An Introduction to Modal Logic

Ordinary logic studies the partition of sentences[1] into two categories, true and false. Modal logic investigates a finer classification. A sentence can be either necessary (true, and it couldn't have been otherwise), contingently true (true, but it might have been false), contingently false (false, but it might have been true). or impossible (couldn't have been true). The informal study of the logical properties of modality goes back at least to Aristotle, but the advent of formal systems of symbolic modal logic dates from the publication of C. I. Lewis's *Survey of Symbolic Logic*[2] in 1918. Lewis started with the sentential calculus, and added symbols to represent "it is necessary that," "it is possible that," and "implies."[3] He then developed deductive calculi by adopting various appealing axioms and deriving their conseqeunces.

What redeemed formal modal logic from empty symbol pushing was the development of possible-world semantics. The idea comes from Leibniz, who thought of God as surveying all possible worlds, selecting the one that was best, and making it actual. The idea that this is the best of all possible worlds is a nutty idea, even by philosophers' standards – see Voltaire's *Candide* – but it proved fruitful. Possible-world semantics for modal sentential calculus were developed by J.C.C. MacKensie and Alfred Tarski in the 1940s,[4] and they were extended to

---

[1] I am restricting attention to sentences used to make assertion, setting aside sentences used to ask questions or make requests or issue promises. A general theory that encompasses these other ways of employing language is developed by John Searle in *Speech Acts* (Cambridge: Cambridge University Press, 1969).

[2] Berkeley, Calif.: University of California Press.

[3] Recall my polemics in Logic I against reading "→" as "implies." The confusion of "implies," which is a transitive verb, with a sentential connective originates with Alfred North Whitehead and Bertrand Russell's *Principia Mathematica* (Cambridge: Cambridge University Press, 1910). See W. V. Quine, "Reply to Professor Marcus," *Synthese* 20 (1961), reprinted in Quine, *The Ways of Paradox* (Cambridge, Mass.: Harvard University Press, 1966).

[4] "On Closed Elements in Closure Algebra," *Annals of Mathematics* 47 (1946): 122-162.

modal predicate calculus by Saul Kripke in the 1960s.[5] We'll only talk about modal sentential calculus here, and that only superficially.[6]

We start with a version of the sentential calculus with infinitely many atomic sentences and add an operator "□," read "it is necessary that"; if ϕ is a sentence, so is $\Box\phi$. We treat Lewis's symbol for possibility, "◊," as defined: $\Diamond\phi =_{Def} \sim \Box \sim \phi$. We'll have no need for Lewis's symbol for implication.

Our interest here is in the interpretation of the modal sentential calculus in which "□" means "It is provable that," so that "◊" means "It is consistent that."

A *Kripke model* is an ordered quadruple <W,R,I,a>, where W, the set of *worlds*, is a nonempty set; R, the *accessibility relation* is a binary relation on W; I, the *interpretation function,* is a function that assigns to each pair $<\phi$,w> with ϕ a sentence and w a world either the value 0 or the value 1; and a $\in$ W is the *actual world.* The triple <W,R,I> is a *frame*. For ϕ and atomic sentence and w a world, ϕ is true in w if and only if I(ϕ,w) = 1. A conjunction is true in w iff both conjuncts are true in w, a disjunction is true in w if and only if one or both disjuncts are true in w, and so on. $\Box\phi$ is true in w iff ϕ is true in every world v with Rwv. A sentence is *true* in the model iff it's true in a.

A sentence is *valid* for a frame or set of frames iff it's true at every world in every member of the set.

---

[5] "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica* 16 (1963): 83-94.

[6] For a fuller treatment, see Brian Chellas, *Modal Logic* (Cambridge: Cambridge University Press, 1980), G. E. Hughes and Max Cresswell, *A New Introduction to Modal Logic* (Routledge, 1996), or J. C. Beall and Bas van Fraassen, *Possibilities and Paradox* (Oxford: Oxford University Press, 2003).

A *normal modal system* is a set of sentences Γ with the following properties:

Tautological consequence: Every tautological consequence of Γ is in Γ.

Necessitation: If ϕ is in Γ, so is $\Box\phi$.

Schema (K): Each instance of the axiom schema

(K) $(\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi))$

is in Γ.

Theorem. For Γ a set of sentences, the following are equivalent:

(i) Γ is a normal model system.

(ii) There is a class of frames such that Γ is the set of sentences valid for every member of the class

(iii) Either Γ is the set of all sentences or there is a frame <W,R,I> such that Γ is the set of sentences valid for <W,R,I>. .

Proof: That (ii) implies (i) is easy to check. That (iii) implies (ii) is immediate; if Γ is the set of all sentences, our class of frames will be the empty class. So we only need to worry about showing that (i) implies (iii). Given Γ a normal modal system, let's say a set of sentences is *Γ-consistent* iff it contains all the members of Γ and it is consistent by the sentential calculus. A *maximal Γ-consistent* set is a Γ-consistent set such that, for every sentence, either the sentence of its negation is in the set. Let *W* be the class of all maximal Γ-consistent sets of sentences; unless Γ is the set of all sentences, *W* will be nonempty. For u and v elements of W, define Ruv iff ϕ is in v whenever $\Box\phi$ is in W. Define $I(\phi,w) = 1$ iff $\phi \in w$, for ϕ atomic. We want to show that, for any sentence ϕ, ϕ is true in w in the frame <W,R,I> iff $\phi \in w$. This will tell us that the folowing are equivalent:

$\phi \in \Gamma$

ϕ is an element of every maximal Γ-consistent set

ϕ is an element of every world w in W

ϕ is true in every world w in W

ϕ is valid for the frame <W,R,I>

The proof that, for any sentence ϕ, ϕ is true in w if and only if $\phi \in$ w proceeds by induction on the complexity of ϕ. The only part of this that isn't entirely routine is to show that $\Box\psi$ is true in w iff its a element of w. Here's the proof of the right-to-left direction: If $\Box\psi$ is an element of w, then, by definition of R, ψ is an element of every world accessible from w. It follows by inductive hypothesis that ψ is true in every world accessible from w, that is, that $\Box\psi$ is true in w.

For the other direction, suppose that $\Box\psi$ isn't an element of w. We want to see that there is a world accessible from w in which ψ isn't true. This means, according to the inductive hypothesis, that we want a world accessible from w that contains ψ. That is, given the definition of R, we want a maximal Γ-consistent set of sentences that includes all the sentences θ with $\Box\theta$ in w but that doesn't include ψ. To get this, it will suffice to show that $\Gamma \cup$ {sentences θ: $\Box\ \theta \in$ w} $\cup$ {$\sim \psi$} is tautologically consistent. If it is, we can expand $\Gamma \cup \{\theta: \Box\ \theta \in w\} \cup \{\sim\psi\}$ to a maximal Γ-consistent set by the familiar technique of marching through the sentences one by one, for each sentence when we come to it adding either it or its negation to the set, preserving Γ-consistency at every stage. Because of Necessitation, if γ is in Γ, $\Box\gamma$ is in Γ, and so $\Box\gamma$ is in w and γ is in $\{\theta: \Box\theta \in w\}$. So it will be enough to show that $\{\theta: \Box\theta \in w\} \cup z\{\sim\psi\}$ is tautologically consistent. If not, then there exist sentences $\theta_1, \theta_2,..., \theta_n$ such that each $\Box\theta$i is in

w and such that $(\theta_1 \rightarrow (\theta_2 \rightarrow ... \rightarrow (\theta_n \rightarrow \psi)...))$ is a tautology. It follows by Tautological Consequence and Necessitation that $\Box(\theta_1 \rightarrow (\theta_2 \rightarrow ... \rightarrow (\theta_n \rightarrow \psi)...))$ is a member of Γ, and hence, by multiple applications of Tautological Consequence and Schema (K), that $(\Box\theta_1 \rightarrow (\Box\theta_2 \rightarrow ... \rightarrow (\Box\theta_n \rightarrow \Box\psi)...))$ isn in Γ, and hence in w. Because w is closed under *modus ponens*, it follows that $\Box\psi$ is in w, contrary to our assumption.

The frame <W,R,I > that we just constructed is called the *canonical frame* for Γ. The principal moral of the theorem is that, if a sentence is outside Γ, then there is a world in the canonical frame in which it is false.☒

Let me write down some axioms schemata; the schemata were named by different people at different times, so the nomenclature is annoyingly haphazard:

(T) $(\Box\phi \rightarrow \phi)$

(4) $(\Box\phi \rightarrow \Box\Box\phi)$

(B) $(\phi \rightarrow \Box\Diamond\phi)$

(5) $(\Diamond\phi \rightarrow \Box\Diamond\phi)$

Let me also write down some notable properties of binary relations:

R is a *reflexive* relation on W iff, for each w in W, we have Rww.

R is *transitive* iff, for each u, v, and w, if Ruv and Rvw, then Ruw.

R is *symmetric* iff, for each u and v, if Ruv, then Rvu.

R is *Euclidean* iff, for each u, v, and w, if Ruv and Ruw, then Rvw.

K is defined to be the smallest normal modal system (that is, every other normal modal system includes K), so that a sentence is an element of K iff it is derivable from instances of schema (K) by the rules TC and Necessitation. A sentence is in K iff it is true in every world in

every frame. Why? The set of sentences valid for every frame is a normal modal system, so it includes K. If ϕ isn't in K, then there is a frame in which there is a world in which ϕ is false, namely, the canonical model for K.

KT is defined to be the smallest normal modal system that includes (T), so that a sentence is an element of KT iff it is derivable from (K) and (T) by the rules TC and Necessitation. A sentence is a element of KT iff it is true in every world in every reflexive frame. Why? Given a model <W,R,I,a>, with R reflexive, if $\Box\phi$ is true in a, then ϕ is true in every world accessible from a; in particular, ϕ is true in a itself; so all instances of schema (T) are true in the model. Consequently, the set of sentences valid for every reflexive frame is a normal modal system that includes (T). Moreover, the canonical frame for KT is reflexive; for any world w in the canonical frame, if $\Box\phi$ is in w, ϕ is in w, so we have Rww. Thus, if ϕ isn't in KT, then there is a reflexive frame in which there is a world in which ϕ is false, namely, the canonical frame for KT.

K4 is defined to be the smallest normal modal system that includes (4). A sentence is an element of K4 iff it's true in every world in every transitive frame. Why? Given a model <W,R,I,a> with R transitive, if $\Box\ \phi$ is true in a and w is a world accessible from a, then every world accessible from w is accessible from a.. Since ϕ is true in every world accessible from a, ϕ must be true in every world accessible from w, so that $\Box\ \phi$ is true in w. We have shown that $\Box\ \phi$ is true in every world accessible from a, so that $\Box\ \Box\ \phi$ is true in a. Thus we see that all instances of schema (4) are true in the model, so that the set of sentences true in every transitive model will be a normal modal system that includes (4). Moreover, the canonical frame for K4 is transitive. If u, v, and w are worlds in the canonical frame for K4 with Ruv and Rvw, then if $\Box\phi$ is in u, $\Box\ \Box\ \phi$ is in u, so that $\Box\phi$ is in v and ϕ is in w. Consequently, Ruw. Thus, if ψ is not in

KT, then the canonical frame for KT will be a transitive frame in which there is a world in which ϕ is false.

KB is the smallest normal system that contains (B). A sentence is in KB iff it's valid for the class of symmetric frames.

K5 is the smallest normal modal system that includes (5). A sentence is in K5 iff it's valid for the class of Euclidean frames.

KT4, which Lewis called "S4," is the smallest normal modal system that includes both (T) and (4). A sentence is is KT4 iff it's valid for the class of reflexive, transitive frames.

KTB is the smallest normal modal system that includes both (T) and (B) A sentence is in KTB iff it's valid for the class of reflexive, symmetric frames.

KT5, which Lewis called (S5), is the smallest normal modal system that includes both (T) and (5). A sentence is in KT5 iff it's valid for the class of reflexive, Euclidean frames. Since a binary relation that is reflexive and Euclidean will also be transitive and symmetric, KT5 is the same as KT4B5.

I could keep doing this for a long time, but you get the point.

The mystery component of the story is the accessibility relation. I have never heard a remotely satisfying explanation of why one possible world should or should not be accessible

from another.[7] As far as I can tell, the accessibility relation is something we tack on *ad hoc* so as to get the pretty relations between frames and axiom systems.

[7] We sometimes think if the march of history as following a forked path through a continually unfolding array of branching possibilities, so that what happens now can constrain what will be possible tomorrow. It could happen that, as of now, it's possible that I should fly to Jamaica tomorrow, but that some untoward event could happen tonight that would render it impossible for me to fly to Jamaica tomorrow; I might, for example, be eaten by a tiger escaped from a circus. So, even though it's not possible for me to fly to Jamaica tomorrow, it might become impossible for it to be possible for me to fly to Jamaica tomorrow, so that some instances of schema (5) can fail. To represent this conception formally, we take a "possible world" to be, not (as we would have expected) a possible complete course of history, but rather an ordered pair consisting of a complete course of history and a time. A statement will be possible at <h,t> if its truth is compatible with the course of history according to h as it has unfolded up to time t. A statement not involving modality will be true in <h,t> iff it's true in h. A pair <h',t'> will be accessible from <h,t> if t' is later than or equal to t and if h and h' agree in their depiction of the history of the world up to time t. The appropriate modal logic will be KT4.