individual windows, but nothing to group them together."

the world) as sets of points, then it cannot be totally unreasonable to represent real events (events in the world) as sets of paths. Such representation has obvious technical advantages. For one thing, the Boolean operations of union and intersection and even complement become well-defined, and so do the operations of relative product and transitive closure. But it also makes it possible more directly to address the question concerning the relationship between events and action, a topic much debated by philosophers. One example from one of the most interesting philosophers of action:

> The notion of a human act is related to the notion of an event, *i.e.*, a change in the world. What is the nature of this relationship? It would not be right, I think, to call acts a kind or species of events. An act *is* not a change in the world. But many acts may quite appropriately be described as the bringing about or *effecting* ('at will') of a change. To act is, in a sense, to *interfere* with 'the course of nature'. ... To every act ... there corresponds a change or an event in the world. [171, 36f., 39]

One wonders why there should be this difference in theoretical status between propositions and events. Given a certain context, propositions may be true of false; but events may occur or not. A proposition perhaps may be made true or made false; but an event may be brought about or avoided. A proposition may be known or be believed or be given a certain normative status by someone competent to do so; but an event can be foreseen or be remembered, be prescribed or proscribed. One may perceive that a proposition is true; but one may also perceive an event. Some authors such as van Benthem have recognised the analogous position of the two categories, propositions and events, and have tried to give them an even-handed treatment. Unfortunately, philosophers have remained unimpressed so far.

3.1.3 Thinking about change

There already exist modal logical modellings containing interpretations of events, not so much in the philosophical as in computer science literature, [149, 163]. In general, we believe that philosophers have much to learn from the theoretical computer scientists, whose assault on conceptual problems is often fresh and undaunted ("never mind what Aristotle said"). But how are we to make philosophical sense of their constructions and avoid *ad hoc*-ishness?

There is an environment, also referred to as *the world*. The world is always in some (total) *state* or other. The states themselves never change, but the state in which the world is (the currently *actual* state) may change from time to time. The way the world changes is influenced, but perhaps not completely determined, by *agents* outside the world. Furthermore, all change is regular: it takes place according to some *change rule*.

Trying to incorporate these ideas into the semantics of traditional modal logic, think of a *system* as a triple (U, A, C), where U represents the world, A is the set of all agents (assumed to be finitely many), and C is a function representing the change rule. Of these three primitives, two are old: the universe U of *points* (representing total world states) and the set A of agents are as above (for simplicity, we shall think of the agents as a set of integers $\{0, 1, \ldots, n-1\}$, where n is nonnegative (if $n = 0$, then A is empty and the system is agentless)). To describe the change function, the one new primitive, is more complicated. The world is always in some total world state or other. Change in the

world (which for simplicity's sake is assumed to be discrete) is completely described by the change function. Any finite change is from a point-before to a point-after, perhaps via a number of intermediate points; an infinite change can be described in the same way except that there is no point-after. Suppose the world is in a state $u \in U$. Suppose the agents, at this point, make individual "contributions" or "inputs" $i_0, \ldots, i_{n-1}$, respectively, and that there are no further contributions from the agents, nor any other kind of interference. The result will be a change in the world (one possible change is of course the null change), which we may represent as a sequence $p = C(u, i_0, \ldots, i_{n-1})$ of points of which u is the first element; we say that p is the *theoretical result* of the input $(i_0, \ldots, i_{n-1})$. Points to be noted. (1) C depends on the currently actual state of the world but not on any other possible state. (2) The nature of the agents' inputs is not specified (in this modelling). But (3) they are assumed to be outside (not a part of) the world. (4) Since the change rule is represented by a given function, the system as described is deterministic. (It would be nondeterministic if the change rule were represented by a function assigning to each $n+1$-tuple $(u, i_0, \ldots, i_{n-1})$, not a path, but a set of paths. Such indeterminacy would be called *ontic*: not due to limited knowledge on our part, but a property of the system itself.) (5) It is possible for an agent to make no proper contribution. When this happens we say, for book-keeping purposes, that his input is the *null input,* and we use the symbol 0 to denote it.

3.1.4 What a modal logic of actions and agency might look like

Suppose we (as outside observers) witness a certain development taking place in a system. What would it be to have a record of it? Any development in the world consists in the succession of one state after the other, therefore to know the sequence of states, in the order they were realised, would be to have a representation of what took place. But knowing the inputs of the agents would yield a fuller understanding. So perhaps we should think of a record as a certain sequence of elements of type $(u, i_0, \ldots, i_{n-1})$. (There are a number of technical details that would need to be addressed in a rigorous exposition. Let us mention one: if $(u, i_0, \ldots, i_{n-1})$ and $(u', i'_0, \ldots, i'_{n-1})$ are consecutive elements of a history, then u' must be the second element of the path $C(u, i_0, \ldots, i_{n-1})$.)

Technically we may think of a record, from now on more often called a *history*, as a function that assigns to each proper subpath q of a certain path p an n-tuple $(i_0, \ldots, i_{n-1})$; we refer to $(i_0, \ldots, i_{n-1})$ as *the agents' input after q*. The path p is called the *trace* of h; in symbols, $tr(h) = p$. Compared to the "thin" histories of classical modal logic, ours are a lot "fatter", but there is an obvious connexion: the trace of a fat history is a thin history.

As for *events*, we think of them as sets of finite paths (a *path* in U is a sequence of points in U). If p is a path in e (that is, if $p \in e$), then we may say that p *realises* e. So if we witness p played out before our eyes, we also witness a realisation of e. Given this terminology, what would it mean to say that the agents bring about or realise a certain event e? This is a question to which at the present time no-one seems to have an answer. One source of difficulty is the complexity of the many-agent case. The effects of the input of one agent can be modified or completely altered if there are inputs of other agents, either at the same time or later. Agents also sometimes change their mind, thus modifying or altering the effects of their own earlier inputs. For these and other reasons it is often difficult, not only in abstract theory but also in real life, to determine agency

and to allocate (causal) responsibility.

Let us examine one particularly simple case. We say that a history h with trace p is *simple* if $p = C(u, i_0, \ldots, i_{n-1})$, where $i_a \neq 0$ for at least some $a < n$, and, furthermore, $h(q) = (0, \ldots, 0)$, for all $q < p$ such that $q \neq \varnothing$. Thus a simple history is one where the trace is the theoretical result of the initial input. In this very special case, it makes sense to make comments such as the following: the action of the agents was allowed to run its course; the agents brought about the path p; the agents' action realised all events e for which there are paths p', p'' and q such that $q \in e$ and $p = p'qp''$; the agents are (causally) responsible for every event realised by their action.

The one-agent case is of special interest. In this case, we treat the other agents, if any, as part of the background. Technically, if a is an agent, then define the *a-reduct of C* as the function C^a, which assigns to each pair (*u*,i) the set

$$C^a(u, i) = \{p : \exists i_0, \ldots, i_{n-1}(c_a = i \ \& \ p = C(u, i_0, \ldots, i_{n-1}))\}.$$

In general, $C^a(u, i)$ is not a singleton set, so in this case we face a certain indeterminacy. Note, however, that the latter may be said to be *epistemic* in kind, in contrast to the possible ontic indeterminacy discussed above.

So far we have left open the question about the nature of the agents' input. In this particular case, however, it would be tempting to think that the input i of the agent a consists in calling up a program or plan or, as it was termed in Segerberg [164], routine $\mathbf{r}$ such that, if $\mathbf{r}$ is started with the world in the state u, then the paths in $C^a(u, i)$ correspond to possible developments according to $\mathbf{r}$ (computations according to $\mathbf{r}$, if $\mathbf{r}$ is a program). A mathematician might even go as far as identifying the input, the routine and the corresponding reduct: $\mathbf{r} = C^a(u, \mathbf{r})$.

Summarising this discussion: there are three entities that should not be confused: the routine $\mathbf{r}$; the event of running $\mathbf{r}$; the result of running $\mathbf{r}$ (on a particular occasion or in general).

Carrying out the routine $\mathbf{r}$ is, in a sense, what the agent "really does". In the case of individual physical human action — the case that dominates analytical philosophical discussion of action — an agent's routines may be identified with his ways of moving parts of his body:

> If we interpret the idea of a bodily movement generously, a case can be made for saying that all primitive actions are bodily movements. [...] We never do more than move our bodies: the rest is up to nature. [154, 49 and 50]

But of course by our bodily movements we accomplish many other actions. If we introduce a distinction (not honoured by ordinary language) between *doing* and *realising* an action, we might reserve the former locution for what the agent "really does" or "does directly", and the latter for what the agent may accomplish by his action or "does indirectly". In a modal language we might accordingly introduce event operators does_i, done_i and $\mathsf{realises}_i$, $\mathsf{realised}_i$ and contemplate appropriate meaning-conditions for them. There are no direct counterparts in natural language to these operators, but we have something like the following in mind. Let e be the event that is the interpretation of α

and a the agent assigned to i:

$\mathsf{does}_i\alpha$: a is just about to do e,
$\mathsf{done}_i\alpha$: a has just finished doing e,
$\mathsf{realises}_i\alpha$: because of a's action, e is just about to be realised,
$\mathsf{realised}_i\alpha$: because of a's action, e has just been realised.

It is a challenge as yet unmet to give fruitful conditions for the latter two operators, which seems like the more important pair. The challenge will not be met in this paper either, but it may be instructive to see what the difficulty is.

Let us say that a history is *well-behaved* if it is the sequence of simple histories. Furthermore, let us say that (h, g) is an *articulated history* if the last element of the trace of h is also the first element of the trace of g; call the latter element the virtual present of (h, g). Note that it is natural to think of h as the past history up to the virtual present moment of (h, g) and g as a possible future history from there on. The following semi-formal meaning-conditions summarise the remarks above concerning the one-agent case. Assume that (h, g) is an articulated history and that both h and g are well-behaved. We assume that a is the agent assigned to i, that α is an event term and that the interpretation of an event term α is an event $[\![\alpha]\!]$:

$(h, g) \vDash \mathsf{does}_i\alpha$ iff g begins with an initial simple subhistory g' such that $C^a(u, i) = [\![\alpha]\!]$ and $tr(g') \in [\![\alpha]\!]$, where $i \neq \varnothing$ is a's contribution after h,
$(h, g) \vDash \mathsf{done}_i\alpha$ iff h ends with an terminal simple subhistory h' such that $C^a(v, i) = [\![\alpha]\!]$ and $tr(h') \in [\![\alpha]\!]$, where $h = h''h'$, for some history h'', and v is the first element of $tr(h')$, and $i \neq \varnothing$ is a's contribution after h''.

For the other pair one might try conditions such as these:

$(h, g) \vDash \mathsf{realises}_i\alpha$ iff $\exists g', g'', g_0, g_1, g_2, p(g = g'g'' \;\&\; g' \text{ is simple} \;\&\; g' = g_0g_1g_2 \;\&\; p = tr(g_1) \;\&\; p \in [\![\alpha]\!])$,
$(h, g) \vDash \mathsf{realised}_i\alpha$ iff $\exists h_0, h_1, h_2, g', g'', p(h = h_0h_1h_2 \;\&\; g = g'g'' \;\&\; h_1h_2g' \text{ is simple} \;\&\; p = tr(h_2) \;\&\; p \in [\![\alpha]\!])$

or

$(h, g) \vDash \mathsf{realises}_i\alpha$ iff $\exists h', h'', g_0, g_1, g_2, u, p, i(h = h'h'' \;\&\; g = g_0g_1g_2 \;\&\; h''g_0g_1 \text{ is simple} \;\&\; p = tr(g_0) \;\&\; p \in [\![\alpha]\!] \;\&\; u \text{ is first in } tr(h'') \;\&\; tr(h''g') = C^a(u, i) \;\&\; \forall q(q \in C^a(u, i) \Rightarrow \exists q_0, q_1, q_2(q = q_0q_1q_2 \;\&\; q_1 \in [\![\alpha]\!])))$,
$(h, g) \vDash \mathsf{realised}_i\alpha$ iff $\exists h_0, h_1, h_2, g', g'', u, p, i(h = h_0, h_1, h_2 \;\&\; g = g'g'' \;\&\; h_1h_2g' \text{ is simple} \;\&\; p = tr(h_2) \;\&\; p \in [\![\alpha]\!] \;\&\; u \text{ is first in } tr(h_1) \;\&\; tr(h_1h_2g) = C^a(u, i) \;\&\; \forall q(q \in C^a(u, i) \Rightarrow \exists q_0, q_1, q_2(q = q_0q_1q_2 \;\&\; q_1 \in [\![\alpha]\!])))$.

Call the former pair of conditions the weak definition and the latter pair the strong definition of action realisation. The weak definition is probably too wide, and the strong definition probably too narrow, for either properly to reflect an intuitive understanding of an action or event being realised. Another complication is that in daily life there is a tendency to consider that an agent has realised an action (whether he intended it or not) if and only if we hold him causally responsible for it; that is, if we can attribute agency to him. Insofar as the attribution of agency is normative in this sense, it is beyond our simple modelling.

3.2 Deontic logic

3.2.1 Background

Deontic logic is the formal study of normative concepts. Here we shall concentrate on the concepts 'obligatory' and 'ought'. There are many 'oughts', and it is well to keep them apart. In particular, we wish to call attention to three distinctions. One that is particularly relevant to this paper is that between 'ought-to-be' (*Seinsollen*) and 'ought-to-do' (*Tunsollen*): 'oughts' that apply to the state of the world, and 'oughts' that apply to actions. That they really are different notions that require different logics was argued particularly forcefully by Castañeda. The distinction itself is older and often associated with Meinong, who seems to have held that, even though they are different concepts, *Tunsollen* is in the final analysis logically reducible to *Seinsollen*. This view was endorsed by Chisholm who gave it a more precise formulation: in his view — dubbed the Meinong/Chisholm thesis by Horty —

(1) it ought to be that i brings it about that φ,

(2) i ought to bring it about that φ,

are logically equivalent [152, 159]. Actually, this thesis involves also another important distinction: that between the *personal* and the *impersonal*. This is seen more clearly if (2) is rephrased as "it is obligatory for i to bring it about that φ" or even "i has an obligation to bring it about that φ". So in a discussion of the Meinong/Chisholm thesis there are actually two distinctions to bear in mind.

Yet another important distinction is that between *standing* versus *one-time* notions of deontic concepts. A standing obligation (permission, prohibition) has a certain scope and covers everything in that scope, while a one-time obligation (permission, prohibition) concerns one particular item (event, occasion, alternative, possibility, or what not). The two kinds of concepts differ in respect to performance of the actions involved. For example, while a one-time obligation is discharged when one performs the particular action it concerns, a standing obligation can be violated but never, within it scope, completely discharged.

To bring analytical order to this field, von Wright 1963 introduced deontic logic. Relatively quickly Standard Deontic Logic emerged, simply classical logic with an extra propositional operator $\mathbf{O}$ with the informal reading of $\mathbf{O}\varphi$ as "it is obligatory that φ" or "it ought to be the case that φ" and satisfying certain extra postulates: all instances of the schemata $\mathbf{O}(\varphi \wedge \psi) \leftrightarrow (\mathbf{O}\varphi \wedge \mathbf{O}\psi)$ and $\mathbf{O}\varphi \rightarrow \neg\mathbf{O}\neg\varphi$ are axioms, and the rule of replacement of provable equivalents holds. Some authors would also include further axiom schemata, for example, $\mathbf{O}\bot$ (making the system normal), $\mathbf{O}\varphi \rightarrow \mathbf{OO}\varphi, \mathbf{OO}\varphi \rightarrow \mathbf{O}\varphi$ and $\mathbf{O}(\mathbf{O}\varphi \rightarrow \varphi)$. (For definitions and criticisms of SDL, see [156, 153].)

Needless to say, this simple theory — which was a great step forward at the time — was unable to deal with the barrage of counterexamples and conundrums posed by moral philosophers. Some have concluded that the dream of an adequate formal deontic logic is a chimæra, other have looked for ways in which to increase the expressiveness of that very primitive object language. In particular, some authors, including von Wright himself, decided that deontic logic needs a logic of action as a base.

3.2.2 Deontic logic within dynamic logic?

Someone who accepts dynamic logic as a logic of action could reasonably try the simple device of directly adding deontic operators. The latter would of course apply to terms, not formulæ. For example, let us write $\mathsf{ob}_a\alpha, \mathsf{pm}_a\alpha$, and $\mathsf{fb}_a\alpha$ for "α is obligatory for a", "α is permitted for a" and "α is forbidden for a", respectively. In order to define these notions, one might resort to a well-known device going back to Stig Kanger and Alan Ross Anderson, independently of one another, of introducing a constant OK (for approval or absence of a sanction) or a constant S (for disapproval or presence of a sanction); the two approaches are equivalent on the assumption that the formula $\mathsf{S} \leftrightarrow \neg\mathsf{OK}$ is valid. [161, 141, 142] or permission and prohibition this would seem to provide a start, at least initially. In fact, two possibilities come to mind:

(1) $\mathsf{pm}_a\alpha \leftrightarrow [\alpha]\mathsf{OK}$ and $\mathsf{fb}_a\alpha \leftrightarrow \neg[\alpha]\mathsf{OK}$,

(2) $\mathsf{pm}_a\alpha \leftrightarrow \langle\alpha\rangle\mathsf{OK}$ and $\mathsf{fb}_a\alpha \leftrightarrow \neg\langle\alpha\rangle\mathsf{OK}$.

Both have a certain plausibility. Alternative (1) is in the spirit of so-called free-choice permission: permission implies that any outcome of doing the permitted will meet with approval. Alternative (2) is more insidious: if the agent has permission to do something there may nevertheless be outcomes of exercising the permission that will incur the sanction. It seems we do have concepts of permission with these features, say, strong permission and weak permission. There are analogous remarks about prohibition; in either case, the formula $\mathsf{fb}_a\alpha \leftrightarrow \neg\mathsf{pm}_a\alpha$, is valid. So far, so good. But for obligation there is a problem: how to express it? In Standard Deontic Logic, φ is obligatory if and only if the negation of φ is not forbidden. So perhaps one might try $\mathsf{ob}_a\alpha \leftrightarrow \neg\mathsf{pm}_a(-\alpha)$, where pm is one of the alternative operators above and $-\alpha$ is "the complement of α". (See [162] for an effort of this kind; cf. [146].) However, there are difficulties with this approach, which seem hard to overcome. The main difficulty is perhaps that, although the notion of the complement of α can be given a precise meaning in the formal semantics, it does not agree well with intuitive notions. If events are binary relations in a set U, then the complement $U \times U - e$ of an event e is of course again a binary relation. But in general the complement may not be recognised as an intuitively well-defined event corresponding to that set-theoretical entity. It is also worth noticing that sanctions and absence of sanctions may apply not to points but to paths: not so much to *what* is done as *how* it is done. It may be all right to drive from one place to another, but if you do so by going in the wrong direction on a one-way street you may find yourself in trouble. Again, one would wish for a more general analysis.

3.2.3 Norms, norm systems and norm functions

There are norms of different kinds. Every time a mode of behaviour is prescribed or proscribed, approved or disapproved, a norm or a norm system is created. Not only do we have moral and legal codes of varying complexity, but in general all standards of behaviour set norms. 'Etiquette', 'decorum', '*savoir-faire*', '*comme-il-faut*' and 'tasteful' exemplify concepts that are meaningful only in relation to some norm system. The norm systems we meet in daily life are usually neither exact nor complete. For any complex norm system, we need experts, pundits, arbiters, judges, connoisseurs or some such authority to implement it. The ten commandments and the Golden Rule form the

basis of a (religiously founded) morality, but we need theologians to explain what they mean, and ministers to tell us how to apply them. The law attempts to give rules for assessing any possible situation that may come up, but lawyers often disagree about what the law says in a particular case; in many countries, even the Supreme Court decides its issues by vote.

What would it be to have a complete norm system? Consider a given, maximal history. If the norm system, call it the Norm, is complete and we (the analysts) have a full understanding of it, then it should be possible for us, at least in principle, to examine the history, from beginning to end, and see whether at any stage there has been a violation of the Norm. If there has, then paint the history red. Otherwise, ask if there is some over all respect in which the history fails to comply with the Norm. If there is, then paint the history yellow. Finally, if after all this the history is not painted either red or yellow, then paint it green. The set of green histories could be called "legal" if the Norm is legal, "moral" if the Norm is moral, "politically correct" if the Norm is political correctness, and so on. Here, to use a neutral, expression, we shall call the green histories *normal*. At this point, we shall not make a distinction between yellow and red but simply call all histories of that colour *non-normal*.

Strictly speaking, it is not enough that a norm system can partition the set of maximal histories into normal and non-normal; for any past history, the set of future histories must be similarly partitioned. For in general — unless the Norm is totally unforgiving or recognises the possibility of so-called tragic dilemmas or, at the other extreme, is totally tolerant or permissive — any past will admit of possible futures that are red or yellow or green in the sense just described.

In order formally to represent a norm system in this sense — there could of course be several, but we shall be dealing with only one — we now introduce the concept of a norm function. Consider the model theory outlined in the section on the logic of action. Let U and T be as in subsection 3.1.1. A (*total*) *T-history* is a function from T to U; a partial function from T to U is *a partial T-history*. A *past* is a partial history h such that $hg \in H$, for some history g. By the same token, a *future* is a history g such that $hg \in H$ for from some history h. If h is a past, we write $cont(h)$ and $cont^{\circ}(h)$ for the set of all *complete continuations* of h in H and the set of all *incomplete continuations* of h in H, respectively; in symbols, $cont(h) = \{g : hg \in H\}$ and $cont^{\circ}(h) = \{g_0 : \exists g_1(hg_0g_1 \in H)\}$. Now, a *norm function* is a function N from the set of all possible pasts to the set of subsets of all possible futures such that, for every possible past $h, N(h) \subseteq cont(h)$. If $g \in N(h)$ and $g = g'g''$, then we say that g' is a *normal continuation of h and that g* is a *complete normal continuation of h*.

We end this subsection by noting a number of modal operators than can be introduced with truth conditions in the modelling of the previous section. First, there are three box operators $[x]$ and corresponding dual diamond operators $\langle x \rangle$, where x is H ("historically"), D ("deontically") or F ("future"):

$(h, g) \vDash [H]\varphi$ iff $\forall g'(g' \in cont(h) \Rightarrow (h, g') \vDash \varphi)$.
$(h, g) \vDash [\mathrm{D}]\varphi$ iff $\forall g'(g' \in norm(h) \Rightarrow (h, g') \vDash \varphi)$.
$(h, g) \vDash [\mathrm{F}]\varphi$ iff $\forall h', g'((hg = h'g' \ \& \ \exists f(f \neq \varnothing \ \& \ h' = hf)) \Rightarrow (h', g') \vDash \varphi)$.

[H] is the operator called "historical necessity" by Chellas and "unavoidability" by Thomason. [D] is a deontic operator that should not be automatically translated as "it is obligatory that"; if a reading other than the literal "for every normal continuation" is insisted

on, we recommend "ideally", but care has to be taken not to read too much into that word. [F] and ⟨F⟩ are Prior's operators G and F, respectively. We are of course able to help us to all the usual temporal operators, including Kamp's UNTIL:

$$(h, g) \vDash (\text{UNTIL}\varphi)\chi \text{ iff either } \forall g', g''(g = g'g'' \Rightarrow (hg', g'') \nvDash \varphi)$$
$$\text{or else } \exists g_1, g_2(g = g_1g_2 \;\&\; (hg_1, g_2)\varphi \;\&\; (1) \;\&\; (2))$$

where (1) and (2) are the following conditions:

(1) $\forall g', g''((g = g'g'' \;\&\; (hg', g'') \vDash \varphi) \Rightarrow \exists g^*(g' = g_1g^*))$,

(2) $\forall k, k'((g = kk' \;\&\; \exists k^*(g_1 = kk^*)) \Rightarrow (hk, k') \vDash \chi))$.

3.2.4 A fragment of dynamic deontic logic

Deontic logic and doxastic logic are often thought to be formally quite similar. Nevertheless, to develop a dynamic deontic logic (DΔL) as a counterpart to the dynamic doxastic logic (DDL) outlined in the previous section would require much effort. Here we shall be content to offer one example of what $D\Delta L$ might look like by considering how an operator of the personal, one-time, ought-to-do type might be definable in our framework.

In other words, what would a meaning-condition for $\mathsf{ob}_i\alpha$ look like if it is to carry something like the intuitive meaning of "e is obligatory for a" or "it is obligatory for a to see to it that e is done" (where e is an event or action and a is an agent)? A careful explication in natural language might run as follows (*first formulation*): "As long as you haven't done whatever it is that you are obligated to do, you are still supposed to do it (if the obligation has not, for some reason, lapsed), never mind violations of the norm that may have taken place in the past; but when you then do it in a normal way, you thereby discharge that particular obligation." Dressing this vernacular suggestion in semi-technical language may give this result (*second formulation*): an event e is *one-time-obligatory for* agent i, given the past history h, if and only if, *if* at the end of any continuation f_0 of h the event e has not yet been realised by i, *then* (i) e is realised in every normal continuation of hf_0, and (ii) if k_0 is an incomplete normal continuation of hf_0 in which e has been realised, then there is a normal continuation of hf_0k_0 in which e is never again realised.

To make this semi-technical version a notch more formal, counterfactually assume that we possess a definition of action realisation (remember that we were not quite able to work one out in the section on action logic). The semi-formal formulation above may the be replaced by the following formal definition (*third formulation*):

$$\begin{aligned}(h, g) \vDash \quad & \mathsf{ob}_i\alpha \text{ iff } \forall f \in \mathit{cont}(h)\forall f_0, f_1((f = f_0f_1 \;\&\; (hf_0, f_1) \nvDash \mathsf{realised}_i\alpha) \Rightarrow \\ & (\forall k \in \mathit{norm}(hf_0)\exists k_0, k_1(k = k_0k_1 \;\&\; (hf_0k_0, k_1) \vDash \mathsf{realised}_i\alpha \;\& \\ & \exists l \in \mathit{norm}(hf_0k_0)\forall l_0, l_1(l = l_0l_1 \Rightarrow (hf_0k_0l_0, l_1) \nvDash \mathsf{realised}_i\alpha))).\end{aligned}$$

The final, fully syntactic version of our definition of personal one-time ought-to-do obligation is the following valid schema (*fourth formulation*):

$$\mathsf{ob}_i\alpha \leftrightarrow [\text{H}](\text{UNTIL}\mathsf{realised}_i\alpha)[\text{D}]\langle\text{F}\rangle(\mathsf{realised}_i\alpha \wedge \langle\text{D}\rangle[\text{F}]\neg\mathsf{realised}_i\alpha).$$

The ob-operator is of course only one of a number of operators explicating a notion of obligatoriness (cf. [169]). Complicated as it is, it nevertheless neglects at least one

important aspect, namely, what may be called the Problem of the Implicit Dead-line: the time within which a one-time obligation should be discharged is often not explicit, but there may still be some time limit that is tacitly understood. (See, for example, [155]).

3.2.5 Normative positions

The modelling presented in the last two sections is limited in numerous ways. Some of them would take much effort to overcome. However, there is one particular shortcoming that is philosophically important and deserves a comment here. Norm systems are usually systematic. If one past history is a (not necessarily normal) continuation of another, then one would expect the normative situation after the former to be in some intimate sense related to the normative situation after the latter: a norm that does not possess a certain minimum of coherence will not be viable. In order to give an example of a possible coherence criterion, let us first augment the modelling of the previous section.

Describing the AGM paradigm above, we introduced sphere systems to model belief states (see subsection 2.2.1). Now we shall use sphere systems to model what we will call the "normative position". Thus we redefine the concept of a *norm function* to be a function N from the set of all possible pasts such that, for every possible past h, $N(h)$ is a sphere system in $cont(h)$ (meaning that $X \subseteq cont(h)$, for every element $X \in N(h)$). It is the sphere system $N(h)$ that we term the *normative position after* h.

We adopt the following technical definitions. If h is a past history, g an incomplete continuation of h, and X is a set of complete continuations of h, then we write X^g for the set of continuations of hg that are final segments of elements of X. Schematically, if h is a past history and $g \in cont^\circ(h)$, then for all $X \subseteq cont(h)$,

$$X^g =_{df} \{f : f \in cont(hg) \ \& \ gf \in X\}.$$

Similarly, if S is a set of subsets of $cont(h)$, then we write S^g for the set of nonempty subsets Y of $cont(hg)$ such that $Y = X^g$, for some $X \in H$. Schematically,

$$S^g = \{X^g : X \in H \ \& \ X^g \neq \varnothing\}.$$

Note that S^g is a sphere system in $cont(hg)$ if S is a sphere system in $cont(h)$.

We are now ready for the definition that is the point of this exercise: we define a norm function as *coherent* if, for all finite continuations g of any past history h,

$$N(hg) = (N(h))^g.$$

Technically, the change from $N(h)$ to $N(hg)$ is related to the notion of irrevocable change discussed in a context of belief revision in [166]. (It is worth recalling that the original concern of Carlos Alchourrón — professor of jurisprudence and one of the fathers of AGM — was, not belief change, but legal change.)

3.2.6 Moral

We have included the non-standard material found in the last few sections — no doubt trying the reader's patience to the limit — in the hope of driving home four theses:

- as first argued by von Wright, deontic logic depends on the logic of action,

- as argued by Castañeda, for logical analysis propositions (formulæ) may not be enough — we also need something like actions (terms),
- the logic of even common normative concepts is more complex than is usually thought,
- modal logic is well equipped to deal with that particular kind of complexity.

ACKNOWLEDGEMENTS

In writing this paper we have benefited from the advice of a select group of colleagues: Joseph Almog, Robert Demolombe, Vincent Hendricks, Jeff Horty. Most of their advice was gratefully followed. The reason for not following all of it — in particular not the most radical suggestion, to go back to square one and write a completely new article — is a familiar one: not enough time and not enough space. We are indebted to them for all their careful, extensive and, on the whole, encouraging comments. Lindström's work was supported by The Bank of Sweden Tercentenary Foundation via the research project "Medvetande, Materialism och Möjlighet" ("Mind, Materialism, and Modality").

BIBLIOGRAPHY

The following bibliography contains not only references in the narrow sense of being explicitly referred to in the main text but also a number of works that it has seemed important to include as a help to those readers who might wish to pursue interests that the authors hope the reading of this essay has stimulated.

References to part 1

[1] J. Almog. Naming without necessity, *The Journal of Philosophy* 83, 210–242, 1986.
[2] A. Anderson. Some new axioms for the logic of sense and denotation: Alternative (0), *Noûs* 14, 217–234, 1980.
[3] A. Anderson. Alonzo Church's contribution to philosophy and intensional logic, *The Bulletin of Symbolic Logic* 4, 129–171, 1998.
[4] R. Ballarin. Validity and necessity, *Journal of Philosophical Logic*, to appear..
[5] R. Barcan Marcus. A functional calculus of first order based on strict implication', *The Journal of Symbolic Logic* 11, 1–16, 1946.
[6] R. Barcan Marcus. The deduction theorem in a functional calculus of first order based on strict implication, *The Journal of Symbolic Logic* 11, 115–118, 1956.
[7] R. Barcan Marcus. The identity of individuals in a strict functional calculus of second order, *The Journal of Symbolic Logic* 12, 12–15, 1947.
[8] R. Barcan Marcus. Essentialism in modal logic, *Noûs* 1, 90–96, 1967. Reprinted in Barcan Marcus [9].
[9] R. Barcan Marcus. *Modalities: Philosophical Essays*, Oxford: Oxford University Press, 1993.
[10] J. van Benthem and A. ter Meulen, eds. *Handbook of Logic and Language*, Amsterdam: Elsevier, 1997.
[11] E. W. Beth. Semantic entailment and formal derivability, *Mededelingen van de Koninklijke Nederlandse Akademie van Wetenschappen, Afdeling Letterkunde*, n.s., Vol.18, no. 13. Amsterdam: 1955, 309–42. Reprinted in Hintikka [43, 9–41].
[12] P. Blackburn, M. de Rijke and Y. Venema. *Modal logic*, Cambridge Tracts in Computer Science 53, Cambridge: Cambridge University Press, 2001.
[13] G. Boolos. *The Logic of Provability*, Cambridge: Cambridge University Press, 1993.
[14] J. P. Burgess. Quinus ab Omni Naevo Vindicatus, in Ali Kazmi (ed.) *Meaning and Reference*, *Canadian Journal of Philosophy*, Supplementary Volume 23, 1997.
[15] J. P. Burgess. Which modal logic is the right one?, *Notre Dame Journal of Formal Logic* 40, 81–93, 1999.
[16] R. Carnap. Modalities and quantification, *The Journal of Symbolic Logic* 11, 33–64, 1946.

[17] R. Carnap. *Meaning and Necessity: A Study in Semantics and Modal Logic,* Chicago: University of Chicago Press, 1947. Second edition with supplements, 1956.
[18] A. Church. A formulation of the logic of sense and denotation (abstract), *The Journal of Symbolic Logic* 12, 31, 1946.
[19] A. Church. A formulation of the logic of sense and denotation, in *Structure, Method, and Meaning: Essays in honor of H. M. Sheffer,* New York: Liberal Arts Press, 1951.
[20] A. Church. Outline of a revised formulation of the logic of sense and denotation (Part I), *Noûs* 7, 24–33, 1973.
[21] A. Church. Outline of a revised formulation of the logic of sense and denotation (Part II), *Noûs* 8, 135–156, 1974.
[22] A. Church. Russellian simple type theory, in *Proceedings and Addresses of the American Philosophical Association* 47, 21–33, 1974.
[23] N. Cocchiarella. On the primary and secondary semantics of logical necessity, *Journal of Philosophical Logic* 4, 13–27, 1975.
[24] J. Copeland. The genesis of possible worlds semantics, *Journal of Philosophical Logic* 31, 99–137, 2002.
[25] J. Divers. *Possible Worlds*, London: Routledge. 2002.
[26] K. Fine. Modality *de re*, in Almog *et al. (*eds.*) Themes from Kaplan*, Oxford: Oxford University Press, 197–272, 1986.
[27] K. Fine. Quine on quantifying in, in C. A. Anderson and J. Owens (eds.) *Propositional Attitudes*, Stanford: CSLI, 1–25, 1991.
[28] M. Fitting. *Types, Tableaux, and Gödel's God*, Boston: Kluwer, 2002.
[29] M. Fitting. Intensional logic–beyond first-order, in Hendricks, V., Malinowski, J. (eds.), *50 Years of Studia Logica*, Dordrecht, Kluwer, 87–108, 2003.
[30] M. Fitting and R. Mendelsohn. *First-Order Modal Logic*, Dordrecht: Kluwer, 1998.
[31] D. Føllesdal. *Referential Opacity and Modal Logic*, New York, NY: Routledge, 2004. Published version of doctoral dissertation, Harvard University, 1961.
[32] G. Frege. Über Sinn und Bedeutung, original in *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50, 1892. English translation in Geach and Black [34].
[33] D. Gallin. *Intensional and Higher-Order Modal Logic*, Amsterdam: North-Holland, 1975.
[34] P. Geach and M. Black, eds. *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell 1982.
[35] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalküls, *Ergebnisse eines mathematischen Kolloquiums*, 4, 39–40, 1933. Reprinted with an English translation ('An interpretation of intuitionistic propositional calculus') in S. Feferman *et al.* (eds.), *Kurt Gödel, Collected Works* Vol. I, Oxford: Oxford University Press, 1986, 300–303.
[36] R. Goldblatt. Mathematical modal logic: a view of its evolution, *Journal of Applied Logic* 1, 309–392, 2003.
[37] A. Hazen. Counterpart-theoretic semantics for modal logic, *The Journal of Philosophy* 76, 319–338, 1979.
[38] J. Hintikka. Form and content in quantification theory, *Acta Philosophica Fennica*, 8, 7–55, 1955.
[39] J. Hintikka. Quantifiers in deontic logic, Helsinki: *Societas Scientarum Fennica, Commentationes Humanarum Literarum* 23, no. 4, 1957.
[40] J. Hintikka. Modality as referential multiplicity, *Ajatus* 20, 49–64, 1957.
[41] J. Hintikka. Modality and quantification, *Theoria* 27, 117–128, 1961.
[42] J. Hintikka. *Knowledge and belief: An Introduction to the Logic of the Two Notions*. Cornell: Cornell University Press, 1962.
[43] J. Hintikka. *The Philosophy of Mathematics*, Oxford Readings in Philosophy, Oxford: Oxford University Press, 1969.
[44] J. Hintikka. *Models for Modalities, Selected Essays*, Dordrecht: D. Reidel, 1969.
[45] J. Hintikka. *The Intentions of Intentionality and Other New Models for Modalities*, Dordrecht: D. Reidel, 1975.
[46] J. Hintikka. Standard vs. nonstandard logic: Higher-order, modal, and first-order logics, in E. Agazzi (ed.) *Modern Logic*, Dordrecht: D. Reidel, 283–296, 1981.
[47] J. Hintikka. Is alethic modal logic possible? In J. Hintikka and M. B. Hintikka: *The Logic of Epistemology and the Epistemology of Logic*. Dordrecht: D. Reidel, 1989.
[48] J. Hintikka and G. Sandu. The fallacies of the new theory of reference, *Synthese* 104, 245–283, 1995.
[49] G. Holmström-Hintikka, S. Lindström, and R. Sliwinski, eds. *Collected Papers of Stig Kanger with Essays on his Life and Work*, Volume I, Synthese Library Vol. 303, Dordrecht: Kluwer, 2001.
[50] G. Holmström-Hintikka, S. Lindström, and R. Sliwinski, eds. *Collected Papers of Stig Kanger with Essays on his Life and Work*, Volume II, Synthese Library Vol. 304, Dordrecht: Kluwer, 2001.
[51] L. Humberstone. Two-dimensional adventures, *Philosophical Studies* 118, 17–65, 2004.

[52] B. Jónsson and A. Tarski. Boolean algebras with operators, *American Journal of Mathematics* 73 (1951), 891–939, vol. 74 (1952), 127–162, 1951.
[53] S. Kanger. *Provability in Logic,* Acta Universitatis Stockholmiensis, *Stockholm Studies in Philosophy* 1, Stockholm: Almqvist & Wiksell, 1957. Reprinted in Holmström-Hintikka *et al.*(2001a).
[54] S. Kanger. The morning star paradox, *Theoria* 23, 1–11, 1957. Reprinted in Holmström-Hintikka *et al.*(2001a).
[55] S. Kanger. A note on quantification and modalities, *Theoria* 23, 133–134, 1957. Reprinted in Holmström-Hintikka *et al.*(2001a).
[56] S. Kanger. On the characterization of modalities, *Theoria* 23, 152–155, 1957. Reprinted in Holmström-Hintikka *et al.*(2001a).
[57] S. Kanger. *New foundations for ethical theory, Part 1*, (mimeographed) Stockholm, 1957, reprinted (with minor changes) in R. Hilpinen (ed.): *Deontic logic: introductory and systematic readings*, Dordrecht: D. Reidel 1971, 36–58. Reprinted in Holmström-Hintikka *et al.*(2001a).
[58] D. Kaplan. *Foundations of Intensional Logic,* dissertation, University of California, Los Angeles, Ann Arbor: University Microfilms International, 1964.
[59] D. Kaplan. Quantifying in, in D. Davidson and J. Hintikka (eds.), *Words and Objections: Essays on the Work of W. V. Quine,* Dordrecht: D. Reidel, 178–214, 1969.
[60] D. Kaplan. How to Russell a Frege-Church, *The Journal of Philosophy* 72, 716-729, 1975.
[61] D. Kaplan. Opacity, in Hahn and Schilpp (eds.), *The Philosophy of W. V. Quine*, *The Library of Living Philosophers*, Volume XVIII, Open Court, La Salle, Illinois, 1986.
[62] J. Kim. *Supervenience and Mind: Selected Philosphical Essays*, Cambridge and New York: Cambridge University Press, 1993.
[63] J. C. King. Structured Propositions, *The Stanford Encyclopedia of Philosophy (Fall 2001 Edition)*, Edward N. Zalta (ed.), `http://plato.stanford.edu/archives/fall2001/entries/propositions-structured/`
[64] S. Kripke. A completeness theorem in modal logic,*The Journal of Symbolic Logic,* 24, 1–14, 1959.
[65] S. Kripke. Semantical analysis of modal logic (abstract) *The Journal of Symbolic Logic* 24, 323–324, 1959.
[66] S. Kripke. Semantical considerations on modal logic, in *Proceedings of a Colloquium on Modal and Many-Valued Logics,* Helsinki, 23–26 August, 1962. *Acta Philosophica Fennica* 16, 83–94, 1963. Reprinted in Linsky [82].
[67] S. Kripke. Semantical analysis of modal logic: I. Normal Modal Propositional Calculi, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9, 67–96, 1963.
[68] S. Kripke. Semantical analysis of modal logic: II. Non-normal modal propositional calculi, in J. W. Addison, L. Henkin, and A. Tarski (eds.), *The Theory of Models* (Proceedings of the 1963 International Symposium at Berkeley), Amsterdam: North-Holland, 206–220, 1965.
[69] S. Kripke. Identity and necessity, in M. Munitz (ed.), *Identity and Individuation,* New York: New York University Press, 135–164, 1971.
[70] S. Kripke. Naming and necessity, in D. Davidson and G. Harman (eds.) *Semantics of Natural Language*. Dordrecht: D. Reidel, 253-355, and 763–769, 1972. Reprinted (with a new preface) as Kripke [71].
[71] S. Kripke. *Naming and Necessity*, Cambridge, Mass: Harvard University Press, 1980.
[72] C. I. Lewis. Implication and the algebra of logic, *Mind* 21, 522–531, 1912.
[73] C. I. Lewis. *Survey of Symbolic logic*. Berkeley: University of California Press, 1918.
[74] C. I. Lewis and C. H. Langford. *Symbolic Logic*, New York: The Century Company, 1932.
[75] D. Lewis. Counterpart theory and quantified modal logic, *The Journal of Philosophy* 65, 113–126, 1968.
[76] D. Lewis. General Semantics, *Synthese* 22, 18–67. Reprinted in Lewis [79], 1970.
[77] D. Lewis. *Counterfactuals*, London: Blackwell, 1973.
[78] D. Lewis. Causation, *Journal of Philosophy* 70, 556–67, 1973.
[79] D. Lewis. *Philosophical Papers, Volume I*, Oxford: Oxford University Press, 1983.
[80] D. Lewis. *On the Plurality of Worlds*. London: Blackwell, 1986.
[81] S. Lindström. An exposition and development of Kanger's early semantics for modal logic, in P. W. Humphreys and J. H. Fetzer (eds.) *The New Theory of Reference — Kripke, Marcus, and Its Origins.* Dordrecht: Kluwer, 1998. Reprinted with minor changes in Holmström-Hintikka et al. (2001b).
[82] L. Linsky. *Reference and Modality*, Oxford: Oxford University Press, 1971.
[83] R. Montague. Logical necessity, physical necessity, ethics and quantifiers, *Inquiry* 4, 259–269, 1960. Reprinted in [84].
[84] R. Montague. *Formal Philosophy: Selected Papers of Richard Montague*. Edited and with an Introduction by Richmond H. Thomason. New Haven: Yale University Press, 1974.
[85] J. Myhill. Problems arising in the formalization of intensional logic, *Logique et Analyse* 1, 78–83, 1958.

[86] S. Neale. On a milestone of empiricism, in Orenstein, A. and P. Kotatko (eds.) *Knowledge, Language, and Logic*, 237–346, 2000.
[87] D. Nolan. *Topics in the Philosophy of Possible Worlds*, London: Routledge, 2002.
[88] C. Parsons. Intensional Logic in Extensional Language, *The Journal of Symbolic Logic* 47, 289–328, 1982.
[89] T. Parsons. Essentialism and quantified modal logic, *The Philosophical Review* 78, 35–52, 1969. Reprinted in Linsky [82].
[90] B. H. Partee and H. L. W. Hendriks. Montague grammar, in van Benthem and ter Meulen [10].
[91] W. V. Quine. Notes on existence and necessity, *Journal of Philosophy* 40, 113–127, 1943.
[92] W. V. Quine. The problem of interpreting modal logic, *The Journal of Symbolic Logic* 12, 43–48, 1947.
[93] W. V. Quine. Reference and modality, in *From a Logical Point of View,* Cambridge, Mass.: Harvard University Press, 1953.
[94] W. V. Quine. Three grades of modal involvement, *Proceedings of the XIth International Congress of Philosophy,* vol. 14, North-Holland, Amsterdam, 65–81, 1953. Reprinted in Quine [95].
[95] W. V. Quine. *The Ways of Paradox and Other Essays,* Cambridge, Mass.: Harvard University Press, 1966.
[96] W. V. Quine. Review of *Identity and Individuation*, M. K. Munitz, ed., New York, 1971. *Journal of Philosophy* 69, 488–97, 1972.
[97] N. Salmon. *Reference and Essence*, Oxford: Basil Blackwell, 1982.
[98] N. Salmon. *Frege's puzzle*, Cambridge, Mass.: MIT Press, 1986.
[99] A. Smullyan. Modality and description, *The Journal of Symbolic Logic* 13, 31–37, 1948.
[100] S. Soames. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*, Oxford: Oxford University Press, 2002.
[101] R. Solovay. Provability interpretations of modal logic, *Israel Journal of Mathematics* 25, 287–304, 1976.
[102] R. Stalnaker. *Context and Content — Essays on Intentionality in Speech and Thought*, Oxford: Oxford University Press, 1999.
[103] R. Stalnaker. *Ways a World Might Be — Metaphysical and Anti-Metaphysical Essays.* Oxford: Oxford University Press, 2003.
[104] Z. G. Szabó. Compositionality, *The Stanford Encyclopedia of Philosophy* (Fall 2004 Edition), Edward N. Zalta (ed.), 2004. `http://plato.stanford.edu/archives/fall2004/entries/compositionality/`
[105] A. Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen, *Studia Philosophica* 1, 261–405, 1936. (English translation: 'The concept of truth in formalized languages', 152–278 in *Logic, Semantics, Metamathematics*, second edition, Hackett Indianapolis, 1983).
[106] A. Tarski. Über den Begriff der logischen Folgerung, *Actes du Congrès International de Philosophie Scientifique*, vol. 7, Paris, 1–11, 1936. (English translation: 'On the concept of logical consequence', 409–420 in *Logic, Semantics, Metamathematics*, second edition, Hackett, Indianapolis, 1983).
[107] G. H. von Wright. *An Essay in Modal Logic*, Amsterdam: North-Holland, 1951.
[108] E. Zalta. Logical and analytic truths that are not necessary, *The Journal of Philosophy*, Vol 85, 57-74, 1988.

References to part 2

[109] C. Alchourrón, P. Gärdenfors and D. Makinson. On the logic of theory change, *The Journal of Symbolic Logic* 50, 510–530, 1985.
[110] H. Arló-Costa. First order extensions of classical systems of modal logic: The role of the Barcan schemas, *Studia Logica* 71, 87–18, 2002.
[111] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*, Cambridge: MIT Press, 1995.
[112] P. Gärdenfors. Imaging and conditionalization, *The Journal of Philosophy* 79, 747–760, 1982.
[113] P. Gärdenfors. *Knowledge in flux: modeling the dynamics of epistemic states.* Cambridge, Mass.: The MIT Press, 1988.
[114] P. Gochet and P. Gribomont. Epistemic Logic, in *Handbook of the History and Philosophy of Logic*, edited by Gabbay, D.M. and Woods, J. Amsterdam: Elsevier Science, 2005.
[115] G. Grahne. Updates and counterfactuals. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, edited by J. A. Allen, R. Fikes & E. Sandewall, 269–276. San Mateo, Calif.: Morgan Kaufmann, 1991.
[116] A. Grove. Two modellings for theory choice, *Journal of Philosophical Logic* 17, 157–170, 1988.

[117] S. O. Hansson. *A Textbook of Belief Dynamics: Theory Change and Data Base Updating*, Dordrecht: Kluwer, 1999.
[118] V. F. Hendricks. *Forcing Epistemology*, forthcoming, Cambridge University Press, 2005.
[119] V. F. Hendricks, S. A. Pedersen, and K. F. Jørgensen, eds. *Knowledge Contributors*, Synthese Library Vol. 322. Kluwer Academic Publishers, 2003.
[120] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, NY: Cornell University Press, 1962.
[121] W. van der Hoek, H. Ditmarsch, and B. Kooi. Concurrent Dynamic Epistemic Logic, in (Hendricks *et al.*, 2003)
[122] H. Katsuno and A. O. Mendelzon. On the difference between updating a knowledge base and revising it. In *Belief revision*, edited by P. Gärdenfors, 183–203. Cambridge: Cambridge University Press, 1992.
[123] A. M. Keller and M. Winslett Wilkins. On the use of an extensional relational model to handle changing incomplete information, *I.E.E.E. transactions on software engineering*, SE-11:7, 620–633, 1985.
[124] H. Levesque. All I know, a study in autoepistemic logic, *Artificial Intelligence* 42, 263–309, 1990.
[125] I. Levi. *The Enterprise of Knowledge*, Cambridge: Cambridge University Press, 1983.
[126] I. Levi. *The Fixation of Belief and its Undoing*, Cambridge: Cambridge University Press, 1991.
[127] D. Lewis. *Counterfactuals*. Oxford: Blackwell, 1973.
[128] D. Lewis. Probabilities of conditionals and conditional probabilities, *Philosophical Review* 85, 297–315, 1976.
[129] B. van Linder, W. van der Hoek, and J.-J. Meyer. Actions that make you change your mind. *In Knowledge and belief in philosophy and artificial intelligence*, edited by A. Laux and H. Wansing, 103–136. Berlin: Akademie-Verlag GmbH, 1995.
[130] S. Lindström and W. Rabinowicz. Epistemic entrenchment with incomparabilities and relational belief revision. In *The logic of theory change*, edited by André Fuhrmann and Michael Morreau, 93–126. *Lecture Notes in Artificial Intelligence*, no. 465. Berlin: Springer-Verlag, 1990.
[131] S. Lindström and W. Rabinowicz. Extending dynamic doxastic logic accommodating iterated beliefs and Ramsey conditionals within DDL. In *For Good Measure: Philosophical Essays Dedicated to Jan Odelstad*, edited by Lars Lindahl, Paul Needham and Rysiek Sliwinski, Uppsala: Uppsala Philosophical Studies 46, 126–153, 1997.
[132] S. Lindström and W. Rabinowicz. DDL Unlimited: Dynamic Doxastic Logic for Introspective Agents', *Erkenntnis* 50, 353–385, 1999.
[133] F. P. Ramsey. *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. Edited by D. H. Mellor. London and Henley: Routledge and Kegan Paul, 1978.
[134] M. de Rijke. Meeting some neighbours: A dynamic modal logic meets theories of change and knowledge representation, in Van Eijck, J. and Visser, A. (eds.), *Logic and Information Flow*, Cambridge: MIT Press, 1994.
[135] H. Rott. *Change, choice and inference: a study of belief revision and nonmonotonic reasoning*. Oxford Logic Guides Vol. 42. Oxford: Clarendon Press, 2001.
[136] K. Segerberg. Two traditions in the logic of belief: bringing them together. In *Logic, Language and Reasoning: Essays in Honour of Dov Gabbay*, edited by Hans Jürgen Ohlbach and Uwe Reyle, 135–147. Dordrecht: Kluwer, 1999.
[137] K. Segerberg. The basic dynamic doxastic logic of AGM. In *Frontiers in Belief Revision*, edited by Mary-Anne Williams and Hans Rott, 57–84. Dordrecht: Kluwer, 2001.
[138] K. Segerberg. Moore problems in full dynamic doxastic logic. In *Essays in Logic and Ontology: Dedicated to Jerzy Perzanowski*, edited by J. Malinowski and A. Pietruszczak, 11–25. Poznań Studies in the Philosophy of the Sciences and the Humanities, Vol. 50. Amsterdam/Atlanta, GA: Rodopi, 2003.
[139] R. Sorensen. *Blindspots*, Oxford: Oxford University Press, 1988.
[140] R. Stalnaker. A theory of conditionals. In *Studies in Logical Theory*, *American Philosophical Quarterly*, Monograph 2, edited by N. Rescher, 98–112. Oxford: Blackwell, 1968.

References to part 3

[141] A. R. Anderson. The formal analysis of normative systems. Technical report. New Haven, CT: Office of Naval Research, 1962.
[142] A. R. Anderson. Logic, norms, and roles, *Ratio* 4, 36–39, 1962.
[143] L. Åqvist. *Introduction to Deontic Logic and the Theory of Normative Systems*. Napoli: Bibliopolis, 1987.

[144] N. Belnap, M. Perloff and Ming Xu. *Facing the Future: Agents and Choice in Our Indeterministic World.* With contributions by Paul Bartha, Michell Green and John Horty. New York, NY: Oxford University Press, 2001.
[145] J. van Benthem. *Exploring Logical Dynamics*, CSLI Publications, Stanford, Calif, 1996.
[146] J. M. Broersen. *Modal Action Logic for Reasoning about Reactive Systems.* Doctoral dissertation, Vrije Universiteit van Amsterdam, 2003.
[147] H.-N. Castãneda. *Thinking and Doing.* Dordrecht: D. Reidel, 1975.
[148] H.-N. Castãneda. The paradoxes of deontic logic: the simplest solution to all of them in one fell swoop. In *New Studies in Deontic Logic: Norms, Actions and the Foundations of Ethics*, edited by R. Hilpinen, 37–85. Dordrecht: D. Reidel, 1981.
[149] P. R. Cohen and H. J. Levesque. Intention is choice with commitment, *Artificial intelligence*, 42, 213–261, 1990.
[150] B. F. Chellas. *The Logical Form of Imperatives.* Stanford, CA: Perry Lane Press, 1969.
[151] B. F. Chellas. Conditional obligation. In S. Stenlund (ed.), *Logical Theory and Semantic Analysis*, 23–33, Dordrecht: D. Reidel, 1974.
[152] R. Chisholm. The ethics of requirement, *American Philosophical Quarterly* 1, 147–153, 1964.
[153] S. Danielsson. What shall we do with deontic logic?, *Theoria* 66, 97–114, 2000.
[154] D. Davidson. *Essays on Actions and Events.* Oxford: Clarendon Press, 1980.
[155] R. Demolombe, P. Bretier and V. Louis, manuscript, 'Formalisation de l'obligation de faire avec délais'.
[156] J. Forrester. *Being Good and Being Logical: Philosophical Groundwork for a New Deontic Logic.* Armonk, NY and London: M. E. Sharpe, 1996.
[157] R. Hilpinen. Actions in deontic logic. In *Deontic Logic in Computer Science: Normative System Specification*, edited by J.-J. C. Meyer & R. J. Wieringa, 85–100. New York, N.Y.: John Wiley & Sons, 1993.
[158] J. Hintikka. Quantifiers in deontic logic, Societas Scientiarum Fennica, *Commentationes humanarium litterarum* 23, no. 4, 1957.
[159] J. Horty. *Agency and Deontic Logic*, New York, NY: Oxford University Press, 2001.
[160] G. Holmström-Hintikka, S. Lindström, and R. Sliwinski, eds. *Collected Papers of Stig Kanger with Essays on his Life and Work*, Volume I, Synthese Library Vol. 303, Dordrecht: Kluwer, 2001.
[161] S. Kanger. *New Foundations for Ethical Theory, Part 1*, (mimeographed) Stockholm, 1957, reprinted (with minor changes) in R. Hilpinen (ed.): *Deontic Logic: Introductory and Systematic Readings*, Dordrecht: D. Reidel 1971, 36–58. Reprinted in Holmström-Hintikka *et al.*(2001).
[162] J.-J. Ch. Meyer. A different approach to deontic logic: deontic logic viewed as a variant of dynamic logic, *Notre Dame Journal of Formal Logic* 29, 109–136, 1988.
[163] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning.* San Mateo, CA: Morgan Kaufmann Publishers, 1991.
[164] K. Segerberg. Routines. *Synthese* 65, 185–210, 1985.
[165] K. Segerberg. Getting started: beginnings in the logic of action, *Studia logica* 51, 347–378, 1992.
[166] K. Segerberg. Irrevocable belief revision in dynamic doxastic logic, *Notre Dame Journal of Formal Logic* 39, 287–306, 1998.
[167] K. Segerberg. Outline of a logic of action. In *Advances in Modal Logic*, Volume 3, Frank Wolter, Heinrich Wansing, Maarten de Rijke, and Michael Zakharyaschev, editors, 57–84. Lecture Notes. Stanford, CA: CSLI Publications, 2001.
[168] K. Segerberg. Modellings for two types of action. In *A Philosphical Smorgasbord: Essays on Action, Truth and Other Things*, edited by K. Segerberg and R. Sliwinski, 151–156. Uppsala: Filosofiska institutionen. 2003.
[169] K. Segerberg. Trying to meet Ross's challenge. In *Logic and Philosophy in Italy: Essays in honor of Corrado Mangione*, edited by Edoardo Ballo and Miriam Franchella, pp. 155–166. Monza: Italy: Polimetrica, 2006.
[170] G. H. von Wright. Deontic logic, *Mind* 60, 1–15, 1951.
[171] G. H. von Wright. *Norm and Action: A Logical Inquiry.* London: Routledge and Kegan Paul, 1963.