UDC 16

G. Sandu, Dr. hab., Professor,
T. Tanninen, PhD-student
University of Helsinki, Finland

THE BIRTH OF MODAL LOGIC IN NORTHERN EUROPE

This articles investigates the genesis of modal logic through the works of G. H. von Wright, Jaakko Hintikka, Stig Kanger and Saul Kripke.

Keywords: Jaakko Hintikka, modal logic, history of logic.

Constituents and distributive normal forms

Hintikka learned about constituents and distributive normal forms from the lectures of his teacher, G. H. von Wright. The lectures took place at the University of Helsinki during 1947-1948. We fix a monadic first-order language. From the primitive predicate symbols of the language, one can generate mutually exclusive predicates (Q-predicates) in an obvious way. Thus if we assume that the language possesses only two monadic predicates, M_1 and M_2, we get 4 Q-predicates

$Q_1(x) = M_1(x) \wedge M_2(x)$
$Q_2(x) = M_1(x) \wedge \neg M_2(x)$
$Q_3(x) = \neg M_1(x) \wedge M_2(x)$
$Q_4(x) = \neg M_1(x) \wedge \neg M_2(x)$.

A *constituent* tells us which Q-predicates are instantiated and which ones are empty in an underlying universe of individuals. Thus the logical form of a constituent (with quantifier depth 1) is:

$C = \pm\exists x Q_1(x) \wedge \ldots. \wedge \pm\exists x Q_4(x)$

Constituents are mutually exclusive and each constituent specifies a "possible world". The disjunction of all constituents is called by von Wright a tautology, which, when presented in this way, is said to be in distributive normal form. Von Wright will later on apply constituents to the study of modal logic (von Wright 1951).

Hintikka, 21 years old, set himself the task to extend distributive normal forms to the entire first-order logic with relation symbols. The project resulted in his doctoral dissertation, *Distributive Normal Forms in the Calculus of Predicates*, Hintikka (1953), where Hintikka showed, among other things, that each formula in first-order logic is equivalent to a disjunction of (canonical) constituents. In the particular case in which the sentence is a consistent generalization (quantificational sentence without individual constants), Hintikka showed that it can be expressed as a finite disjunction of constituents (each generalization has a finite quantificational depth.) Hintikka's results are better known to the community from Hintikka (1964).

Constituents and distributive normal forms became the methodological pillar of what later on came to be known as Hintikka's school in inductive logic and philosophy of science, which involved, in addition to Hintikka himself, his students R. Tuomela, R. Hilpinen and I. Niiniluoto. Beginning with (1955), Hintikka developed the tool of model sets and applied them to alethic and epistemic logic. We will survey some of the main results in comparison to similar treatments by G. H. von Wright, Stig Kanger, and Saul Kripke.

Model sets

In Hintikka (1955) the author introduced models sets as a new tool in logical semantics, and constructed a new proof of the completeness of first-order logic. A model set is a set of sentences in the relevant logical language which constitutes a partial description of a possible state of affairs.

One starts with a first-order language L and assumes it has an infinite number of individual constants (actually Hintikka did not use the expression "individual constants, but what he called "free individual variables" or sometimes "free individual symbols". However, Hintikka often emphasizes that free individual variables cover names and other singular terms which purport to refer to well defined objects; see e.g. Knowledge and Belief, p.93.). A model set μ is any set of sentences of L which satisfies some very intuitive closure conditions:

(i) For any atomic sentence A, not both $A \in \mu$ and $\neg A \in \mu$
(ii) If $A \wedge B \in \mu$, then both $A \in \mu$ and $B \in \mu$
(iii) If $A \vee B \in \mu$, then either $A \in \mu$ or $B \in \mu$
(iv) If $\neg\neg A \in \mu$, then $A \in \mu$.
(v) If $\neg(A \wedge B) \in \mu$; then $\neg A \in \mu$ or $\neg B \in \mu$
(vi) If $\neg(A \vee B) \in \mu$; then $\neg A \in \mu$ and $\neg B \in \mu$

The clauses for quantifiers introduce further complications:

(C.E) If $\exists x A \in \mu$, then $A(x/b) \in \mu$ for at least one constant b of L

(C.U) If $\forall x A \in \mu$, and if b occurs in at least one member of μ, the $A(x/b) \in \mu$.

(C. ¬E) If $\neg\exists x A \in \mu$, then $\forall x \neg A \in \mu$.

(C. ¬U) If $\neg\forall x A \in \mu$, then $\exists x \neg A \in \mu$.

Identity requires additional rules:

(C.=) If A is an atomic formula or its negation, and $A \in \mu$ and if B is exactly like A except that a and b have been interchanged in one or several places, then $B \in \mu$.

(C.self≠) For no constant b: $b \neq b \in \mu$.

Sometimes Hintikka prefers the following rule to (C.self≠):

(C.self=) If b occurs in the formulas of μ, then $b = b \in \mu$.

Hintikka's purpose in studying the notion of model set is expressed in the following passage:

The basic notion of a semantic theory is normally the notion of truth. In so far as we are not interested in truth under some *particular* interpretation of logical formulae but rather in the question of whether there are *any* interpretations which make a give set of formulae true (in short, if we are not interested in any one interpretation more than in the others), the basic concept of a semantical theory may also be chosen to be that of *satisfiability*. If the negation of a formula A is not satisfiable, A is said to be valid. (Hintikka, 1961, p. 119.)

Hintikka defines the notion of satisfiability by reference to the Carnapian notion of state-description:

- A set of formulae λ is satisfiable if and only if there is a state description in which all the members of λ are true.

For a single sentence A we say that A is satisfiable if and only if $\{A\}$ is satisfiable. Thus a sentence is satisfiable if and only if A is true in a state description.

A state description, the ancestor of the notion of model set, is a set of formulae which satisfies the following conditions:

(C.1) If A is an atomic sentence (or an identity) then not both $A \in \mu$ and $\neg A \in \mu$

(C.2) If A is an atomic sentence (or an identity) then either $A \in \mu$ or $\neg A \in \mu$

(C.3) If A is an atomic sentence (or an identity) or the negation of an atomic sentence (identity), and if $A \in \mu$ and $a = b \in \mu$ and if B is exactly like A except that a and b have been interchanged in one or several places, then $B \in \mu$.

(C.4) Not $\neg(b = b) \in \mu$.

Thus essentially, state-description is a set of atomic sentences or their negations. In order to understand the above definition of satisfiability, we still need to understand the notion "a state description makes all the members of λ true". One way to proceed, following Hintikka (1961), is to give necessary and suffficient conditions for a set of sentences μ to be the set of *all* sentences which are true in a state description. The set of conditions includes, in addition to (C.1)-(C.4) the following:

(C.5) If $A \wedge B \in \mu$, then both $A \in \mu$ and $B \in \mu$.

(C.6) If both $A \in \mu$ and $B \in \mu$, then $A \wedge B \in \mu$.

(C.7) If $A \vee B \in \mu$, then either $A \in \mu$ or $B \in \mu$.

(C.8) If either $A \in \mu$ or $B \in \mu$ and all the individual constants occurring in $(A \vee B)$ occur in the other formulae of μ, then $A \vee B \in \mu$.

(C.9) If $\exists xA \in \mu$, then $A(x/b) \in \mu$ for at least one constant b

(C.10) If $A(x/b) \in \mu$ for at least one constant b, then $\exists xA \in \mu$.

(C.11) If $\forall xA \in \mu$, and if b occurs in at least one member of μ, then $A(x/b) \in \mu$.

(C.12) If $A(x/b) \in \mu$ for every individual constant b which occurs in the formulae of μ, then $\forall xA \in \mu$.

Thus conditions (C.1)-(C.4) make sure that μ is a state-description, and the other conditions constitute a recursive definition of what it is for a non-atomic sentence to be true in a state description. The clauses for negation are missing, because it is assumed that negation occurs only infront ot atomic sentenced. But if this assumption were dropped, they could be easily added, e.g.

(C.13) If $\neg(A \wedge B) \in \mu$, then either $\neg A \in \mu$ or $\neg B \in \mu$.

(C.14) If either $\neg A \in \mu$ or $\neg B \in \mu$, then $\neg(A \wedge B) \in \mu$, etc.

We shall disregard them in what follows.

Now we return to the above definition of satisfiability of a set of sentences (or formulae) and reformulate it, following Hintikka, as:

- A set of formulae λ is satisfiable if and only if λ can be embedded in a set which satisfies conditions (C.1)-(C.12).

One of Hintikka's basic insights in his early work is the observation that the right-to-left conditions are redundant for his purpose, and among those, condition (C.2) is also redundant. He ends up only with the left-to-right conditions (C.1), (C.3), (C.4), (C.5), (C.7), (C.9) and (C.11) and calls any set which satisfies them a *model-set*. Indeed, we notice that they (together with similar left-to-right conditions for negation) constitute his definition of a model set at the beginning of this section. (Hintikka, 1961.)

Hintikka is then able to prove that a set of formulae is satisfiable if and only if it can be embedded in a model set.

He expresses informally this result in the following way:

The result may perhaps be expressed intuitively by saying that a model set is the formal counterpart of a possible state of affairs (of a 'possible world'.) (It is, however, large enough a description to make sure that the state of affairs in question is really possible.) For it is natural to say that a set of sentences is satisfiable if it can be embedded in a (partial or exhaustive) description of possible states of affairs; and this is just what we demonstrated if model sets are interpreted as such descriptions. (Hintikka, 1961, p. 122.)

Later on Hintikka will extend the notion of satisfiability (consistency) to sets of sentences which contain modal operators. Before describing his ideas, let us look at the work on modalities done by Hintikka's teacher, G. H. von Wright.

Von Wright: An Essay in Modal Logic

With model sets in place, one of the major challenges Hintikka took was to see how the notion of satisfiability could be generalized to sets of sentences containing alethic (*it is necessary, it is possible*), deontic (*it is obligatory, it is permitted*) and epistemic (the agent knows, believes) modalities. The context of Hintikka's work was provided by C.I. Lewis' and von Wright's work on modal logic.

C.I. Lewis (1932) considered alethic principles like

(a) If necessarily A, and A entails B, then necessarily B

$$\frac{\Box A \qquad\qquad \Box(A \to B)}{\Box B}$$

(b) Whatever is a logical law is necessary

(c) If it is necessary that A; then it is necessary that it is necessary that A

$$\Box p \to \Box\Box p$$

and investigated various modal systems to deal with them.

Von Wright (1951) investigates four groups of modalities:

- alethic modalities *(necessary, possible, contingent, impossible*)
- epistemic modalities (*verified or known to be true, undecided, falsified or known to be false*)
- deontic modalities (*obligatory, permitted, forbidden, indifferent*)
- existential modalities (*universal, existing, empty*)

The starting point of von Wright's investigations was the observation that the formal relations between concepts in one group are analogous to those of the concepts in the other groups. For instance, in the class of deontic modalities, if a proposition is obligatory, then its negation is forbidden. Its counterpart in alethic modalities is 'if a proposition is necessary, then its negation is impossible', which also holds. Von Wright develops his former technique on constituents into a method which decides, together with the truth-tables, whether a modal sentence expresses a "truth of logic" or not. By the latter von Wright means a sentence whose truth depends "upon the specific logical nature of modal concepts", e.g.

$\Diamond A \wedge \Box(A \to B) \to \Diamond B$

(Von Wright, 1951, p. 10.)

Here is an illustration of von Wright's technique for the modal system he calls M_1 which studies M_1 – sentences, that is, truth-functional compounds of atomic M_1 – sentences and/or atomic N_1 – sentences, where:

- Atomic M_1 – sentences, are atomic sentences prefixed with $\Diamond$ or truthfunctional compounds of atomic sentences, where the compound is prefixed with $\Diamond$
- Atomic N_1 – sentences, are atomic sentences prefixed with $\Box$ or truthfunctional compounds of atomic sentences, prefixed with $\Box$.

Von Wright shows how the modal principles

(I) If $\Diamond(A \vee B) \leftrightarrow (\Diamond A \vee \Diamond B)$

(II) If A and B are logically equivalent, then $\Diamond A$ and $\Diamond B$ are logically equivalent (i.e. they have the same truth-values)

provide, in combination with the truth-table method, a decision procedure for each M_1 – sentence, that is a mechanical procedure which shows whether an M_1 – sentence is a tautology of modal logic or not. It goes like this.

- Each propositional formula A has a disjunctive normal form, that is, it can be expressed as a disjunction of conjunctions of atomic sentences or their negations.
- By principle (II), $\Diamond A$ is equivalent to $\Diamond B$ where B is the disjunction normal form of A.
- By principle (I), $\Diamond A$ is equivalent to the disjunction of, say, m conjunctions, each prefixed with $\Diamond$. The latter are (modal) constituents.
- So it seems that the truth-value of each atomic M_1 – sentence could be determined from the truth-values of its

constituents by the truth-table method, provided that the constituents can appear in the truth-tables in any combination of truth-values (i.e. are independent).

To understand this later requirement, consider the modal formula $\Diamond(A \vee \neg A)$.

The disjunctive normal form of $(A \vee \neg A)$ (when the list of atomic formulas consists only of A) is $(A \vee \neg A)$. By (I) it is equivalent to $(\Diamond A \vee \Diamond\neg A)$. Its constituents are $\Diamond A$ and $\Diamond\neg A$. But given that $(A \vee \neg A)$ is a tautology, then it cannot happen, according to von Wright, that both $\Diamond A$ and $\Diamond\neg A$ are false. Thus the following principle is still needed in addition to (I) and (II):

(III) Any propositional formula A is itself possible or its negation is possible.

Now the principles (I)-(III) in combination with the truth-table method establish that $\Diamond(A \vee \neg A)$ is a logical truth in the system M_1.

By (I), $\Diamond(A \vee \neg A)$ is equivalent with $(\Diamond A \vee \Diamond\neg A)$. Its constituents are $\Diamond A$ and $\Diamond\neg A$. Thus its truth-table is:

$\Diamond A$	$\Diamond\neg A$	$(\Diamond A \vee \Diamond\neg A)$
T	*T*	*T*
T	*F*	*T*
F	*T*	*T*
F	*F*	*F*

By (III) the row in its truth table in which both A and :A are false, is deleted. Then $\Diamond(A \vee \neg A)$ comes out as "logically true in the system M_1".

What about $\Diamond(A \wedge \neg A)$? The disjunctive form of $(A \wedge \neg A)$ is empty, i.e. it is a 0-term disjunctive-sentence. We would like the truth table for $\Diamond(A \wedge \neg A)$ to be always *F* but we can't get this result from the principles listed so far and the truth-tables. Von Wright adds another principle to his list:

(IV) If a proposition is a tautology, then the proposition that it is necessary is a tautology too.

(IV) ensures that $\Box\neg(A\wedge\neg A)$ is a tautology. But $\Box\neg(A\wedge\neg A)$ is an abbreviation of $\neg\Diamond(A \wedge \neg A)$. By the truth-table method, $\Diamond(A \wedge \neg A)$ is logically false in the system M_1.

A similar method applies to atomic N_1 sentences and then to any M_1 sentence. Finally von Wright shows that these principles combined with the truthtable method shows that $\Diamond A \wedge \Box (A \rightarrow B) \rightarrow \Diamond B$ is a logical truth in the system M_1.

Von Wright (1951) (Chapter IV) also constructs a system of epistemic modalities by using epistemic counterparts of the principles (I)-(IV). They are obtained by replacing "possible" by "not falsified" and then by defining the other epistemic modalities in terms of "falsified". Thus A is falsifed, FA, expresses the same proposition as the proposition that the negation of A is verified, $V \neg A$: And A is undecided can be expressed by $\neg VA\wedge\neg V\neg A$ or equivalently by $\neg FA\wedge\neg F\neg A$. Thus from the point of view of "formal behaviour" "the verified corresponds to the necessary, the undecided to the contingent, and the falsified to the impossible."

Von Wright notices the analogy between the alethic "it is true that *p* but not neccessary that *p*" which expresses the contingency of *p* and the epistemic "it is true that *p* but not known (verified) that *p*" which expresses the epistemical contingency of *p*: But he also notices a difference between them:

Now certainly a proposition may be true without being known to be true. And certainly someone may intelligibly say "it is true that *p*, though nobody knows it". But if he said "It is true that *p*, though nobody knows it, not even I" we should feel there was something linguistically wrong. (von Wright, 1951, p. 32)

We recognize today that von Wright's example is an illustration of the so-called Moore's paradox. In his review of von Wright (1951), Strawson (1953), takes the mentioned difference between alethic and epistemic notions to throw doubts on the whole enterprise of epistemic logic: "Facts of this kind may lead us to wonder how far a system of epistemic modalities can contribute to the philosophical elucidation of words like "know" ". Later on in *Knowledge and Belief*, Hintikka (1962) offers a solution to "Moore's" paradox (cf. below.)

Von Wright also deals with combinations of epistemic and existential modalities, that is, quantified epistemic logic. Of these combinations he is particularly interested in epistemic-existential sentences (*de dicto*), e.g. "It is known that something is red", existential-epistemic sentences (*de re*), e.g. ""Something is known to be red" and the system which combines both. He points out that the first two notions require no new governing principles, but the third one requires two new principles (idem, p. 49):

(IV) If it is known that everything possesses a certain property, then everything is known to possess that property

(V) If there is a thing which is known to possess a certain property, then it is known that something possesses this property

Von Wright points out that none of these principles is convertible. Later on in Knowledge and Belief Hintikka will show that these principles are valid using the technique of model sets and model systems.

The decision method for epistemic modalities is completely similar to the previous one, i.e. we reduce the original V E-sentence to a truth-function of atomic constituents, the only difference being that the atomic constituents have now the form FC where C is a constituent in a monadic predicate language (see section 1), that is, a specification of a possible world built up from disjoint unary predicates of the underlying language and the existential quantifiers or their negations. Skipping over many details, the normal form of the V E-sentence $VEA \vee \neg FUA$ (here EA is an abbreviation of $\exists xA$ and UA of $\forall xA(x)$) turns out to be

$$\neg(\neg F(\neg EA\wedge E\neg A)\vee\neg F(\neg EA\wedge\neg E\neg A))\vee$$
$$(\neg F(\neg E\neg A\wedge EA)\vee\neg F(\neg E\neg A\wedge\neg EA))$$

which is a truth-function of the atomic V E-constituents $F(\neg EA\wedge E\neg A)$, $F(\neg EA\wedge\neg E\neg A)$ and $F(\neg E\neg A\wedge EA)$. Thus we can check, by the truth-table method whether this formula is a logical truth or not. The only restriction on the distribution of truth-values (which does not apply to this case), is that if a sentence has a maximal number of V E- constituents (the disjunction of the correponding E-constutents is a tautology), then not all of them can be falsified.

Finally von Wright investigates "higher-order" modalities (e.g. "it is possible that it is necessary that *p*") for which he needs a new principle of reduction:

(VI) If it is possible that a certain property is possible, then the property is possible.

Von Wright shows that, if this principle is adopted, then higher-order modal sentences can be shown to be equivalent to truth-functional complexes of of first-order modal properties.

In Appendix II, von Wright investigates various axiomatic systems and compares them to C.I.Lewis's systems. Von Wright points out that if 'verified' or 'known to be true' refer to the actual knowledge of some particular person, then the counterparts of Lewis' principles may fail. We will see later on that Hintikka interpreted these notions in the same way as von Wright: they refer to idealized agents.

Modality and Quantification (Hintikka 1961)

Von Wright's analysis of modal notions did not appeal to the notion of possible worlds as alternatives to our actual world. Hintikka took a different route. He did not use constituents but model sets. In Hintikka (1957a), (1957b) and (1961) Hintikka extends the notion of satisfiability of sets of formulas of predicate logic to sets of formulas with modal operators such as "it is possible" and "it is necessary" and deontic operators like "it is obligatory that". One of his main insights is that the satisfiability of a set of sentences involving modal notions forces us to consider sets of model sets:

In our definition of satisfiability, we therefore have to consider sets of model sets. Such sets of sets we shall call *model systems*. (Hintikka, 1961, p. 122)

Hintikka inquires into the question of what conditions must model systems be subject to:

"Suppose that $\Diamond A \in \mu \in \Omega$ where Ω is a model system (and where $\Diamond$ is to be read 'possibly'). Then clearly we have to require that *A*;which is perhaps not true in the state of affairs described by μ, must nevertheless be true in some other state of affairs which could have been realized instead the one described by μ. Descriptions of such states of affairs will be called alternatives to μ. In other words, the following condition must be satisfied:

(C.M*) If $\Diamond A \in \mu \in \Omega$; then there is in Ω at least one alternative v to μ such that $A \in v$" (Hintikka, 1961, p. 123)

Similar conditions are associated with $\Box$A:

(C.N) If $\Box A \in \mu \in \Omega$, and if $v \in \Omega$ is an alternative to μ, then $A \in v$.

The combination of modal notions and model sets bring in difficulties of their own. Consider a formula with an occurrence of an individual constant (or free individual symbol), which belongs to a model set μ. We can safely assume, following Hintikka, that the individual constant in question stands for an individual which exists in the state of affairs described by μ. But if the formula in question belongs to several model sets at the same time, the previous assumption has more severe consequences. It implies that an individual may exist in more than one 'possible world':

The presence of a free individual variable in the formulae of μ, we may thus say, is the formal counterpart to the existence of its value in the state of affairs described by μ. From this it follows that when a formula *A* is transferred from a model set μ to one of its alternatives say v- we have to heed the free individual variables *A* contains. If one of them does not occur in the other formulae of v, then the adjunction of *A* to v is legitimate only if the relevant values of this free individual variable are assumed to exist not only in the state of affairs described by v but also in that described by v. In general this assumption cannot be made. Individuals which de facto exist may possibly fail to do so. (Hintikka, 1961, p. 125.)

Hintikka then considers the following variant of (C.N):

(C.N*) If $\Box A \in \mu \in \Omega$, and if $v \in \Omega$ is an alternative to μ, and if each free individual variable of *A* occurs in at least one other formula of v, then $A \in v$.

Modal principles like (C.M*), (C.N) and (C.N*) suffice, accordingly to Hintikka, for a minimal modal logic. Once they are in place, he is able to formulate the definition of satisfiability for sets of modal sentences:

- A set λ of formulae is satisfiable if and only if there is a model system (Ω; *R*) such that $\lambda \subseteq \mu$ for some member μ of Ω.

For emphasis: a model system is a pair (Ω; *R*) where the first member Ω is a set of model sets and the second member *R* is the relation of alternativeness which satisfies (C:M*) and (C.N*).

In Hintikka (1961) it is mentioned that the semantical system thus obtained (interesting enough, Hintikka considers the pair (Ω; *R*) a *semantical* system) is equivalent to von Wright's system M (von Wright, 1951.) Hintikka also mentions that by requiring the relation *R* to be transitive, we obtain a stronger system which is equivalent to Lewis' s system S4, and by requiring it to be symmetric, we obtain a semantical system whose syntactical twin is obtained by adding to the system M the Brower's axiom

$A \to \Box\Diamond A.$

Further on, he notices that by requiring *R* to be transitive and symmetric, we obtain a system which is equivalent to Lewis's S5. Hintikka (1961) adds, however: "I shall not prove these results here". Instead he mentions that the principle

$\exists x \Box F(x) \to \Box \exists x F(x)$

is not satisfiable in a model system which obeys (C.N*) (Hintikka, 1961, p. 124.)

Hintikka does not present a formal argument, but it is easy to build one.

Here it is.

Suppose there is a model set μ in a model system Ω such that:

1. $\exists x \Box F(x) \in \mu$
2. $\neg \Box \exists x F(x) \in \mu$

By a series of equivalent transformations on (2), we get

3. $\Diamond \forall x \neg F(x) \in \mu$

which, by (CM*), implies

4. $\forall x \neg F(x) \in \mu^*$

where μ^* is an alternative to μ. From (1) and (C.E) we get

5. $\Box\, F(b) \in \mu$.

Now if we applied (C.N) to (5) we would get

6. $F(b) \in \mu^*$

which together with (4) would give $\neg P(b) \in \mu^*$, contradiction.

However, when we replace (C.N) with (C.N*), we cannot any longer make the transition from (5) to (6) given that (the individual denoted by) *b* does not occur in (other formulae of) μ^*. In other words, we can prove the validity of

$\exists x \Box P(x) \to \Box \exists x P(x)$

if we assume that if (the individual referred to by) *b* exists in μ (recall that for Hintikka a free individual variable or individual constant occurring in a formula in a model set is the formal counterpart of an individual in the possible world described by the model set), then it also exists in μ^*; but this is precisely what Hintikka denies.

Hintikka contemplates the possibility to restore the validity of $\exists x \Box F(x) \to \Box \exists x F(x)$ by requiring that "whatever exists in a possible state of affairs exists in all the alternative states of affairs; in short, that whatever exists exists necessarily." (Hintikka, 1961, p. 125). But he does not go for it (Kripke also discusses this solution in Kripke, 1963.) He also mentions a condition on model sets which "formulates exhaustively the assumption that free individual variables are transferable from a model set to its alternatives":

(C.self=*) If *b* occurs in at least one formula of μ and if v is an alternative to μ, then $a = a \in v$. (Hintikka, 1961, p. 125).

Let us take stock. Hintikka (1961) considers two kinds of modal systems. One of them, which satisfies (C.self=*), embodies the assumption that all actually existing individuals exist necessarily; the other one, which satisfies (C.N*), dispenses with this assumption. We shall see that in *Knowledge and Belief* (1962), Hintikka points out an interesting difference between the alethic principle $\exists x \Box P(x) \to \Box \exists x P(x)$ that Hintikka denies, and its epistemic counterpart

$\exists x K_a F(x) \to K_a \exists x F(x)$

that Hintikka, like von Wright, endorses (cf. von Wright principle V.)

Knowledge and Belief (Hintikka 1962)

Hintikka (1962) investigates the satisfiability of sets of sentences involving knowledge and belief in the context of model sets:

...we are led to ask how the properties of model sets are affected by the presence of the notions of knowledge and belief; how, in other words, the notion of model set can be generalized in such a way that the consistency (defensibility) of a set of statements remains tantamount to its capacity of being embedded in a model set. What additional conditions are needed when the notions of knowledge and belief are present? (Hintikka, 1962, p. 34)

The basic concepts are now "the agent *a* knows that *A*", symbolized by K_aA; and "it is possible, for all the agent knows that *A*", symbolized by P_aA: In the new context Hintikka does not speak any longer of consistency and inconsistency of a formula or a set of formulae, but of *defensibility* and *indefensibility*, respectively; and instead of valid sentences he talks about *self-sustaining* sentences. Thus to show that a set of sentences is defensible one has to show that it is embeddable onto a model system ($\Omega;R$) where Ω is a set of model sets and *R* is the alternativeness relation. And to show that a set of sentences λ is indefensible, one has to show that there is no model model set $\mu \in \Omega$ of a model system ($\Omega;R$) such that $\lambda \subseteq \mu$.

The notions K_aA and P_aA, introduce new requirements on model systems. Some of them are simply counterparts of their alethic relatives, e.g.:

(C.K) If K_aA belongs to a model set μ (in a model system Ω), and if μ^* is an alternative to μ (with respect to the agent a) in Ω, then *A* belongs to μ^*.

(C.¬K) If $\neg K_aA$ belongs to a model set μ, then $P_a\neg A$ belongs to μ.

(C.P) If P_aA belongs to a model set μ, then there is at least one alternative μ^* to μ in Ω such that *A* belongs to μ. Etc.

But there are new requirements which reflect the specific properties of knowledge and belief. For knowledge, it is required that the alternative relation be at least reflexive and transitive:

(C.K*) If K_aA belongs to a model set μ, then A also belongs to μ

(C.KK*) If K_aA belongs to a model set μ in some model system Ω, and if μ^* is an alternative to μ (with respect to the agent *a*) in Ω, then K_aA belongs to μ^*.

The latter says that everything the agent *a* knows in the state of affairs described by μ, is also known in every *a*-alternative state of affairs described by μ^*. It correspond to the knowledge-axiom:

$$K_aA \rightarrow K_aK_aA.$$

The purpose of (C.KK*) is to enforce a robust, infaillible notion of knowledge. It can be shown that in the absence of (C.KK*), there is a model set μ in a model system Ω such both $K_aA \in \mu$ and $K_a(B \rightarrow \neg K_aA) \in \mu$. That is, in such situations, the agent knows that *A* but he also knows that if *B* is the case, he will loose the knowledge that *A*. Hintikka rejected this "faillibilist" conception of knowledge. (C.KK*) rules out model sets of this kind.

Hintikka's defense of the (C.KK*) principle makes it clear that Hintikka is concerned with *virtual knowledge*, that is, knowledge of cognitively perfect agents who are sufficiently clever to be able to carry out the implications of what they know. In accordance with this line, Hintikka's interpretation of all the principles (C.K)-(C.KK*) is that for a cognitively ideal agent it is irrational (indefensible) to claim that e.g. he knows that *A* and to deny, on the same occasion, that *A*.

Knowledge and Belief contains many indefensibility arguments. The proof of the indefensibility of a statement *A* is interpreted, in the spirit of the model set technique, as an aborted attempt to describe a state of affairs in which *A* would be true; and in the same spirit "every proof of the fact that a statement *p* implies epistemically another statement q is, intuitively speaking, an aborted attempt to describe consistently a state of affairs (with alternatives) in which *p* would be true but *q* false." (Hintikka, 1962, p. 45).

Here is one of Hintikka's examples of a self-sustaining principle. We show that

$K_aA \land K_aB \rightarrow K_a(A \land B)$

is self-sustaining by trying to build up a model set in which the antecedent is true (i.e. it belongs to a model set) and the consequent is false (i.e. its negation belongs to the same model set).

Suppose there is a model set μ in a model system Ω such that

1. $K_aA \land K_aB \in \mu$ (assumption)
2. $\neg K_a(A \land B) \in \mu$ (assumption)

From (2) and (C.¬K) we get

3. $P_a\neg(A \land B) \in \mu$

and thus by

4. $\neg(A \land B) \in \mu^*$

for some alternative μ^* to μ.

Skipping over a couple of steps, which lead to $K_aA \in \mu$ and $K_aB \in \mu$, we infer by (C.K):

5. $A \in \mu^*$
6. $B \in \mu^*$
7. $(A \land B) \in \mu^*$ (5,6 and logic).

We have derived a contradiction, which shows that the negation of $K_aA \land K_aB \rightarrow K_a(A \land B)$ is indefensible and thus this sentence itself is self-sustaining.

Using this technique, Hintikka is able to show how the epistemic counterparts of C.I Lewis S4 are self-sustaining. He also gives a solution to some traditional puzzles, like Moore's paradox. Finally, Hintikka defends his program in epistemic logic against Quine's criticisms of modal logic by showing that substitutivity of identity and existential generalization make sense in modal contexts, provided certain assumptions are fulfilled. Let me shortly say few words about some of these matters.

Moore's paradox

In Hintikka (1962) he discusses Moore's paradox of "saying and disbelieving". He starts by noticing that there is something logically queer about someone asserting

1. *A* but I do not believe that *A*

even if it is not self-contradictory (indefensible) according to the criteria he set up. He offers the following explanation of the absurdity of (1).

It is expected from anyone (say b) who asserts the sentence

2. *A* but a does not believe that *A*

"that it is possible for him to believe what he says, that is, it would be defensible for him to say

3. I believe that the case is at follows: *A* but a does not believe that *A*". (idem p. 52)

This sentence is of the form

4. $B_b(A \land \neg B_aA)$

while (1) is of the form

5. $B_a(A \land \neg B_aA)$.

Now Hintikka shows that (5), unlike (4), is indefensible in his system. To show this, he follows the usual reductio ad absurdum proof, and supposes (5) belongs to a model set. Then using the transitivity of belief, he derives a contradiction (p. 52). Hintikka mentions that he has offered a solution to Moore's puzzle which does not invoke any additional principles to the ones he has so far introduced. Perhaps a short critical remark should be considered at this point. True, Hintikka does not strengthen the *logical* principles that govern knowledge and belief. He does introduce, however, without noticing, an extra-assumption, which is a norm of assertion: assert a sentence only if you believe it (i.e it is defensible).

Quantifiers and identity in epistemic logic: 'knowing who'

The combination of epistemic notions with quantifiers and identity leads to problems analogoue to those we encountered in alethic quantified systems. These matters have been extensively debated and we will not explore them in great details here. We shall focus on Hintikka's notion of 'knowing who' and the way he perceived the difference between the logical treatment of alethic and epistemic notions.

The presence of quantifiers, identity and knowledge operators allows Hintikka to represent in his logical setting the notion "*a knows who b* is". For instance, he renders "*a* knows who Mr. Hyde is" as

$$\exists x K_a(x=h)$$

This notion introduces requirements of its own on model systems:

(C.EK=EK=*) If $\exists x K_a(b=x) \in \mu$, and μ^* is an epistemic alternative to μ with respect to *a*; then $\exists x K_a(b=x) \in \mu^*$.

(C.EK=) If $\exists x K_a(b=x) \in \mu$, then $\exists x(b=x) \in \mu$.

The second condition tells us that if *a* knows who (the individual referred to by) *b* is in the possible world *μ*, then *b* exists in *μ*. The first condition tells us that if *a* knows who *b* is in the possible world *μ*, then *a* knows who *b* is in all *a*'s epistemic alternatives.

Hintikka's justification of these principles is based on his decision to take 'knowing who' to behave logically in the same way as 'knowing that' (Hintikka, 1962, p. 116). Thus (C.EK=) may be seen as the counterpart of the principle

$$K_a A \rightarrow A$$

for *knowing who*. And analogously, (C.EK=EK=*) may be seen as the counterpart for knowing who of (C.K.K*), which ensures the validity of the axiom

$$K_a A \rightarrow K_a K_a A.$$

We pointed out earlier that this axiom (and its semantical counterpart (C.K.K*)) ensure a robust notion of knowledge. In the same way, (C.EK=EK=*) ensures that if an agent knows who (the individual denoted by) *b* is, then he is not going to loose this knowledge in any of his epistemic alternatives.

More generally, existential and universal quantifiers, have, in non-epistemic contexts, rules of instantiation which are completely analogous to the rules (C.∃) and (C.U). But the interaction of quantifiers and epistemic operators produces additional problems, as already witnessed by (C.EK=EK=*) and (C.EK=). Hintikka compares these problems with their counterparts in alethic contexts referring back to Hintikka (1961). When we discussed that paper in an earlier section, we pointed out that constructions of the form

$$(\exists x \Box \ldots x \ldots) \in \mu \quad (\exists x \Diamond \ldots x \ldots) \in \mu$$

raise the question of whether an individual *existing* in a model set *μ* also exists in the alternatives that □ or ◊ forces us to consider.

In *Knowledge and Belief*, Hintikka considers the analogue constructions

$$(\exists x K_a \ldots x \ldots) \in \mu \qquad (\exists x P_a \ldots x \ldots) \in \mu$$

but he interprets them in a different way. The quantifiers in these constructions "range", not over only individuals existing in *μ* but over individuals existing in *μ* which are also known (in the sense of knowing who, that is, identified). Thereby the problem these constructions raise is whether an individual known by *a* in the model set *μ*, is also known by a in *a*-alternatives to *μ*. And given Hintikka's notion of *knowing who* and the analogy he draws between this notion and *knowing that*, his answer is positive. That is, in the general case, the following constraints on modal systems are added:

(C.E$_{ep}$) If $\exists x A \in \mu$, then $A(x/b) \in \mu$ and $\exists x K_a(x = b) \in \mu$ (it is supposed that *A* contains an occurrence of the operator K_a or P_a and '*x*' occurs within the scope of one of them in *A* but not within the scope of any other epistemic operator).

(C.U$_{ep}$) If $\forall x A \in \mu$ and $\exists x K_a(x = b) \in \mu$, then $A(x/b) \in \mu$ (with the same assumptions as in (C.E$_{ep}$)).

We note that both $(\exists x K_a \ldots x \ldots) \in \mu$ and $(\exists x P_a \ldots x \ldots) \in \mu$ fall under the incidence of (C.E$_{ep}$). They both generate a substitutional instance $(K_a \ldots b \ldots) \in \mu$ and $(K_a \ldots b \ldots,) \in \mu$, respectively, such that *a* knows who *b* is in *μ*. (C.EK=EK=*) further ensures that *a* knows who *b* is also in *a*'s epistemic alternatives.

With the help of these principles, Hintikka is able to show the self-sustenability of the principle

$$\exists x K_a F(x) \rightarrow K_a \exists x F(x)$$

whose counterpart in alethic logic

$$\exists x \Box F(x) \rightarrow \Box \exists x F(x)$$

he rejects. Here is Hintikka's argument (1962, p. 117).

Suppose there is a model set *μ* in a model system such that

1. $\exists x K_a F(x) \in \mu$, and
2. $\neg K_a \exists x F(x) \in \mu$.

By equivalent transformations on (2) we get:

3. $P_a \forall x \neg F(x)$

which togetner with (C.P) implies

4. $\forall x \neg F(x) \in \mu^*$

where μ^* is an *a*-alternative to *μ*.

From (1) and (C.E$_{ep}$) we get

5. $K_a F(x/b) \in \mu$
6. $\exists x K_a(x = b) \in \mu$.

From (6) we obtain using by (C.EK=EK=*)

7. $\exists x K_a(b = x) \in \mu^*$

from which we derive, using (C.EK=)

8. $\exists x(b = x) \in \mu^*$.

From (5) and (C.K) we get:

9. $F(x/b) \in \mu^*$

and from (6), by (C.EK=EK=*) we obtain

10. $\exists x K_a(b = x) \in \mu^*$

Now that *a* knows in μ^* who (the referent of) *b* is, we can instantiate the formula in (4) and get

10. $\neg F(x/b) \in \mu^*$.

We have obtained the desired contradiction which shows the indefensibility of the negation of $\exists x K_a F(x) \rightarrow K_a \exists x F(x)$ and thereby the self-sustainability of the formula itself.

Hintikka concludes the argument with the following observation:

> The self-sustenance of $[\exists x K_a F(x) \rightarrow K_a \exists x F(x)]$ shows that there is an interesting difference between the logical behavior of the notion of knowledge and that of the notion of necessity toward quantifiers, in spite of the fact that the two are closely similar in many respects. For the notion of necessity the analogue of $[\exists x K_a F(x) \rightarrow K_a \exists x F(x)]$ is not valid. From the fact alone that there exists an individual which cannot help having a certain property it does not follow that there necessarily is an individual with this property. For the individual first mentioned might conceivably not exist. (Hintikka, 1962, p. 117)

In a footnote Hintikka refers to the argument against the self-sustenance of

$$\exists x \Box P(x) \rightarrow \Box \exists x P(x)$$

that he presented in his earlier paper (Hintikka, 1961). Indeed, as we recall from our earlier section, the proof of the validity of this sentence required the assumption that if (the individual referred to by) *b* exists in *μ*, then it also exists in μ^*. Hintikka rejected it, and consistently with that, he also rejects it in *Knowledge and Belief*, as the quote above indicates.

Hintikka on Quine's criticism of modal logic

Hintikka's work in epistemic logic went against Quine's arguments to the effect that quantifier rules like existential generalization and substitutivity of identity are misguided in alethic contexts. Hintikka acknowledges that none of these rules holds uniformly in epistemic contexts. That is, one cannot always infer

1. *a* knows that Dr. Jekyll is a murderer (i.e., $K_a(M(j))$) from the premises
2. *a* knows that Mr. Hyde is a murderer (i.e., $K_a(M(h))$) and
3. Dr. Jekyll is the same man as Mr. Hyde (i.e. $j = h$).

Neither can one infer

4. $(\exists x)K_a(M(x))$

from (2).

For Quine, the failure of substitutivity in the first example indicates the referential opacity of the position occupied by the term "Mr. Hyde". This feature is also responsible for the impossibility of existential generalization in the second example. Quine's solution was to restrict these rules to referentially transparent contexts.

For Hintikka (1962), the *failures are not failures* of referentiality, that is, they are not due, as Quine sometimes seems to suggest, to the way in which our singular terms refer to objects. The source of the failures has to do rather with *multiple referentiality*, that is, with the fact that a has to consider several epistemic alternatives to the current one. In some of these "possible worlds" the proper names "Dr. Jekyll" and "Mr. Hyde" refer to two distinct men (p. 102). For Hintikka substitutivity of identity makes perfectly good sense in epistemic contexts, provided that a knows that Mr. Hyde is the same man as Dr. Jenkyll, a requirement that Hintikka formulates as

$K_a(h = j)$

In an analogous way, Hintikka goes on, "quantifying in" that is, moving from

$K_a \ldots h \ldots$

to

$\exists x K_a \ldots x \ldots$

goes smoothly whenever *a knows who* Mr. Hyde is, that is, whenever $\exists x K_a(x = h)$ also holds (p. 112).

Semantically speaking, Hintikka (1962) interprets clauses of the form $K_a(h = j)$ as saying that the two names refer to the same individual in every *a*-epistemic alternative; and he interprets clauses of the form $\exists x K_a(x = h)$ as ensuring that *h* names the same individual in every relevant epistemic alternative (pp. 111-112). Yet, I would like to claim, against Hintikka, that none of the rules Hintikka proposes ensures that '*b*' refers to one and the same individual in every possible world in which *b* exists. This can be seen in the following way.

Suppose that $\exists x K_a(b = x) \in \mu$, and μ^* is an epistemic alternative to μ. From this one can derive, using (C.EK=EK=*) and (C.EK=), that

$$\exists x(b = x) \in \mu$$

and

$$\exists x(b = x) \in \mu^*:$$

The most we can now get from these conditions, using the model sets technique based on the substitutional interpretation of quantifiers, is that $b = c \in \mu$ and $b = d \in \mu^*$ for some constants *c* and *d*. The two conditions are compatible with both the "descriptive" interpretation of individual constants according to which the referent of such a constant may vary from world to world, and with the "rigid" interpretation according to which the interpretation remains fixed. In other words, the non-referential semantics with its substitutional interpretation of quantifiers the technique of model sets relies on, cannot enforce that '*b*' refers to one and the same individual in every relevant possible world. In our particular example, '*b*' and '*c*' could very well refer to one and the same individual, say *e*, in the possible world described by μ, and, on the other side, '*b*' and '*d*' could refer to the individual $d \neq e$ in the world described by μ^*. Hintikka came to realize this later on, or so we would like to think. For instance, in Hintikka and Sandu (1995) the authors claim that when the quantifiers are interpreted objectually (and extensionally), then $\exists x K_a(b = x)$ and $\exists x \Box(b = x)$ express that '*b*' is a "rigid designation" in epistemic and alethic contexts, respectively (p. 181 in Hintikka 1998).

Hintikka, Kanger and Kripke

Kanger's reconstruction of Hintikka's early work (Kanger, 1972)

As mentioned earlier, Hintikka's model sets share common features with Carnapian state descriptions. Carnap (1946, 1947) defines a notion of universal modality "it is necessary that *A*", $\Box A$, in a straightforward way:

$\Box A$ is true at a state description Σ if and only if *A* is true at all state descriptions Σ'.

An essential new ingredient in Hintikka's work, compared to Carnap, is the alternativeness relation *R* and the notion of model system (Ω, R). By varying *R* Hintikka is able to model various modal and epistemic notions, as we have seen in earlier sections. Model systems appear in print in Hintikka's early work in 1957, 1961, and 1962. But a striking difference between Hintikka and Carnap, as well as, as we shall see, between Hitnikka and Kripke, and Hintikka and Kanger, is that Hintikka never presented explicitly in his early work a recursive definition of the notion of truth in a model (model set, possible world) for a logical language which combines both quantifiers and modalities. That is, he never presented a definition of the form

A is true in a model (possible world, model set) iff....

where A runs over modal and quantified formulas. Hintikka was concerned with the semantical notion of consistency (satisfiability, defensibility) of a sentence or set of sentences, and not with the notion of truth in a particular model, as he often emphasized (e.g. Hintikka, 1961). This had some important consequences.

In model sets (and state descriptions) quantifiers are treated substitutionally. The substitutive interpretation of quantifiers dispenses with the notion of model and with the notion of reference. For this reason, although Hintikka often speaks of free variables and individual constants referring to an individual or another in the possible world describe by the relevant model set, this talk remains at an informal level and the assumptions behind it are never made explicit.

It is interesting, against this background, to compare Hintikka's ideas to the work of two of his contemporaries, Stig Kanger and Saul Kripke, who worked on the same problems during the same time as Hintikka. Kanger is the only one of the three who presented a detailed comparison of his own framework with Kanger's and Hintikka's systems. In his comparison, Kanger (1972) extracts from Hintikka's theory a recursive truth-definition of the notion "A is true in a state-description *S* of a model system (Φ, R)" but confesses that

"We shall here formulate Hintikka's theory so that its relationship with Carnap becomes explicit- or maybe over-explicit. (In fact, we are depriving Hintikka's theory of one of its virtues.)" (Kanger, 1972, p. 115).

This is an over-explicitation indeed, as I tried to point out that due to his special interests, Hintikka took as the basic concept of his semantical theory the notion of satisfiability.

In Kanger's reconstruction, he attributes to Hintikka the notion truth in an interpretation(Ω, H) that he denotes by $T(A, (\Omega, H))$. Here Ω is a nonempty class of state descrption (!) and H is a member of Ω. It is useful to recall the definition of a state-description:

(C.1) If A is an atomic sentence (or an identity) then not both $A \in \mu$ and $\neg A \in \mu$

(C.2) If A is an atomic sentence (or an identity) then either $A \in \mu$ or $\neg A \in \mu$

(C.3) If A is an atomic sentence (or an identity) or the negation of an atomic sentence (identity), and if $A \in \mu$ and $a = b \in \mu$ and if B is exactly like A except that a and b have been interchanged in one or several places, then $B \in \mu$.

(C.4) Not $\neg(b = b) \in \mu$.

Kanger defines $T(A, (\Omega, H))$ recursively for *sentences*:

(a) $T(P(a_1,\ldots, a_n); (\Omega, H))$ is the truth-value T (true) iff $P(a_1,\ldots, a_n) \in H$;

(b) $T(a_1 = a_2; (\Omega, H))$ is T iff $a_1 = a_2 \in H$

(c) $T(\neg A, (\Omega, H))$ is T iff $A \notin H$

The clauses for other extensional connectives are standard.

(d) $T(\forall xB, (\Omega, H))$ is the truth-value T iff $T(B(x/b); (\Omega, H))$ is T for each individual constant b; etc

The truth for modal sentences uses the relation of accessibility:

- $T(\Box A; (\Omega, H))$ is T iff $T(\Box A, (\Omega, H))$ is T for every $H' \in \Phi$ such that HRH'.

Kanger's "over-explicitation" or rather "reconversion" of Hintikka's model systems into state-sescriptions and his attribution to Hintikka of the notion of "truth in a state-description" are useful. It illustrates the strategy of how one can take satisfiability as the basic semantical notion and then extract a recursive definition of the derived notion of "truth in a possible world". It leaves out, however, the restrictions on model systems that played such a major role in Hintikka's thought.

Kripke and Hintikka

In contrast to Hintikka, both Kanger (1957, 1972) and Kripke (1959, 1963) take "truth in a possible world (model)" as the basic semantical notion, and define satisfiability in terms of it. Also both of them use, not substitutional quantification, as Hintikka and Carnap did, but the Tarskian notion of satisfaction. In other words, quantifiers are defined objectually, the range of quantifiers is made explicit and so is the notion of reference and model.

Kripke (1959) considers a modal language which contains individual and propositional variables, predicate symbols and the modal operator $\Box$.

Given a non-empty domain D of individuals, for each formula (!) A in the object language one defines the notion of a complete assignment for A in D, which is a function that assigns:

- to every free individual variable of A an individual in D,
- to every propositional variable which is a subformula of A either the truthvalue T or F, and
- to every n-place predicate symbol P occurring in A an n-place relation on D.

A *model of A* and D is an ordered pair (G, K) of complete assignments for A in D, where $G \in K$ and all the assignments of K agree on the free variables of A. The assignment G is supposed to play the role of the actual world and the set K is to be thought of as the set of all possible worlds. Notice that there is no accessibility relation on K.

Given a model (G, K) for A and D, every subformula B of A receives the value T or F relatively to an arbitrary assignment $H \in K$ in a recursive way:

(i) If B is an atomic formula $P(x_1,\ldots, x_n)$, then it receives the value T if and only if the n-tuple $(a_1,\ldots, a_n)$ assigned by H to the free variables $x_1,\ldots, x_n$ belongs to the extension of P as given by H; otherwise it is assigned the value F.

(ii) If B is $x_1 = x_2$, then it receives the value T if and only if the individual in D assigned to x_1 by H is the same as that assigned to x_2. Otherwise it receives the value F. The clauses for the extensional connectives are standard.

(iii) B is $\forall xC(x)$, then it receives the value T if and only if $C(x)$ is assigned the value T for every assignment of an element of D to x; otherwise it receives the value F.

(iv) If B is $\Box C$, then it receives the value T if and only if every member of K assigns the value T to C; otherwise it receives the value F.

Few things need to be emphasized in Kripke's definition:

- Possible worlds are truth-value assignments
- They all share a common domain of individuals D
- A model is relativized to a modal formula A
- A free variable x is assigned an element of the commonly shared domain D (and thus its interpretation remains "rigid")
- On the other side, the interpretation of a predicate symbol P may vary from world to world, that is, there are assignments in K which assigns to P different extensions in the domain D
- Given that there is no accessibility relation, $\Box$ expresses an universal (S5) notion of necessity, like in Carnap.

A formula A is said to be *valid in a model* (G, K) of A and D if A is assigned the value T by G (informally: A is true in the actual world). Actually later on in Kripke (1963) he acknowledges that a better notion than "valid in a model" is "true in a model". A *is valid in D* simpliciter if A is valid in every model of A on D. A is *satisfiable* if there is a non-empty domain D and a model of A on D such that A is valid in this model. Finally, A is *universally valid* if A is satisfiable in every non-empty domain D. The formula B is *semantically entailed* by $A_1, A_2, \ldots, A_n$ if and only if $(A_1 \wedge A_2 \wedge \ldots \wedge A_n) \rightarrow B$ is universally valid. Notice that if $n = 0$ this amounts to B being universally valid.

Kripke proves a completeness theorem which shows that B is semantically entailed by $A_1, A_2, \ldots, A_n$ if and only if the semantical tableau construction where $A_1, A_2, \ldots, A_n$ are on the left side and B is on the right side of the tableau closes. We will not enter into these details here but we take note of the con-nection between semantic tableaux (model sets) and the semantical notion of entailment and universal validity which are defined in terms of truth in a model.

In Kripke (1963), this picture is radically changed. In a footnote at the beginning of the paper, Kripke tells the readers that: The authors closest to the present theory appear to be Hintikka and Kanger. The present treatment of quantification, however, is unique as far as I know, althought it derives some inspiration from acquaintance with the very different methods of Prior and Hintikka. (Kripke 1963, Footnote 1, page 83.)

What is this "unique treatment of quantification"? Essentially, it is obtained by imposing a quantificational structure on a set of possible worlds (and an accessibility relation). This happens by

- relativizing the range of a quantifier to a possible world; in order to do this, each possible world is endowed with its own universe.
- providing a semantic value for free variables à la Tarski through the notion of assignment; the objects assigned to the free variables may come from any of the individual universes.
- relativizing the notion of satisfaction to a possible world and an assignment.

Here are the technical details.

The starting point is the notion of model structure (m.s.) for a propositional modal language. It is a triple (G, K, R), where K is the set of possible worlds, G is the actual world, $G \in K$, and R is an accessibility relation on K that Kripke interprets as follows:

- For every $H_1, H_2 \in K$, H_1RH_2 means that H_2 is possible relative to H_1, that is, every proposition true in H_2 is possible in H_1. (Kripke, 1963, p. 84.)

Kripke notices that reflexivity of R is a natural requirement and mentions that one may impose additional requirements, corresponding to various axioms of modal logic.

Given a model structure (G,K,R), a model for the propositional modal language is a binary function φ which assigns to each atomic formula P and possible world H in K, a truth-value $\varphi(P, H)$ which is T of F. We recognize in the notion of a model the ancestor of what nowadays is called a *Kripke-model* for a modal propositional language.

Given a model, one can then assign by induction truth-values for complex propositional formulas. The clause which interests us is:

- $\varphi(\Box A,H) = T$ iff $\varphi(A, H') = T$ for every $H' \in K$ such that HRH'. Informally: A is necessary in H iff A is true in all worlds H' possible relative to H.

A *quantified model structure* (q.m.s.) is a model structure (G,K,R) together with a function ψ which assigns to every possible world H in K its own domain $\psi(H)$, that is, the set of individuals existing in H. We are told that:

Notice, of course, that $\psi(H)$ need not be the same set for different arguments H, just as, intuitively, in worlds other than the real one, some actually exising individuals may be absent while new individuals, like Pegasus, may appear. (Kripke, 1963).

Let U be the set of all individuals which exist in some world or another in K (i.e. $U = \bigcup_{H \in K} \psi(H)$.) A *quantificational model* on a q.m. is now defined as a binary function $\varphi(P^n, H)$ where the second variables ranges over possible worlds in K and the first variable over predicate symbols of the underlying language. When $n = 0$, P^n is a propositional letter and thus $\varphi(P^n, H)$ is T or F. For $n \geq 1$, $\varphi(P^n, H)$ is a subset of U^n, that is, an n-place relation which is the extension of P^n in the possible world H.

Kripke defines inductively $\varphi(A, H)$, the truth value of the formula A in the possible world H *relative to an assignment* of individuals in U to the free variables of A:

(i) The case of propositional variables has been taken care of.

(ii) If A is $P^n(x_1,\ldots, x_n)$ $(n \geq 1)$, and the assignment assigns the individuals $a_1,\ldots, a_n$ from U to the variables $x_1,\ldots, x_n$, then $\varphi(P^n, H) = T$ if the n-tuple $(a_1,\ldots, a_n)$ belongs to $\varphi(P^n, H)$.

The inductive steps for the propositional connectives are straightforward.

(iii) If A is $\Box B$, $\varphi(A, H) = T$ relatively to the assignment is T if and only if $\varphi(B, H') = T$ for all the possible worlds H' such that HRH' (relative to the same assignment).

(iv) Assume now we have a formula $A(x, y_1,\ldots, y_n)$ where $x, y_1,\ldots, y_n$ are the only free variables present. Assume also that $\varphi(A(x, y_1,\ldots, y_n), H)$ has been defined for each possible assignment to the free variables $x, y_1,\ldots, y_n$. Then we define $\varphi(\forall x A(x, y_1,\ldots, y_n);H) = T$ relative to an assignment $a_1,\ldots, a_n$ of elements of U to the free variables $y_1,\ldots, y_n$ if $\varphi(A(x, y_1,\ldots, y_n), H) = T$ for every assignment of $b, a_1,\ldots, a_n$ to the free variables $x, y_1,\ldots, y_n$, where b is also an element of $\psi(H)$. As already mentioned, the last restriction means that we quantify only over the individuals existing in H.

Kripke illustrates the above definitions by giving counter-examples to two familiar formulas:

$$\forall x\Box F(x) \rightarrow \Box\forall xF(x);\ \Box\forall xF(x) \rightarrow \forall x\Box F(x)$$

The formula on the left is known as the *Barcan formula*, and that on the right as the *converse of the Barcan formula*. I will consider here only Kripke's counterexample to the converse of the Barcan formula. It is a model structure (G, K, R) where K consists of two worlds, the actual world G and a second world H. The accessibility relation R is the universal relation

$$R = \{(G, H), (H, G); (G, G); (H, H)\}.$$

The quantificational model structure on (G, K, R) is formed by endowing each possible world with its own domain. In the present case we take: $\varphi(G) = \{a, b\}$ and $\psi(H) = \{a\}$. Finally, to obtain a model, we have to define an extension of the predicate symbol P in each possible world. Following Kripke we let: $\varphi(P, G) = \{a, b\}$ and $\varphi(P, H) = \{a\}$.

The first observation is that $\forall xF(x)$ is true in both worlds, (relative to the empty assignment), that is, $\varphi(\forall xF(x), G) = T$ and $\varphi(\forall xF(x), H) = T$, given that any assignment of an element of $\psi(G)$ to x is a member of $\varphi(F, G)$ and similarly for $\varphi(\forall xF(x), H) = T$. Thus $\Box\forall xF(x)$ is true in G. On the other side, $\forall x\Box F(x)$ is true in G iff $\Box\psi(G)$ is true in G for every individual in G assigned to x iff $\Box\psi(G)$ is true in G when a is assigned to x and when b is assigned to x. The first claim holds iff a belongs to the extension of F in both G and in H. This is true. The second claims holds iff b belongs to the extension of F in both G and in H. This is false, given that b does not belong to the extension of F in H.

It is interesting to compare Kripke's treatment of the converse of the Barcan formulas to Hintikka's treatment of the counterpart of this formula in epistemic logic.

Hintikka (1962) shows that

$$K\forall xF(x) \rightarrow \forall xKF(x)$$

is self-sustainable by showing that its negation is indefensible. We follow Hintikka's argument (and his numbering) which is essentially the same as the earlier argument which established the self-sustainability of

$$\exists xK_aF(x) \rightarrow K_a\exists xF(x).$$

Suppose there is a model set μ in a model system such that

(134) $K_a\forall xF(x) \in \mu$ and

(135) $\neg\forall xKF(x) \in \mu$.

By a series of equivalent transformations, (135) is reduced to

(136) $\exists xP_a\neg F(x) \in \mu$

which, by the rules (C.Ep), implies

(137) $P_a\neg F(x/b) \in \mu$, for an individual constant b, and

(138) $\exists xK_a(x = b) \in \mu$.

From (137) we get

(139) $\neg F(x/b) \in \mu^*$

where μ^* is an epistemic alternative to μ, and from (138) we also get using (C.EK=EK=*)

(140) $\exists xK_a(x = b) \in \mu^*$

We apply (C.EK=) to (140) to obtain:

(141) $\exists x(x = b) \in \mu^*$.

From (134) we get

(142) $\forall xF(x) \in \mu^*$

and given (141), we can apply (C.Uep) and derive

(143) $F(x/b) \in \mu^*$.

We ended in a contradiction and conclude that the negation of the converse ofvthe Barcan formula is indefensible.

We notice that the contradiction is obtained by first deriving a substitutional instance $P_a\neg F(x/b) \in \mu$ of (136) from which we get that (the individual denoted by) b, introduced in μ, is not F in μ^*. On the other side, from (134) we know that all individuals in μ^* are F in μ^*. But b exists in μ^*, by (141), and therefore by instantiating with b in μ^* we get that b is F in μ^*.

In Kripke's setting one cannot get a contradiction by assuming that $\Box\forall xF(x)$ and the negation of $\forall x\Box F(x)$ (i.e., $\exists x\neg F(x)$) are true in G. A contradiction is avoided because (the individual denoted by) *b in G* (think of *G* as μ and of *H* as μ^*) does not have the property *F* in *H*, given that it does not exist in *H* (an individual which does not exist in a world cannot have a property at that world, because the extension of a predicate is formed only from the individuals existing at that world!) But in Hintikka's model systems, the rules (C.Ep), (C.EK=EK=*), and (C.EK=) have the consequence that (the individual denoted by) *b* in μ exists in μ^* and thereby falls under the incidence of the universal quantifier in μ^*.

We witness here, one more time, the difference that exists between Hintikka's treatment of epistemic notions (in the context of model sets), and Kripke's treatment of alethic notions (in a model-theoretical setting). The source of the difference does not lie, we would say, in the substitutional versus the Tarskian interpretation of quantifiers, but in the principle (C.EK=EK=*):

(C.EK=EK=*) If $\exists x K_a(b = x) \in \mu$, and μ^* is an epistemic alternative to μ with respect to *a*; then $\exists x K_a(b = x) \in \mu^*$

that, recalling our earlier discussion, Hintikka associates with his notion of 'knowing who'.

On a more critical note, let us note that Hintikka's argument depends on whether we accept his representation of "a knowing who b is" as $\exists x K_a(b = x)$.

It is not obvious to us that the latter is the correct representation of the former, as also Lemmon remaked in his review of *Knowledge and Belief*. We think that Hintikka was driven to this interpretation and to (C.EK=EK=*) by his substitutional interpretation of quantifiers, but we will not discuss this matter here. Instead we will stay content with the following remarks.

In his review of Hintikka (1962), Chisholm (1963) points out that Hintikka's idea of multiple reference pushes him towards metaphysics (essentialism), for it presupposes a method of *cross-identification* on the basis of which one would have to be able to establish when an individual in one world is the same as an individual in another world. Chisholm reviewed several criteria of crossidentifications, including essential properties, but did not find any of them fully acceptable. Chisholm (1967) ended up on a rather sceptical note: if we had a satisfactory answer to the question of *knowing who*, we would also have criteria to distinguish essential from non-essential properties.

Chisholm's criticisms (and similar criticisms coming from Castaneda motivated Hintikka to develop methods of cross-identification in the years to come. In Hintikka (1969) he introduces the distinction between *public* and *perspectival* identification. I may have heard of Barack Obama, know who he is (the President of US) but have never seen him. When I finally see him, I identify him perspectivally, that is, I place him on my visual map. Or, I may be in a situation in which I have seen him, but fail to associate him with Barack Obama, i.e. fail to identify him publicly. When this happens I know who Barack Obama is. Hintikka developed the distinction between "two modes of identification" in Hintikka (1969). Corresponding to two modes of identification, the representation in the logical language of "knowing who *b* is" bifurcates now into $\exists x K_a(b = x)$ (public) and $ExK_a(b = x)$ (perspectival).

Kripke and Hintikka on existence

Kripke (1963) considers the possibility of blocking the counterexamples to the Barcan formulas in alethic contexts by restricting the domains of the model structure. According to Kripke's proposal the counter-example to the converse of the Barcan formula could be blocked by requiring that whenever *HRH'*, we must also have $\psi(H) \subseteq (H')$. This type of solution leads (when the inclusion is also formulated for the other direction) to accepting that all actually existing individuals exist necessarily, a principle that Hintikka rejects. Hintikka, as we pointed out earlier, preferred to go for an alethic system with existential presuppositions, enforced through the principle (C.N*). Accepting this principle, however, leads to some restriction on the rule of substitution.

Perhaps with an eye on Hintikka's suggestion, Kripke (1963) considers also the possibility to introduce existence. He introduces existence as a predicate in a modal system (quantified M) based on a quantification theory in which, following a proposal by Quine, only closed formulae are asserted. (Kripke 1963, p. 89). We will not present here the axioms of quantified M. But it is worth mentioning that the existence predicate avoids the principle that everything exists necessarily that bothered Hintikka. Let us follow Kripke and see how.

The existence predicate $E(x)$ has to satisfy, for each model φ on a m.s. (G, K, R) the condition $\varphi(E, H) = (H)$, for each possible world *H*. In other words, everything in the domain of *H* exists. As Kripke remarks, this condition can be also given an axiomatic form, as the closures of formulae of the form: $\forall x A(x) \wedge E(y) \rightarrow A(y)$ and $\forall x E(x)$. But then "necessarily everything exists, $\Box \forall x E(x)$ becomes a theorem of the system. Yet, as Kripke shows, existence differs from the tautological predicate $A(x) \vee \neg A(x)$. This predicate is had by every individual necessarily, i.e. $\forall x \Box (A(x) \vee \neg A(x))$ is a theorem of the system, $\forall x \Box E(x)$ is not. (Kripke, 1963, p. 90.)

Kanger (1957)

In Kanger (1957) "a modification and extension of Tarski's theory was made with the purpose of obtaining semantics for modal formulas." (Kanger, 1972, p. 114)

That is, like Kripke, Kanger has a full blown model-theoretical treatment of modal notions, with "truth in a possible world (system)" as the basic semantical notion. He assigns extensions to predicate symbols relatively to a domain. These extensions may vary when we move from one domain to another. The various domains are to be thought of as possible worlds related by an accessibility relation. The main difference with Kripke (1959) and (1963), we would say, is the fact that the interpretation of a variable (in the standard sense) varies from a domain to another. In other words, variables are treated intensionally.

Less informally, a system is a triple (U, W, V) where:

- *U* is a universe (domain)
- *W* is a binary function which assigns extensions to predicate symbols in every universe, i.e., $W(P^n, U) \subseteq U^n$for every n-place predicate symbol and universe *U*.
- *V* is a binary function which assigns a value to every individual variable and universe *U*, that is, $V(x, U) \in U$ for every individual variable x and universe *U*.

Here are Kanger's clauses for the recursive truth-definition of the notion "*A* is true (false) in the system (U, W, V)", in symbols $T(A, (U, W, V)) = T$. The clauses for the non-modal formulas go like this:

(i) $T(x = y, (U, W, V)) = T$ iff $V(x, U) = V(y, U)$; otherwise it is *F*.

(ii) $T(R(x, y), (U, W, V)) = T$ iff $((V(x, y), V(y, U)) \in W(R, U)$; otherwise it is *F*

(iii) $T(\forall x F(x), (U, W, V)) = T$ iff $T(F(x), (U, W, V')) = T$ for every *V'* such that (a) $V'(y, U) = V'(y, U)$ for each *U* and each individual variable *y* other than *x*; and (b) $V'(x, U') = V(x, U')$ for every *U'* other than *U*

The truth-conditions for modal formulas are given using an accessibility relation over universes:

(iv) $T(\Box A, (U, W, V)) = T$ iff $T(A, (U', W, V)) = T$ for every universe *U'* such that *URU'*; otherwise it is *F*.

Kanger notices that by varying the properties of *R* one obtains truth-conditions for various modalities.

Few things need to be emphasized.

- the interpretation of a variable varies with possible worlds
- in the clause of the universal quantifier, the restriction (b) ensures that the interpretation of *x* is kept constant in all the other universes U' and thus the only variation in the value of the quantified variable *x* can come from V' assigning to *x* different individuals in the universe *U*. Thus this clause guarantees that the range of the universal and existential quantifier is the universe *U*.

Kanger notices that the Barcan formula

$\forall x \Box F(x) \to \Box \forall x F(x)$

does not hold in his system. We will not run through the technical argument, but mention instead the following informal considerations which illuminates the relationships with Kripke and Hintikka.

Suppose $\forall x \Box F(x)$ is true in the system (U, W, V). Then for every individual $a \in U$ which is assigned to *x*, *a* belongs to the extension of *F* in every alternative universe U': But this does not guarantee that in every alternative universe U' to *U*; all the individuals in U' are in the extension of *F* in U'.

On the other side, the converse of the Barcan formula

$$\Box \forall x F(x) \to \forall x \Box F(x)$$

holds. Recall Kripke's counter-example to this formula where we have two universes $U = \{a, b\}$ and $U' = \{a\}$ such that the extension of *F* in each of them is the whole universe and the accessibility relation is universal. We will content ourselves to point out why such a counterexample cannot arise in Kanger's setting.

Consider a system (U, W, V) such that *W* assigns to the predicate symbol *F* in the universe U the same extension F has in Kripke's example, that is, $W(P, U) = \{a, b\}$ and $W(P.U') = \{a\}$. Let *V* be such that $V(x, U) = b$ and $V(x, U') = a$ (recall the condition: $V(x, U) \in U$ for every *U*.) Any other *V* will do the job. It is easy to see that $\forall x F(x)$ is true in both systems (U, W, V) and (U', W, V) and thus by (iv), $\Box \forall x P(x)$ is true in (U, W, V). In order to show that $\Box \forall x F(x)$ is true in (U, W, V) consider any V' which satisfies conditions (a) and (b) of clause (iii). We are not interested in other variables than *x* and thus it is only condition (ii) that matters. We have only two possibilities. $V'(x, U) = V(x, U) = b$, which means, given condition (ii), that we must also have $V'(x, U') = V(x, U') = a$. The other possibility is $V''(x,U) = a$, which, by condition (ii), requires that we must also have $V''(x,U') = V(x, U') = a$. Now it is straightforward to check that $\Box F(x)$ is true in (U, W, V') and in (U, W, V''), that is, $F(x)$ is true in (U, W, V'), (U', W, V'), (U, W, V'') and (U', W, V''). For truth in (U, W, V'), notice that $V'(x, U) = b \in W(F, U) = \{a, b\}$; for truth in (U', W, V') we similarly notice that $V'(x, U') = a \in W(F, U') = \{a\}$; the last two cases are similar. We see that condition (b) is crucial here: it rules out the cases (such as Kripke's counter-example to the converse of the Barcan formula) in which an individual assigned to a free variable *x* does not exist in the universe *U* in which a formula containg an occurrence of *x* is evaluated. In Kanger's models whenever a formula is evaluated in a given universe, all the individual assigned to its free variables exist in that universe. One could perhaps say that the individuals assigned to a free variable in different universes play the role of that variable in those universes. And a similar conception could be defended for individual constants. In this case clause (i) would become $T(a = b, (U, W, V)) = T$ iff $V(a, U) = V(b, U)$ and we would say that $a = b$ is true in (U, W, V) if and only if the individual who plays the role of a in U is the same as the individual who plays the role of *b*.

References:

1. Carnap, R., 1946, Modalities and quantification, The Journal of Symbolic Logic, 11, pp. 33-64.
2. Carnap, R., 1947, Meaning and necessity, Chicago, 1947
3. Chisholm, R., 1963, 'The Logic of Knowing', Journal of Philosophy 60: 753–795
4. Chisholm, R., 1967, 'Identity through Possible Worlds: Some Questions', Noûs, 1(1): 1–8.
5. Hintikka, J., 1953, Distributive Normal Forms in the Calculus of Predicates, Acta Philosophica Fennica, reprinted as 'Distributive Normal Form in First-order logic' in J. N. Crossley and M. Dummett, eds. Formal Systems and Recursive Functions, pp. 47-90, 1964.
6. Hintikka, J., 1955, 'Two papers on symbolic logic', Acta Philosophica Fennica, 8.
7. Hintikka, J., 1957a, 'Quantifiers in Deontic logic. Commentationes humanarum litterarum' 23(4), 1-23, Societas Scientiarum Fennica, Helsinki.
8. Hintikka, J., 1957b, 'Modality as Referential Multiplicity', Ajatus 20, 49–64.
9. Hintikka, J.,1961, 'Modality and Quantification', Theoria 27, 119–128.
10. Hintikka, J., 1962/2005, Knowledge and Belief. An Introduction to the Logic of the Two Notions. Ithaca, NY: Cornell University Press. The references here are to the edition prepared by V. Hendriks and J. Symons, 2005.
11. Hintikka, J., 1967a, 'Existence and Identity in Epistemic Contexts', Theoria 33: 138–147.
12. Hintikka, J., 1967b, 'Individuals, Possible Worlds, and Epistemic Logic', Noûs 1: 33–62.
13. Hintikka, J., 1969, 'On the Logic of Perception' in J. Hintikka, Models for Modalities. Selected Essays, Dordrecht: Reidel, pp. 151–183.
14. Hintikka, J. and G. Sandu, 1995, 'The fallacies of the new theory of reference', Synthese, 104, Issue 2, pp 245–283.
15. Hintikka, J., 1998, Paradigms for Language Theory and Other Essays. Selected Papers, volume 4, Kluwer Academic publishers.
16. Kanger, S., 1957, 'New foundations of ethical theory, Stockholm, 1957, reprinted in Hilpinen 1971, pp. 36-58
17. Kripke, S., 1959, 'A completeness theorem in modal logic', The journal of symbolic logic, 24, pp. 1-14.
18. Kripke, S., 1963, 'Semantical considerations on modal logic', Acta philoosphica Fennica, 16, Helsinki, pp. 83-94.
19. Hilpinen, R., 1971 (ed.) Deontic Logic: Introduction and systematic readings, Dordrecht, Holland, 1971.
20. Leblanc, H., 1983, 'Alternatives to Standard First-Order Semantics', in D. Gabbay and F. Guenthner (eds.), Handbook of Philosophical Logic, vol. I, 189-274., D. Reidel Publishing Company.
21. Lewis, C. I., and Langford, C. H., 1932, Symbolic Logic, London: Century. 2nd edition 1959, New York: Dover.
22. Strawson, P. 1953, Reviewed Work: An Essay in Modal Logic by Georg H. von Wright, Philosophy Vol. 28, No. 104, pp. 76-79.
23. Von Wright, G.H., 1951, An Essay in Modal Logic, North-Holland Publishing Company, Amsterdam.

Received Editorial Board 10.04.17

Г. Санду, д-р філос. наук., проф.,
Т. Таннінен, асп.
Університет Гельсінкі, Гельсінкі, Фінляндія

ЗАРОДЖЕННЯ МОДАЛЬНОЇ ЛОГІКИ В ПІВНІЧНІЙ ЄВРОПІ

У статті досліджено генезис модальної логіки у роботах Г. Х. фон Врігта, Яакко Хінтики, Стіга Канжера та Саула Кріпке.

Ключові слова: Яакко Хінтика, модальна логіка, історія логіки.

Г. Санду, д-р филос. наук, проф.,
Т. Таннинен, асп.
Университет Хельсинки, Хельсинки, Финляндия

РОЖДЕНИЕ МОДАЛЬНОЙ ЛОГИКИ В СЕВЕРНОЙ ЕВРОПЕ

В статье исследуется генезис модальной логики в работах Г. Х. фон Вригта, Яакко Хинтикки, Стига Кангера и Саула Крипке.

Ключевые слова: Яакко Хинтика, модальная логика, история логики.