31

Deontic, Epistemic, and Temporal Modal Logics

RISTO HILPINEN

1 Modal Concepts

Modal logic is the logic of modal concepts and modal statements. Modal concepts (modalities) include the concepts of necessity, possibility, and related concepts. Modalities can be interpreted in different ways: for example, the possibility of a proposition or a state of affairs can be taken to mean that it is not ruled out by what is known (an *epistemic* interpretation) or believed (a *doxastic* interpretation), or that it is not ruled out by the accepted legal or moral requirements (a *deontic* interpretation), or that it has not always been or will not always be false (a *temporal* interpretation). These interpretations are sometimes contrasted with *alethic* modalities, which are thought to express the ways ('modes') in which a proposition can be true or false. For example, logical possibility and physical (real or substantive) possibility are alethic modalities.

The basic modal concepts are represented in systems of modal logic as propositional operators; thus they are regarded as syntactically analogous to the concept of negation and other propositional connectives. The main difference between modal operators and other connectives is that the former are not truth-functional; the truth-value (truth or falsity) of a modal sentence is not determined by the truth-values of its subsentences. The concept of possibility ('it is possible that' or 'possibly') is usually symbolized by $\Diamond$ and the concept of necessity ('it is necessary that' or 'necessarily') by $\Box$; thus the modal formula $\Diamond p$ represents the sentence form 'it is possible that p' or 'possibly p,' and $\Box p$ should be read 'it is necessary that p.' Modal operators can be defined in terms of each other: 'it is possible that p' means the same as 'it is not necessary that not-p'; thus $\Diamond p$ can be regarded as an abbreviation of $\neg\Box\neg p$, where $\neg$ is the sign of negation, and $\Box p$ is logically equivalent to $\neg\Diamond\neg p$. Systems modal propositional logic or quantification theory (predicate logic) are obtained by adding the symbols $\Diamond$ and $\Box$ (and possibly other modal signs), together with appropriate rules of sentence formation (e.g. if A is a formula, $\Diamond A$ and $\Box A$ are formulas), to a system of (non-modal) propositional logic or quantification theory.

2 The Semantics of Modalities and Systems of Modal Logic

As was observed above, modal sentences are not truth-functional: the truth-value of a modal sentence is not determined by the truth-values of its constituent. Given a true proposition p, 'it is necessary that p' may be true or false, depending on what p states (the *content* of p), and if p is false, 'possibly p' may be true or false, depending on the content of p. Consequently the logical relationships among modal propositions cannot be explained solely by means of possible truth-value assignments to simple (atomic) sentences, as in non-modal (truth-functional) propositional logic. A more complex semantics is needed. Since antiquity, modal concepts have been regarded as analogous to the quantifiers 'some' and 'all,' and modal propositions have been regarded as involving quantification over possible cases or possibilities of some kind. 'It is necessary that p' can be taken to mean that p is true (or it is true that p) no matter how things turn out to be, and 'it is possible that p' can be interpreted as saying that things may turn out to be or might have turned out to be in such a way that p is true. If the ways in which things can turn out to be are called *possible scenarios*, *situations*, or *possible worlds*, this account can be formulated as the standard possible worlds interpretation of modalities:

(CTN1) $\Box$p is true if and only if p is true in all possible worlds (situations),

and

(CTM1) $\Diamond$p is true if and only if p is true in some possible world (situation).

The possible worlds analysis of modalities goes back (at least) to the fourteenth century; for example, it seems to have been the basis of Duns Scotus's (1265–1308) modal theory (Knuuttila 1993: 143–5). G. W. Leibniz's use of the concept of possible worlds in the seventeenth century suggests a similar analysis, even though Leibniz himself did not analyze the concepts of necessity and possibility in this way. In the formal semantics of modal logic, the truth of a sentence is truth at (or relative to) a possible world, and modal formulas (sentences) are interpreted by means of a valuation function which assigns a truth-value to each sentence at each possible world. Non-modal propositional logic can be regarded as a limiting case in which only one possible world (the actual world) is considered.

In many applications of modal logic, the modal status of a given proposition depends on the situation in which it is evaluated. Many modal statements are contingent: what is possible or necessary depends on the point of evaluation. For example, what is epistemically possible for an individual depends on what the individual in question knows, and this varies from situation to situation. Thus the interpretation of modal sentences should also depend on a relation of *relative possibility* among worlds. The worlds which are possible relative to a given world (or situation) u are called the *alternatives to* u or worlds *accessible from* u. Consequently conditions (CTN1) and CTM1) should be reformulated as follows:

(CTN2) $\Box q$ is true at a world u if and only if q is true in all alternatives to u,

and

(CTM2) $\Diamond q$ is true at a world u if and only if q is true in some alternative to u.

The alternativeness relation was introduced into modal semantics in the 1950s by Marcel Guillaume (1958), Jaakko Hintikka (1957a, 1957b), Stig Kanger (1957), Saul Kripke (1963), Richard Montague (1960), and others. According to (CTN2) and (CTM2), an interpretation or a *model* of a (propositional) modal language is a triple M = ⟨W, R, V⟩, where W = {u,v,w, . . .} is a set of possible worlds (also called the *points* of the model), R is a two-place alternativeness relation defined on W, and V is an interpretation function or a *valuation function* which assigns to each sentence A a truth-value (1 for truth and 0 for falsity) at each possible world u. The pair ⟨W, R⟩ is called the *frame* of the model; thus a model consists of its frame and its valuation function. ‘V(A,u) = 1’ (the truth of A at u in M) is expressed ‘M, u ⊨ A,’ briefly ‘u ⊨ A;’ if A is not true at u, it is false at u. (I shall use below A, B, etc., as metalogical symbols which represent arbitrary formulas of a formal language of modal logic.) A sentence is called *valid* (logically true) if and only if it is true at every world u ∈ W for any interpretation M, and A is valid in a model M if and only if it is true at every point of the model. A sentence B is a logical consequence of A if and only if there is no interpretation M and world u such that M, u ⊨ A and not M, u ⊨ B. The valuation function is subject to the usual Boolean conditions which ensure that the truth-functional compounds of simple sentences receive appropriate truth-values at each possible world, in other words:

(C¬) $u \models \neg A$ if-if (if and only if) not $u \models A$,
(C&) $u \models A \,\&\, B$ if-if both $u \models A$ and $u \models B$,
(C∨) $u \models A \vee B$ if-if $u \models A$ or $u \models B$ or both, and
(C⊃) $u \models (A \supset B) = u \models \neg A$ or $u \models B$ or both.

The truth-conditions of simple modal sentences are expressed in terms of the alternativeness relation R as follows:

(CN) $u \models \Box A$ if and only if $v \models A$ for every $v \in W$ such that R(u,v),

and

(CM) $u \models \Diamond A$ if and only if $v \models A$ for some $v \in W$ such that R(u,v).

This semantics validates (for example) the following modal schemata:

(K) $\Box(A \supset B) \supset (\Box A \supset \Box B)$;
(2.1) $\Box(A \,\&\, B) \supset (\Box A \,\&\, \Box B)$; (The conjunctive distributivity of $\Box$.)
(2.2) $(\Box A \,\&\, \Box B) \supset \Box(A \,\&\, B)$; (The aggregation principle for $\Box$.)
(2.3) $\Box A \supset \Box(A \vee B)$;

(2.4) $\Box(A \supset B) \supset (\Diamond A \supset \Diamond B)$;
(2.5) $\Diamond A \supset \Diamond(A \vee B)$;
(2.6) $\Diamond(A \vee B) \supset (\Diamond A \vee \Diamond B)$; (The disjunctive distributivity of $\Diamond$.)
(2.7) $\Diamond (A \& B) \supset \Diamond A$.

This system, called the system K (from Kripke), can be characterized axiomatically by a set of axioms (or axiom schemata) for propositional logic, the Modus Ponens rule, the axiom schema (K) given above, the definition

(D$\Diamond$) $\Diamond A \equiv \neg\Box\neg A$,

and the modal 'rule of necessitation'

(RN) From A, to infer $\Box$A.

Rule (RN) means that if A is provable, so is $\Box$A.

The system K involves no assumptions about the structural properties of the alternativeness relation R. Different assumptions about the properties of R lead to different extensions of K, that is, systems of modal logic including K. For example, the assumption that R is a *serial* relation, that is satisfies the condition

(CD) For every $u \in U$, R(u,v) for some $v \in U$,

validates the principle that whatever is necessary is possible:

(D) $\Box A \supset \Diamond A$.

It is clear that this principle holds for most 'standard' concepts of necessity and possibility. A counterexample to this principle would be a situation in which both a proposition and its negation are necessary ($\Box A \& \Box\neg A$); most interpretations of modal expressions clearly exclude this. Alethic and epistemic modalities should also satisfy the schema

(T) $\Box A \supset A$,

which is equivalent to

$A \supset \Diamond A$.

Whatever is necessary is true, and a proposition cannot be known (to be true) in the proper sense of the word unless it is in fact true. Principle (T) distinguishes alethic and epistemic modalities from deontic and doxastic interpretations. It is true in all frames in which R is a reflexive relation, that is,

(CRefl) For every $u \in W$, R(u,u).

Moreover, if R is transitive,

(4S) $\Box A \supset \Box\Box A$

is valid, and the assumption that R is symmetrical validates the schema

(B) $\Diamond\Box A \supset A$.

The schema

(E) $\Diamond A \supset \Box\Diamond A$.

holds in all symmetrical and transitive frames. By making various assumptions about R it is thus possible to generate a great variety of modal systems. There is no single 'correct' system of a modal logic, but different systems are appropriate for different purposes and applications. Modal systems can be characterized semantically by the properties of the R-relation, and syntactically by their characteristic axioms (or axiom schemata), for example:

System KD (or briefly D): K + D;
System KT (briefly, T): K + T;
System KT4 (S4): KT + 4S; and
System KT5 (S5): KT + E or KT + 4S + B.

The expressions 'S4' and 'S5' are due to C. I. Lewis, who investigated in the 1910s the concept of strict (necessary) implication, and developed five alternative axiom systems for strict implication, S1–S5 (Lewis and Langford 1932). Lewis's system S4 can be characterized semantically by means of reflexive, and transitive frames, and the semantics of Lewis's S5 can explained by means of models in which R is an equivalence relation (a reflexive, transitive, and symmetric relation). (For different systems and interpretations of modal logic, see Chellas 1980: ch. 4; van Benthem 1988; Hughes and Cresswell 1996: 23–71.)

3 Modality and Quantification

The systems characterized above are systems of propositional logic. When modal operators are added to predicate logic (quantification theory), possible worlds can serve their interpretive function only if they are thought of as having a structure of individuals, properties, and relations. Thus the models of quantified modal logic provide, for each world w, the set D(w) of individuals existing in that world, and a valuation function which assigns an *extension* (an object, a set of objects, or a relation) to each nonlogical expression at each possible world. In other words, a valuation function assigns to each nonlogical expression a function from possible worlds to extensions. Such functions are called the *intensions* of individual terms, predicates, or relational expressions.

The truth conditions of the sentences of modal quantification theory can be interpreted and formulated in different ways. For quantifiers, perhaps the most natural choice is to let them range over the world-relative domains (rather than over all possible individuals). According to this approach, the semantic rules for $\forall$ and $\exists$ can be formulated, in a simplified and self-explanatory notation, as follows:

(CM$\forall$) M, u $\models \forall x A(x)$ if and only if for all individuals d $\in$ D(u), M, u $\models$ A(d),

and

(CM$\exists$) M, u $\models \exists x A(x)$ if and only if for some individual d $\in$ D(u), M, u $\models$ A(d).

The semantic rules for modalities must also be revised, and as in the case of the quantifier rules, different revisions are possible here. Perhaps the most reasonable interpretation of the necessity operator is to regard a sentence of the form $\Box A$ as true at u if and only if A is true in all alternatives to u whose domain contains the individuals denoted by the individual terms in A (including those assigned to individual variables) (van Benthem 1988: 16).

The validity of various principles involving modalities and quantifiers depends on the properties of the frames, in particular, on the relationships among the domains for different worlds. Of particular interest are in this context the following operator exchange principles:

(3.1) $\Box \forall x Ax \supset \forall x \Box Ax$
(3.2) $\forall x \Box Ax \supset \Box \forall x Ax$
(3.3) $\Box \exists x Ax \supset \exists x \Box Ax$
(3.4) $\exists x \Box Ax \supset \Box \exists x Ax$

If nothing is assumed about the domains of different possible worlds, only principle (3.1) is valid, and the rest of the formulas are invalid. However, formula (3.2) (called the Barcan formula, see Barcan (1946: 2)) is valid if the following inclusion principle holds for the domains D(u),

(3.5) If R(u,w), D(w) $\subseteq$ D(u),

and principle (3.4) is valid in all frames satisfying the condition

(3.6) If R(u, w), D(u) $\subseteq$ D(w).

The acceptability of (3.1)–(3.4) depends on the interpretation of the modal operators.

Above, the antecedents of (3.1) and (3.3) and the consequents of (3.2) and (3.4) are *de dicto* propositions, which means that the modal operator is attached to a complete proposition or *dictum*. The consequents (3.1) and (3.3) and the antecedents of (3.2) and (3.4) are called modal propositions *de re*: the modal operators are attached to expressions which contain a free individual term, thus the modality in question is ascribed to the object or thing (*res*) to which the term is regarded as being applicable. Sentences (3.1)–(3.4) describe possible relationships among *de dicto* and *de re* modalities.

4 Deontic, Epistemic, and Temporal Modalities

If modal propositions are understood in terms of the possible worlds semantics, their interpretation as deontic, epistemic, or temporal propositions depends on the interpretation of possible worlds and the alternativeness relation between possible worlds. It is often interesting to consider different (kinds of) modalities simultaneously; for example, a statement of the form

If it is true that p, it is possible to know that p,

or more briefly, 'if p is true, it is knowable,' contains an alethic concept of possibility and an epistemic modality ('to know'). An analysis of such sentences requires models which represent more than one concept of necessity and possibility, with a corresponding multitude of alternativess relations. In such situations different modalities (that is, different concepts of necessity and possibility) require special symbols. **O** and **P** are often used for deontic necessity (the concept of ought or obligation) and possibility (the concept of permissibility), the expressions $\mathbf{K}_i$ and $\mathbf{P}_i$ for the concept of (propositional) knowledge ('i knows that . . .') and the associated concept of epistemic possibility ('it is possible, for all that i knows, that . . .'), and $\mathbf{B}_i$ and $\mathbf{C}_i$ for the concepts of belief and doxastic possibility. (If only one person's knowledge or beliefs are being considered, the subscript can be omitted.) It is possible to define several temporal readings of □ and ◊, for example, 'it has always been the case that' and 'it was at some time the case that,' or 'it will always be the case that' and 'it will at some time be the case that.' The pairs of operators mentioned above are interdefinable in the same way as □ and ◊. The latter symbols are usually reserved for alethic modalities. (Sometimes alethic necessity is expressed by **N** or **L** and possibility by **M**.)

5 Epistemic Logic

The study of epistemic logic, like many other areas of philosophical logic, goes back (at least) to the late scholastic philosophy. Many fourteenth-century treatises on philosophical logic included a section on the logic of knowledge, often entitled *De scire et dubitare* ('On knowing and doubting'), which discussed sophisms and paradoxes involving the concepts of knowledge, belief, and doubt (Boh 1993: ch. 4). At the beginning of the twentieth century Charles Peirce analyzed the semantics of modal notions, and proposed an epistemic interpretation of modality, according to which a proposition is possible if and only if "it is not known to be false in a given state of information." Peirce distinguished this epistemic concept of possibility from what he called "substantive possibility" (alethic possibility), and regarded modalities as quantifiers over "possible cases" or "possible states of things" (Peirce 1931–35: vol. II, paragraph 2.347; vol. V, paragraphs 5.454–455). Peirce and his scholastic predecessors regarded epistemic concepts as modal concepts, but epistemic logic was not developed in a systematic way as a branch of modal logic before Jaakko Hintikka's *Knowledge and Belief* (1962), the first book-length study of the subject.

The epistemic alternatives to a given possible world (or knowledge situation) u are the worlds (situations) not ruled out by what is known (or by what a certain person knows) at u. The concept of doxastic alternativeness is related to the concept of belief in a similar way. The most obvious logical difference between the concepts of knowledge and belief is that the former should satisfy the T-axiom,

(KT) $\mathbf{K}A \supset A$,

in other words, epistemic alternativeness relations must be reflexive, but the T-principle does not hold for the concept of belief. In this respect doxastic modalities resemble deontic modalities. The assumption that the epistemic alternativeness relation is transitive validates the principle that knowing entails knowing that one knows (the KK-thesis),

(4SK) $\mathbf{K}A \supset \mathbf{KK}A$,

This thesis has sometimes been called, sometimes misleadingly, "the positive introspection axiom" (Fagin et al. 1995: 32). The transitivity of the epistemic R-relation means that sentences of the form **K**p can be transferred from a given world to its epistemic alternatives: if **K**p holds at u, then **K**p (and not only p) holds in the epistemic alternatives to u. The acceptability of the KK-thesis is sensitive to variations in the meaning of 'know.' The thesis has been part of many philosophers' conception of knowledge since antiquity, and it has sometimes been thought to characterize a "strong" concept of knowledge (knowledge based on conclusive grounds). On the other hand, if knowledge is regarded simply as true belief, the validity of the thesis depends on the validity of the corresponding thesis about belief,

(4SB) $\mathbf{B}A \supset \mathbf{BB}A$.

This principle seems to hold at least for some varieties of belief. It helps to understand G. E. Moore's paradox of "saying and disbelieving." If is obvious that a sentence of the form

(5.1) $p \,\&\, \neg\mathbf{B}p$

is not inconsistent, but a first-person utterance of (5.1) seems inconsistent or paradoxical. If the BB-schema is valid (i.e. if the doxastic alternativeness relation is transitive), the proposition

(5.2) $\mathbf{B}(p \,\&\, \neg\mathbf{B}p)$

is inconsistent, in other words, a person cannot sincerely assert (5.1) about oneself if sincere assertion is regarded as an expression of belief (Hintikka 1962: 64–9). If knowledge is regarded as true and conclusively justified belief, the KK-thesis means that a person knows that p only if he is also justified in claiming that he knows that p, in other words, the evidence for p is epistemically conclusive only if it justifies the corresponding knowledge-claim. The acceptance of this principle together with the epistemic versions of the rules and axioms of the modal system T amounts to the view that the logic of knowledge corresponds to the Lewis system S4. On the other hand, the epistemic versions of the modal axioms E and B do not seem to hold for the concept of knowledge: a person cannot be expected to be fully informed about his ignorance. The concept of belief obviously fails to satisfy principle T, but the doxastic counterpart of the principle D,

(DB) $\mathbf{B}A \supset \neg\mathbf{B}\neg A$,

should hold at least for the concept of consistent or rational belief.

The meaningfulness of quantifying into a modal context – that is, the interpretation of *de re* modal sentences – depends on the assumption that it is possible to make modal assertions about individuals (objects) independently of how they are described. For example, a sentence of the form

(5.3) $\exists x\Box Fx$

states that there is an individual which is F in all possible worlds (in which it exists). The existential quantifier identifies an individual across possible worlds or connects the 'appearances' of the same individual in different situations. Some philosophers have regarded such identifications as conceptually problematic. Epistemic modalities do not seem to be subject to such conceptual difficulties. The epistemic variant of (5.3),

(5.4) $\exists x\mathbf{K}_i Fx$,

says that some individual x is F in all situations not ruled out by (compatible with) what i knows in a given situation, in other words, someone (or something) is known (by i) to be F. This is of course quite different from saying that i knows that someone is F ($\mathbf{K}_i\exists xFx$). The latter sentence is true but the former false in a situation in which it is known that there are spies, but their identity is unknown – it is not known who they are. In ordinary language, (5.4) can be expressed by saying that i knows who is F. In the same way, the sentence

(5.5) $\exists x\mathbf{K}_i(x = c)$

can be taken to mean that i knows who c is (Hintikka 1989: 20). Some more complex sentences involving quantifiers and epistemic operators do not have any counterparts in the standard first-order modal quantification theory. For example,

(5.6) Alma knows whom everyone admires most,

where every person may admire a different person (for example, his or her mother) cannot be represented in standard first-order epistemic logic. The representation of such sentences requires second-order epistemic logic or an independence-friendly logic in which logical operators (for example, quantifiers and epistemic operators) can be independent of each other (see Hintikka 1989: 27–8).

Systems of epistemic logic based on S4, or any K-system, contain the rule of inference

(RNK) $\vdash A / \vdash \mathbf{K}A$,

where $\vdash$ is the sign of provability, as well as the rule

(RMK) $\vdash A \supset B / \vdash \mathbf{K}A \supset \mathbf{K}B$.

The validity of these rules creates 'the problem of logical omniscience' for epistemic logic: according to the epistemic interpretations of the K-systems, an inquirer knows the logical consequences of whatever he knows, and belief systems are closed with respect to logical deduction. These results motivated Hintikka's reinterpretation of the concept of logical consistency as (logical) defensibility or "immunity to [logical] criticism" (Hintikka 1962: 31), and I. Levi's interpretation of the logic of belief as the logic of doxastic (or epistemic) commitments (rather than "active" beliefs, Levi 1997). Another way to deal with the problem of logical omniscience is to place suitable syntactic restrictions on knowledge-preserving deductive arguments (Hintikka 1989). On the semantical (model-theoretic) side, similar results can be obtained by generalizing the concept of possible scenario or situation to 'seemingly possible' scenarios, represented by so-called urn models (Rantala 1975).

In the past 30 years, epistemic logic has developed into a relatively autonomous field of research, directed at problems and applications with no counterparts in other areas of modal logic (see Fagin et al. 1995; Meyer and van der Hoek 1995). Epistemic logic has been applied in interesting ways to philosophical semantics, epistemology and the philosophy of science. For example, it forms the logical basis of the interrogative theory of inquiry in which questions are treated as requests for knowledge or epistemic imperatives (Hintikka 1976, 1999).

6 Deontic Logic

The logic of normative concepts began to be investigated as a branch of modal logic in the fourteenth century, when some scholastic philosophers observed the analogies between deontic and alethic modalities, and studied the deontic (normative) interpretations of various laws of modal logic (Knuuttila 1993: ch. 5). In the seventeenth century, G. W. Leibniz (1930) called the deontic categories of the obligatory, the permitted, and the prohibited "legal modalities" ("Iuris modalia"), and observed that the basic principles of modal logic hold for the legal modalities. In fact, Leibniz suggested that deontic modalities can be analyzed in terms of the alethic modalities: he suggested that the permitted (*licitum*) is "what is possible for a good man to do," and the obligatory (*debitum*) "what is necessary for a good man to do." In the twentieth century the study of deontic logic as a branch of modal logic was initiated by Georg Henrik von Wright's pioneering work in the early 1950s (1951a, 1951b).

A simple system of deontic logic can be obtained by reading Leibniz's definition of the concept of obligation (ought) as

(O.Lbnz1) A is obligatory for b if and only if A is necessary for b's being a good person,

that is,

$\mathbf{O}_b A$ if and only if $\Box(G(b) \supset A)$,

where **O** is the deontic counterpart of □ and 'G(b)' means that b is 'good' (in the sense intended by Leibniz). If the explicit reference to an agent is deleted, we obtain the definition:

(O.Lbnz2) $\mathbf{O}A \equiv \Box(G \supset A)$.

The corresponding Leibnizian concept of permission is expressed by

(P.Lbnz2) $\mathbf{P}A \equiv \Diamond(G \,\&\, A)$.

These schemata can be regarded as partial reductions of deontic logic to alethic modal logic. In the twentieth-century deontic logic, the Leibnizian analysis of the concepts of obligation and permission was rediscovered by the Swedish philosopher Stig Kanger in 1950. Kanger (1971: 53) interpreted the constant G as "what morality prescribes." According to Kanger, **O**A (it ought to be the case that A) means that A follows from the requirements of morality.

If the alethic □-operator satisfies the axioms and the rules of inference of the modal system called KT (see above), the ought-operator defined by (O.Lbnz2) satisfies the deontic K-principle

(KD) $\mathbf{O}(A \supset B) \supset (\mathbf{O}A \supset \mathbf{O}B)$

and the rule of 'deontic necessitation'

(RND) $\vdash A \,/\, \vdash \mathbf{O}A$.

The additional assumption that being good is possible,

(DG) $\Diamond G$,

yields the deontic D-schema (the principle of deontic consistency),

(DD) $\mathbf{O}A \supset \neg\mathbf{O}\neg A$.

The system of (propositional) deontic logic obtained by adding to propositional logic the axiom schemata KD and DD and the rule RND is usually called the "standard system of deontic logic," abbreviated "SDL" (Føllesdal and Hilpinen 1971: 13–15). The theorems and the (derived) rules of inference of the standard system include the deontic variants of the schemata (1)–(7) and the rule

(RMD) $\vdash A \supset B / \vdash \mathbf{O}A \supset \mathbf{O}B$.

This modal system is often called the system KD or simply D (Chellas 1980: 114).

The sentences of SDL can be interpreted in terms of possible worlds (or world states) and an alternativeness relation between possible worlds in the same way as other modalities. The deontic alternatives to a given world u are worlds (or situations) in which everything that is obligatory at u is the case; thus the worlds related to u by R may be termed *deontically perfect* or *ideal* worlds (relative to u); they are worlds in which all obligations are fulfilled. If possible worlds are regarded as possible courses of events or histories which are partly constituted by an agent's actions, the semantics of SDL divides such histories into deontically acceptable and deontically unacceptable histories. An action is permitted if and only if it is part of some deontically acceptable course of events or if there is some deontically acceptable way of performing the action, and an action is obligatory if and only if no course of events is acceptable unless it exemplifies the action in question. The set of acceptable courses of action (relative to a given action situation) may be termed the *field of permissibility* (Lewis 1979). According to the deontic consistency principle (DD), the field of permissibility is never empty: some action is permissible in any situation. Additional structural assumptions about the R-relation validate further deontic principles. It is clear that sentences of the form

(6.1) $\mathbf{O}p \supset p$

are not logical truths, and therefore R cannot be regarded as a reflexive relation. However, the schema

(6.2) $\mathbf{O}(\mathbf{O}A \supset A)$

seems to hold for the concept of ought (or the concept of obligation): it ought to be the case that whatever ought to be the case is the case. The validity of (6.1) follows from the assumption that the deontic alternativeness relation is secondarily reflexive, in other words,

(C.OO) If R(u,v) for some u, then R(v,v).

SDL is quite a simple system, and cannot do justice to many complexities of normative discourse. This has been shown by various 'paradoxes' which result from attempts to formalize complex normative statements by means of SDL. (For discussions of the paradoxes of deontic logic, see Føllesdal and Hilpinen (1971: 21–6), and the articles in Hilpinen 1981). For example, SDL does not suffice for the representation of many *conditional* norms – and conditional norms abound in normative discourse. The following example about the inadequacy of SDL is analogous to an example given by Chisholm (1963); a situation of this kind is sometimes called 'Chisholm's paradox':

(Ch1) Bertie ought to confess.
(Ch2) Bertie ought to warn Corky if he is going to confess.
(Ch3) If Bertie does not confess, he ought not to warn Corky.
(Ch4) Bertie does not confess.

(Ch1)–(Ch4) seem to form a consistent set of (logically) mutually independent sentences, but in SDL they cannot be represented as such. If (Ch2) is represented as having the form

(6.3) $O(s \supset r)$,

where ‘s’ is taken to mean that Bertie confesses and ‘r’ means that Bertie warns Corky, (Ch1) and (Ch2) entail

(6.4) Or

If (Ch3) is regarded as having the same form as (Ch2), that is,

(6.5) $O(\neg s \supset \neg r)$,

it is (in SDL) a logical consequence of (Ch1), and if it represented as

(6.6) $\neg s \supset O\neg r$,

(Ch3) and (Ch4) entail

(6.7) $O\neg r$,

which, according to SDL, is inconsistent with (6.4); thus the choice of (6.6) as the representation of (Ch3) would make the set (Ch1)–(Ch4) inconsistent. On the other hand, if (Ch2) is formalized as

(6.8) $s \supset Or$,

it is a logical consequence of (Ch4), which is also unacceptable.

Sentence (Ch3) tells what Bertie ought to do in a situation where he has failed to fulfill his obligation to confess; thus it can be said to express a *contrary-to-duty* obligation (abbreviated ‘CTD’): Chisholm’s paradox may also be called the paradox of contrary-to-duty obligation.

7 Temporal Frames

Some authors have proposed to avoid the inconsistency of between (6.4) and (6.7) by relativizing the concept of obligation (or the concept of ought) time: it has been suggested that (6.4) and (6.7) hold at different points of time (Åqvist and Hoepelman 1981). It is obvious that what is obligatory or permitted changes over time; thus it is natural to assume, quite independently of the paradox of contrary-to-duty obligation, that deontic concepts should be analyzed by means of temporally structured systems of possible worlds, and that deontic logic should be based on tense logic (Thomason 1981, 1984; Horty 2001). The temporal structures required for the semantics of deontic modalities should involve a set W of world states or situations and a partial ordering $<$ on W such that for any u, v, $w \in W$, if $u < w$ and $v < w$, then $u < v$ or $v < u$ or $u = w$. The relation $<$ represents the temporal precedence among world-states. According to the $<$-relation, time has a branching, tree-like structure: each world-state has a unique past, but several possible futures. Temporal frames of this kind can also be used in epistemic logic for the representation of epistemic and doxastic changes. Maximal sets of linearly ordered world-states from W are called *histories* through the tree $T = (W, <)$; a set S is linearly ordered whenever for any u, v, $w \in S$, either $u < v$ or $v < u$ or $u = v$. Let H(u) be the set of histories that pass through u. The histories in H(u) represent the possibilities open in (or accessible from) the situation u. The truth-conditions of modal sentences can be defined for world-history pairs u/h such that $h \in H(u)$ (for details, see Thomason 1984; Horty 2001: ch. 2). For example, a temporal necessity operator $\Box$ and a future tense operator **F** can be defined in this framework as follows:

(CNtemp) $M, u/h \models \Box A$ iff $M,u/g \models A$ for every $g \in H(u)$;
(CFtemp) $M, u/h \models \mathbf{F}A$ iff $M,v/h \models A$ for some v such that $u < v$.

According to (CNtemp), it is clear that if there is an $h \in H(u)$ such that $\Box p$ holds at u/h, $\Box p$ holds at u/g for any history $g \in H(u)$; thus alethic modal sentences are *determinately* true or false at (temporary) world states or situations. The truth of $\Box p$ at u can be taken to mean the truth of p is settled or fixed at u, or that p is "settled true" at u (Horty 2001: 10). The deontic alternativeness relation R may be construed as a relation between a situation u and a history $g \in H(u)$: R(u,g) can be taken to mean that g is one of the deontically preferred or deontically acceptable histories passing through u. Relative to each situation u, the field of permissibility consists of the acceptable histories in H(u). The truth-conditions of O-sentences can be defined as follows:

(COtemp) $M,u/h \models \mathbf{O}A$ iff $M,u/g \models A$ for every g such that R(u,g).

According to (COtemp), p is obligatory in a given situation u if and only if p holds in every deontically acceptable history in H(u). Like alethic sentences, deontic sentences are determinately true or false at each $u \in W$. In interesting cases (e.g. in Chisholm-type examples) the proposition in the scope of **O** is not determinately true or false at the situation of evaluation, but refers to the future, for example to the options available to the agent (see Åqvist and Hoepelman 1981: 192). In the above example, (Ch1) (i.e. **O**s) and (6.4) hold as long as confessing is one of the options available to Bertie, but as soon this option is excluded and it is 'settled' that Bertie is not going to confess, (6.7) is true.

8 Conditional Obligations and Rules of Detachment

There are also non-temporal versions of the CTD-paradox. For example, consider the following example (due to Prakken and Sergot 1997): Assume that dogs are not permitted in a certain village, but if anyone has a dog, there ought to be a warning sign about it in front of the owner's house. Moreover, warning signs ought not to be posted without sufficient reason; thus there ought to be no warning sign if there is no dog. This example is formally analogous to Chisholm's example, and an attempt to formalize it in SDL leads to a similar inconsistency (Prakken and Sergot 1997; Carmo and Jones 2000).

The deduction of a contradiction from (6.4) and (6.7) depends on the principle of normative consistency (DD),

$\mathbf{O}A \supset \neg\mathbf{O}\neg A.$

This principle can be criticized independently of Chisholm's example: (DD) excludes the possibility of normative conflicts, but such conflicts are not unusual in morality and law, and it may be argued that they do not amount to paradoxes (Chellas 1974: 24). If the consistency principle is rejected, the deontic version of the aggregation principle (2.2),

$\mathbf{O}A \,\&\, \mathbf{O}B \supset O(A \,\&\, B),$

should be rejected as well, because the latter principle undermines the distinction between a conflict between obligations and the existence of a self-contradictory obligation. Normative conflicts can be distinguished from self-contradictory (impossible) obligations. Thus logicians have developed systems of deontic logic in which (DD) and the aggregation principle do not hold (Chellas 1980: 201–10, 272–5). Such systems represent CTD-situations as involving conflicting obligations, but they do not offer any analysis of CTD-obligations and their relationship to the 'primary' obligations.

As was observed above, the semantics of SDL is based on a division of worlds or situations into acceptable (deontically perfect) and unacceptable worlds, and the O-sentences describe how things are in the deontically faultless worlds. But sentence (Ch3) does not tell how things are in a deontically faultless world; it tells what the agent (Bertie) ought to do under imperfect conditions, that is, in situations in which Bertie does not act in accordance with his obligations. The situation could be described by saying that among the (less than ideal) scenarios where Bertie does not fulfill his obligation to confess, those in which he does not warn Corky are deontically preferable to the circumstances in which he (falsely) warns her. Thus Chisholm's example requires a distinction between different degrees of deontic perfection. (Ch2) can be taken to mean that in deontically perfect circumstances where Berie confesses, he warns Corky, and (Ch3) says that in the best worlds where he does not confess, he does not warn her (Hansson 1969). Let us express these conditional obligations by

(8.1) $O(r / s)$

and

(8.2) $O(\neg r / \neg s)$,

respectively. Let us call the worlds where p is true, 'p-words,' and let the p-worlds which are normatively least objectionable relative to a given situation u be called deontically optimal p-worlds relative to u. The concept of a deontically optimal p-world is a generalization of the concept of a deontically perfect world of SDL, and the assumption that for any consistent proposition p, there is a nonempty set of deontically optimal p-worlds, is a generalization of the SDL principle that any world has a nonempty set of deontic alternatives. The truth of a conditional ought-statement O(q / p) at u can be taken to mean that q is true in all deontically optimal p-worlds (relative to u). According to this interpretation of conditional obligations, the principle of 'deontic detachment,'

(DDet) $\mathbf{O}(B / A) \supset (OA \supset OB)$,

is a valid principle for conditional obligations, but the principle of 'factual detachment,'

(FDet) $\mathbf{O}(B / A) \supset (A \supset OB)$,

does not hold. If (Ch2) and (Ch3) are interpreted in this way, (Ch1)–(Ch4) do not lead to a contradiction: (Ch1) and (Ch2) entail the obligation **O**r, but (Ch3) and (Ch4) do not entail **O**¬r.

Chisholm's paradox can also be avoided by replacing the truth-functional conditional in (6.6) and (6.8) by an intensional (subjunctive) conditional without introducing a special concept of conditional obligation (Mott 1973). In the representation of our example in SDL, the logical asymmetry between (6.3) and (6.6) is required by the assumption of the logical independence of (Ch1)–(Ch4), and this leads to the inconsistency (6.4)–(6.7). If the two conditionals are expressed as intensional conditionals, this problem does not arise. An intensional conditional (e.g. a subjunctive conditional) 'q if p' can be regarded as true in a situation u if and only if q is true in all possible worlds (situations) in which p is true but which resemble u in other respects as much as possible (Lewis 1973). The truth of such a conditional is not a consequence of the falsity of p (or of the truth of q).

If 'q if p' is symbolized 'p > q,' and (Ch2) and (Ch3) are represented (respectively) by

(8.3) $s > Or$

and

(8.4) $\neg s > O\neg r$,

no contradiction will arise. If the *modus ponens* rule (the rule of factual detachment) holds for the conditional connective, (Ch3) and (Ch4) entail (6.7), but (Ch1) and (Ch2) do not entail (6.4). The former analysis of conditional obligations leads in our example to the result that Bertie ought to warn Corky, but the second analysis gives the result that Bertie ought not to warn Corky. Thus the two analyses involve two different concepts of ought (or 'obligation'): the first interpretation of (Ch1)–(Ch3) takes the statements in question as expressions of *prima facie*, defeasible (ideal or sub-ideal) obligations: (Ch1)–(Ch2) can be regarded as saying that in so far as Bertie ought to confess, he ought to warn Corky. On the other hand, if he is in fact not going to confess (or if this is regarded as being *settled*), he has an actual or practical ('all-out') obligation not to warn Corky; the second analysis concerns obligations of the latter type. The dependence of the latter type of obligation (ought) on the former presents an interesting problem for deontic logic and the theory of practical reasoning (Loewer and Belzer 1983). The paradoxes of conditional obligation and attempts to represent various CTD-obligations and other conditional obligations in formal systems of deontic logic have generated an extensive literature on the subject. (See Carmo and Jones 2000 and the articles in Nute 1997.)

As was observed above, deontic propositions are often future oriented and relative to time. This depends on another distinctive feature of deontic concepts, namely, that they are usually applied to acts, and acts normally involve change and take place in time. Philosophers and logicians have represented the concept of action in deontic logic in different ways (Hilpinen 1993, 1997). First, deontic modalities have been combined with action modalities, represented by modal operators which can be read 'i brings it about that p' or 'i sees to it that p' (Belnap 1991; Horty 2001). Another approach is to make a distinction between propositions, represented by propositional symbols, and actions, represented by action terms (action descriptions), and construe deontic concepts as operators which turn action terms into deontic propositions. The latter approach has been adopted in dynamic deontic logic (Segerberg 1982). Both approaches are based on temporal models involving temporally ordered world-states. Like epistemic logic, deontic logic has developed during the past 20–30 years into an autonomous discipline, with applications to computer science, legal informatics, moral philosophy, and other fields (see the papers in McNamara and Prakken 1999).

References

Åqvist, L. and Hoepelman, J. (1981) Some theorems about a "tree" system of deontic tense logic. In Hilpinen (1981), 187–221.

Barcan (Barcan-Marcus), R. C. (1946) A functional calculus of first order based on strict implication. *The Journal of Symbolic Logic* 11, 1–16.

Belnap, N. (1991) Backwards and forwards in the modal logic of agency. *Philosophy and Phenomenological Research*, 51, 8–37.

van Benthem, J. (1988) *A Manual of Intensional Logic*, 2nd edn. Menlo Park, CA: CSLI/SRI International.

Boh, I. (1993) *Epistemic Logic in the Later Middle Ages*. London and New York: Routledge.

Carmo, J. and Jones, A. J. I. (2000) Deontic logic and contrary-to-duties. In D. Gabbay (ed.) *Handbook of Philosophical Logic*, rev edn, vol. 4. Dordrecht: Kluwer Academic Publishers.

Chellas, B. (1974) Conditional obligation. In S. Stenlund (ed.), *Logical Theory and Semantic Analysis: Essays Dedicated to Stig Kanger on his Fiftieth Birthday*. 23–33. Dordrecht: D. Reidel.

Chellas, B. (1980) *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

Chisholm, R. M. (1963) Contrary-to-duty imperatives and deontic logic. *Analysis*, 24, 33–36.

Fagin, R., Halpern, J., Moses, Y. and Vardi, M. (1995) *Reasoning about Knowledge*. Cambridge, MA: MIT Press.

Føllesdal, D. and Hilpinen, R. (1971) Deontic logic: an introduction. In Hilpinen (1971), 1–35.

Guillaume, M. (1958) Rapports entre calculs propositionnels modaux et topologie impliqués par certaines extensions de la méthode des tableaux sémantiques. *Comptes rendus hebdomaires des sèances de l'Academie des Sciences (Paris)*, 246, 1140–42, 2207–2210; 247, 1282–83. Paris: Gauthiers-Villars.

Hansson, B. (1971) An analysis of some deontic logics. *Noûs* 3 (1969), 373–98. Reprinted in Hilpinen (1971), 121–47.

Hilpinen, R. (ed.) (1971) *Deontic Logic: Introductory and Systematic Readings*. Dordrecht: Reidel.

Hilpinen, R. (ed.) (1981) *New Studies in Deontic Logic: Norms, Actions and the Foundations of Ethics*. Dordrecht: Reidel.

Hilpinen, R. (1993) Actions in deontic logic. In J.-J. Ch. Meyer and R. J. Wieringa (eds.), *Deontic Logic in Computer Science: Normative System Specification* (pp. 85–100). Chichester and New York: John Wiley & Sons.

Hilpinen, R. (1997) On states, actions, omissions, and norms. In G. Holmström-Hintikka and R. Tuomela (eds.), *Contemporary Action Theory*, vol. I: *Individual Action* (pp. 83–108). Dordrecht: Kluwer Academic Publishers.

Hintikka, J. (1957a) Quantifiers in deontic logic. *Societas Scientiarum Fennica, Commentationes Humanarum Litterarum*, 23:4 (Helsinki).

Hintikka, J. (1957b) Modality as referential multiplicity. *Ajatus*, 20, 49–64.

Hintikka, J. (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, NY: Cornell University Press.

Hintikka, J. (1989) Reasoning about knowledge in philosophy: the paradigm of epistemic logic. In J. and M. Hintikka (eds.), *The Logic of Epistemology and the Epistemology of Logic* (pp. 17–35). Dordrecht and Boston: Kluwer Academic Publishers. Reprinted from *Reasoning about Knowledge: Proceedings of the 1986 Conference*. Los Altos, CA: Morgan Kaufmann, 1986, 63–8.

Hintikka, J. (1976) *The Semantics of Questions and the Questions of Semantics*, *Acta Philosophica Fennica*, 28:4 Helsinki: Societas Philosophica Fennica.

Hintikka, J. (1999) *Inquiry as Inquiry: A Logic of Scientific Discovery*. Dordrecht: Kluwer Academic Publishers.

Horty, J. (2001) *Agency and Deontic Logic*. Oxford and New York: Oxford University Press.

Hughes, G. E. and Cresswell, M. J. (1996) *A New Introduction to Modal Logic*. London and New York: Routledge.

Kanger, S. (1957) *Provability in Logic* (*Stockholm Studies in Philosophy* 1). Stockholm: Almqvist & Wiksell.

Kanger, S. (1971) New Foundations for Ethical Theory. In Hilpinen (1971), 36–58 (originally published in monographic form in 1957).

Knuuttila, S. (1993) *Modalities in Medieval Philosophy*. London and New York: Routledge.

Kripke, S. (1963) Semantical analysis of modal logic I. Normal modal propositonal calculi. *Zeitschrift für Mathematische Logic und die Grundlagen der Mathematik*, 9, 67–96.

Leibniz, G. W. (1930) *Elementa iuris naturalis*, in G. W. Leibniz, *Sämtliche Schriften und Briefe. Sechste Reihe: Philosophische Schriften* (pp. 431–85). Vol. 1. Darmstadt: Otto Reichl Verlag.

Levi, I. (1997) *The Covenant of Reason: Rationality and the Commitments of Thought*. Cambridge: Cambridge University Press.

Lewis, C. I. and Langford, C. H. (1932) *Symbolic Logic*. New York: Dover Publications.

Lewis, D. K. (1973) *Counterfactuals*. Oxford: Basil Blackwell.

Lewis, D. K. (1979) A problem about permission. In E. Saarinen et al. (eds.), *Essays in Honour of Jaakko Hintikka on the Occasion of His Fiftieth Birthday* (pp. 163–75). Dordrecht: Reidel.

Loewer, B. and Belzer, M. (1983) Dyadic deontic detachment. *Synthese*, 54, 295–318.

McNamara, P. and Prakken, H. (eds.) (1999) *Norms, Logics and Information Systems: New Studies in Deontic Logic and Computer Science*. Amsterdam: IOS Press.

Meyer, J. J. Ch. and van der Hoek, W. (1995) *Epistemic Logic for AI and Computer Science*. Cambridge: Cambridge University Press.

Montague, R. (1960) Logical necessity, physical necessity, ethics, and quantifiers. *Inquiry*, 4, 259–69. Reprinted in Montague, R. (1974), *Formal Philosophy*, R. H. Thomason (ed.). New Haven, CT: Yale University Press.

Mott, P. L. (1973) On Chisholm's paradox. *Journal of Philosophical Logic*, 2, 197–211.
Nute, D. (ed.) (1997) *Defeasible Deontic Logic*. Dordrecht and Boston: Kluwer Aacademic Publishers.
Peirce, C. S. (1931–35) *Collected Papers of Charles Sanders Peirce*, vols. I–VI, ed. C. Hartshorne and P. Weiss. Cambridge, MA: Harvard University Press.
Prakken, H. and Sergot, M. (1997) Deontic logic and contrary-to-duty obligations. In D. Nute (ed.), *Defeasible Deontic Logic* (pp. 223–62). Dordrecht: Kluwer Academic Publishers.
Rantala, V. (1975) Urn models: a new kind of non-standard model for first-order logic. *Journal of Philosophical Logic*, 4, 455–74.
Segerberg, K. (1982) A deontic logic of action. *Studia Logica*, 41, 269–82.
Thomason, R. (1981) Deontic logic founded on tense logic. In Hilpinen (1981), 165–76.
Thomason, R. (1984) 'Combinations of tense and modality. In D. Gabbay and F. Guenther (eds.), *Handbook of Philosophical Logic*, vol. II (pp. 135–65). Dordrecht: Reidel.
von Wright, G. H. (1951a) Deontic logic. *Mind*, 60, 1–15. Reprinted in G. H. von Wright, *Logical Studies*. London: Routledge and Kegan Paul, 1957, 58–74.
von Wright, G. H. (1951b) *An Essay in Modal Logic.* Amsterdam: North-Holland.