LOGICS

JOHN NOLT
University of Tennessee, Knoxville

Modal Logics

Wadsworth Publishing Company
I(T)P® An International Thomson Publishing Company

Belmont, CA • Albany, NY • Bonn • Boston • Cincinnati • Detroit • Johannesburg
London • Madrid • Melbourne • Mexico City • New York • Paris • San Francisco
Singapore • Tokyo • Toronto • Washington

CHAPTER 11

LEIBNIZIAN MODAL LOGIC

11.1 MODAL OPERATORS

Prominent among philosophically important operators that are apparently inexpressible in predicate logic are **alethic** modifiers, such as 'must', 'might', 'could', 'can', 'have to', 'possibly', 'contingently', 'necessarily'. The term 'alethic' comes from the Greek word for truth, *alethea*. These words are said to express alethic modalities—that is, various modes of truth. **Modal logic,** in the narrowest sense, is the study of the syntax and semantics of these alethic modalities.

But the term is also used in a broader sense, to designate the study of other sorts of propositional modalities. These include **deontic** (ethical) **modalities** (expressed by such constructions as 'it ought to be the case that', 'it is forbidden that', 'it is permissible that', etc.); **propositional attitudes** (relations between sentient beings and propositions, expressed by such terms as 'believes that', 'knows that', 'hopes that', 'wonders whether', and so on); and **tenses** (e.g., the past, present, and future tenses as expressed by the various modifications of the verb 'to be': 'was', 'is', and 'will be').

The extension of the term 'modal' to these other forms of modality is no fluke; they share important logical properties with alethic modalities. For one thing, all of them can be regarded as operators on propositions. Consider, for example, these sentences, all of which involve the application of modal operators (in the broad sense) to the single proposition 'People communicate':

Alethic Operators

It is possible that people communicate.
It must be the case that people communicate.
It is contingently the case that people communicate.
It could be the case that people communicate.
It is necessarily the case that people communicate.

Deontic Operators

It is obligatory that people communicate.
It is permissible that people communicate.
It is not allowed that people communicate.
It should be the case that people communicate.

Operators Expressing Propositional Attitudes

Ann *knows that* people communicate.
Bill *believes that* people communicate.
Cynthia *fears that* people communicate.
Don *supposes that* people communicate.
Everyone *understands that* people communicate.
Fred *doubts that* people communicate.

Operators Expressing Tenses

It was (at some time) the case that people communicated.
It was always the case that people communicated.
It will (at some time) be the case that people communicate.
It will always be the case that people communicate.

There are, of course, many more operators in each category. And some of those listed, such as 'it is possible that' and 'it could be the case that' are, at least in some contexts, semantically identical or synonymous. With the exception of the operators expressing propositional attitudes, all of those listed here are monadic; they function syntactically just like the negation operator 'it is not the case that', prefixing a sentence to produce a new sentence. Thus, for example, the operators 'it is necessary that', usually symbolized by the box '□' and 'it is possible that', usually symbolized by the diamond sign '◇',[1] are introduced by adding this clause to the formation rules:

If Φ is a formula, then so are $\Box\Phi$ and $\Diamond\Phi$.

The operators expressing propositional attitudes, however, are binary. But unlike such binary operators as conjunction or disjunction, which unite a pair of sentences into a compound sentence, propositional attitude operators take a name and a sentence to make a sentence. The place for this name may be quantified, as in 'Everyone understands that people communicate'.

Many modal operators have **duals**—operators which, when flanked by negation signs, form their equivalents. The operators '□' and '◇', for example, are duals, as the following sentences assert:

$$\Box\Phi \leftrightarrow \sim\Diamond\sim\Phi$$
$$\Diamond\Phi \leftrightarrow \sim\Box\sim\Phi$$

That is, it is necessary that Φ if and only if it is not possible that not-Φ, and it is possible that Φ if and only if it is not necessary that not-Φ.

There are other duals among these operators as well. Consider the deontic operator 'it is obligatory that', which we shall symbolize as '**O**', and the operator 'it is permissible that', which we shall write as '**P**'. These are similarly related:

$$\mathbf{O}\Phi \leftrightarrow \sim\mathbf{P}\sim\Phi$$
$$\mathbf{P}\Phi \leftrightarrow \sim\mathbf{O}\sim\Phi$$

That '**O**' and '**P**' should thus mimic '□' and '◇' is understandable, since obligation is a kind of moral necessity and permission a kind of moral possibility.

There are also **epistemic** (knowledge-related) duals. The operator 'knows that' is dual with the operator 'it is epistemically possible, for . . . that'—the former representing epistemic necessity (knowledge) and the latter epistemic possibility. (Something is **epistemically possible** for a person if *so far as that person knows* it might be the case.) Symbolizing 'knows that' by '**K**' and 'it is epistemically possible for . . . that' by '**E**', we have:

$$p\mathbf{K}\Phi \leftrightarrow \sim p\mathbf{E}\sim\Phi$$
$$p\mathbf{E}\Phi \leftrightarrow \sim p\mathbf{K}\sim\Phi$$

In English: p knows that Φ if and only if it is not epistemically possible for p that not-Φ; and it is epistemically possible for p that Φ if and only if p does not know that not-Φ ('p', of course, stands for a person).

There are temporal duals as well. Let '**P**' mean 'it was (at some time) the case that' and '**H**' mean 'it has always been the case that'. Then:

$$\mathbf{H}\Phi \leftrightarrow \sim\mathbf{P}\sim\Phi$$
$$\mathbf{P}\Phi \leftrightarrow \sim\mathbf{H}\sim\Phi$$

[1] Sometimes 'L' is used instead of '□' and 'M' instead of '◇'. These abbreviate the German terms for logical (*logische*)—that is, necessary—truth and possible (*mögliche*) truth.

Here 'H' represents a kind of past tense temporal necessity and 'P' a kind of past tense temporal possibility. A similar relationship holds between 'it always will be the case that' and 'it sometimes will be the case that' and between other pairs of temporal operators.

These systematic logical relationships bear a striking resemblance to two familiar laws of predicate logic:

$\forall x\Phi \leftrightarrow \sim\exists x\sim\Phi$
$\exists x\Phi \leftrightarrow \sim\forall x\sim\Phi$

Are these pairs of dual operators somehow analogous to quantifiers?

11.2 LEIBNIZIAN SEMANTICS

Leibniz, who was among the first to investigate the logic of alethic operators, in effect suggested that they are. His semantics for modal logic was founded upon a simple but metaphysically audacious idea: Our universe is only one of a myriad possible universes, or possible worlds. Each of these possible worlds comprises a complete history, from the beginning (if there is a beginning) to the end (if there is an end) of time.

Such immodest entities may rouse skepticism, yet we are all familiar with something of the kind. I wake up on a Saturday; several salient possibilities lie before me. I could work on this book, or weed my garden, or take the kids to the park. Whether or not I do any of these things, my ability to recognize and entertain such possibilities is a prominent feature of my life. For ordinary purposes, my awareness of possibilities is confined to my doings and their immediate effects on the people and things around me. Yet my choices affect the world. If I spend the day gardening, the world that results is a different world than if I had chosen otherwise. Leibnizian metaphysics, then, can be seen as a widening of our vision of possibility from the part to the whole, from mere possible situations to entire possible worlds.

Possible worlds figure most notoriously in Leibniz's theodicy. God, in contemplating the Creation, surveyed all possible worlds, says Leibniz, and chose to actualize only the best—ours. Since ours is the best of all possible worlds, the degree of evil or suffering it contains is unavoidable—as we would see if only we had God's wisdom.[2]

What interests the logician, however, is not how Leibniz used possible worlds to rationalize actual miseries, but how he used them to adumbrate an alethic modal semantics. On Leibniz's view:

$\Box\Phi$ is true if and only if Φ is true in all possible worlds.

[2] This has given rise to the quip that the optimist is one who, like Leibniz, thinks that ours is the best of all possible worlds, whereas the pessimist is one who is sure of it.

and

$\Diamond\Phi$ is true if and only if Φ is true in at least one possible world.

The operators '$\Box$' and '$\Diamond$' are thus akin, respectively, to universal and existential quantifiers over a domain of possible worlds. So, for example, to say that it is necessary that $2 + 2 = 4$ is to say that in all possible worlds $2 + 2 = 4$; and to say that it is possible for the earth to be destroyed by an asteroid is to say that there is at least one possible world (universe) in which an asteroid destroys the earth.

Generalizing where Leibniz did not, we can extend his analysis to other modalities. Deontic operators are like quantifiers over morally possible (i.e., permissible) worlds—worlds that are ideal in the sense that within them all the dictates of morality are obeyed (exactly *which* morality is a question we shall defer!). Epistemic operators are like quantifiers over epistemically possible worlds—that is, over those worlds compatible with our knowledge (or, more specifically, with the knowledge of a given person at a given time). And tense operators act like quantifiers too—only they range, not over worlds, but over moments of time.

Time and possibility: an odd juxtaposition, yet illuminating, for there are rich analogies here. For one thing, just as there is a specific temporal moment, the present, which is in a sense uniquely real (for the past exists no longer, the future not yet), so there is a specific possible world, the actual world, which (for us at least) is uniquely real.

A second point of analogy is that in nonpresent moments objects have different properties from those they do now. I, for example, am now seated in front of a computer, whereas an hour or two ago I was riding my bike. Not all of what was true of me then is true of me now. In the same way, objects have properties different from those they actually have in nonactual worlds. I am a philosophy professor, but I could have been a farmer; that is, in some possible world I have the property of being a farmer, a property I do not actually have.

And just as an object (or a person) is typically not a momentary phenomenon, but has temporal duration—is "spread out," so to speak, through time—so too is an object "spread out" through possibilities. I am not just what I am at the moment; rather, I am an entire life, a yet-uncompleted history, from birth to death. Likewise, or so the analogy suggests, I am not merely what I actually am, but also my possibilities—what I could have been and could still be.[3]

Thus time and possibility share certain structural features, and their respective logics ought to reflect this fact. In Section 13.2 we shall see that to some extent they do. But in the meantime, we have run way ahead of Leibniz's conception of alethic modality. To Leibniz we now return, but with an anachronistic twist. We shall reformulate his insight about alethic operators in contemporary metatheoretic terms.

To begin, observe that a valuation for predicate logic in effect models a single world. It consists of a domain and assignments of appropriate extensions to pred-

[3] Cf. Martin Heidegger's contention that *Dasein* (human existence) *is* its possibilities and thus is more than it factually is; *Being and Time,* trans. John Macquarrie and Edward Robinson (New York: Harper & Row), pp. 68, 183–84, 185.

icates and names within that domain. In modal logic, we posit many possible worlds. A model for modal logic, then, should contain many "worlds," each with its own domain. And because the facts differ from world to world, that model should assign to each predicate not just a single extension, but an extension in each world. To keep things manageably (but preposterously) simple, consider a model representing just three possible worlds, w_1, w_2, and w_3. And (still oversimplifying) let's suppose that w_1 contains exactly four objects, α, β, γ, and δ; w_2 contains exactly two objects, β and γ; and w_3 contains exactly three objects α, δ, and ε:

World	**Domain**
w_1	{α, β, γ, δ}
w_2	{β, γ}
w_3	{α, δ, ε}

Now suppose we want to interpret the one-place predicate 'B', which for the sake of definiteness we may suppose means "is blue." Since a thing may be blue in one world but not in another, we cannot assign this predicate a single set (the set of blue things), as we would have in predicate logic. Rather, we need to assign it a separate set in—or "at" (either preposition may be used)—each world. For each world w, the set assigned to 'B' at w then represents the things that are blue in w. Suppose we assign to 'B' the set {α, β} in w_1, { } in w_2, and {α, δ, ε} in w_3. Then, according to our model there are two blue things in w_1 and none in w_2, and in w_3 everything is blue.

Because extensions differ from world to world (i.e., are world-relative) in modal logic, a valuation $\mathcal{V}$ now must take into account not only predicates, but also worlds, in assigning extensions. Thus we write

$\mathcal{V}(\text{'B'}, w_1) = \{\alpha, \beta\}$
$\mathcal{V}(\text{'B'}, w_2) = \{\ \}$
$\mathcal{V}(\text{'B'}, w_3) = \{\alpha, \delta, \varepsilon\}$

to indicate that at world w_1 the set of things that satisfies the predicate 'B' (i.e., the set of blue things) is {α, β}, and so on.

Truth, too, is now world-relative. Blue things exist in w_1 but not in w_2; thus the formula '∃xBx' ought to be true at w_1 but not at w_2. That is, $\mathcal{V}(\text{'}\exists x Bx\text{'}, w_1)$ = T, but $\mathcal{V}(\text{'}\exists x Bx\text{'}, w_2)$ = F. Accordingly, when we assign truth values to sentence letters, we shall have to assign each letter a truth value for each world. Let 'M', for example, mean "there is motion." We might let 'M' be true in w_1 but not in w_2 or w_3. Thus $\mathcal{V}(\text{'M'}, w_1)$ = T, but $\mathcal{V}(\text{'M'}, w_2) = \mathcal{V}(\text{'M'}, w_3)$ = F.

We shall assume, however, that names do not change denotation from world to world. Thus we shall assign to each name a single object, which may inhabit the domains of several possible worlds, and this assignment will not be world-relative. This models the metaphysical idea that people and things are "spread out" through possibilities, just as they are "spread out" through time. With respect to time, for example, the name 'John Nolt' refers to me now, but also to me when I was a child and to the old man whom (I hope) I will become. I occupy many

moments, and my name refers to me at each of these moments. Analogously, I have many possibilities, and my name refers to me in each. When I consider that I could be a farmer, part of what makes this possibility interesting to me is that it is *my* possibility.[4] It is I, John Nolt, who could be a farmer; my name, then, refers not only to me as I actually am, but to me as I could be. I am a denizen of possibilities (that is, possible worlds), as well as times, and my name tracks me through these possibilities, just as it does through the moments of my life.

Names, then, as we shall understand them, are **rigid designators;** they refer to the same object in each world in which they refer to anything at all. The idea that names designate rigidly, due to Ruth Marcus and Saul Kripke,[5] is now widely, though not universally, accepted. Other semantic interpretations of names have been offered, but we shall not consider them here.

In our semantics we shall model rigid designation by representing the value assigned to a name α simply as $\mathcal{V}(\alpha)$, rather than as $\mathcal{V}(\alpha, w)$, which would represent the value assigned to α *at a world w*. The omission of the world variable indicates that the denotations of names are not world-relative.

The concept of rigid designation harbors a metaphysical presupposition: the doctrine of **transworld identity.** This is the idea that the same object may exist in more than one possible world. It is modeled in our semantics by the fact that we allow the same object to occur in the domains of different worlds. Most logicians who do possible worlds semantics take transworld identity for granted, though there are exceptions.[6]

Though a rigidly designating name refers to the same object in different worlds, that object need not be "the same" in the sense of having the same properties. I would have quite different properties in a world in which I was a farmer, but I would still be the same person—namely, me.

These ideas are reflected in the model introduced above. Object β, for example, exists in w_1 and w_2. It therefore exhibits transworld identity. Moreover, it is in the extension of the predicate 'B' in w_1, but not in w_2. Thus, though it is the same object in w_1 as it is in w_2, it is blue in w_1 but not in w_2. If we think of w_1 as the actual world, this models the idea that an object that is actually blue nevertheless *could be* nonblue (it is capable, for example, of being dyed or painted a different color, yet retaining its identity).

Suppose now that we use the name 'n' to denote object β, that is, let $\mathcal{V}(\text{'n'})$ = β. (Note the absence of a world-variable here; the denotation of a rigidly designat-

4 Of course not all possibilities are *my* possibilities. In a world in which my parents had never met, I would never have existed, and the name 'John Nolt' would not refer to anything in that world (unless, of course, there were a different person with that name—but then the name would simply be ambiguous; that person would not be me). My existence, in other words, is contingent. In our models, this contingency is represented by the fact that an object need not occur in the domain of each world.

5 See Kripke's *Naming and Necessity* (Cambridge: Harvard University Press, 1972).

6 Most notably David Lewis, in "Counterpart Theory and Quantified Modal Logic," *Journal of Philosophy* 65 (1968): 113–26.

ing name, unlike truth or the denotation of a predicate, is not world-relative.) Then we would say that the statement 'Bn' ("n is blue") is true in w_1, but not in w_2, that is, $\mathcal{V}(\text{'Bn'}, w_1) = \text{T}$, but $\mathcal{V}(\text{'Bn'}, w_2) = \text{F}$.

But what are we to say about the truth value of 'Bn' in w_3, wherein β does not exist? Consider some possible (but nonactual) stone. Is it blue or not blue in the actual world? Both answers are arbitrary. Similarly, it seems arbitrary to make 'Bn' either true or false in a world in which 'n' has no referent.

This problem cannot be satisfactorily resolved without either abandoning bivalence (so that 'Bn', for example, may be neither true nor false) or modifying the logic of the quantifiers. The first approach is perhaps best implemented by means of supervaluations, which are discussed in Section 15.3; the second by free logics, which are covered in Section 15.1. Discussion of either method now would perhaps complicate things beyond what we could bear at the moment. We shall therefore leave the question unsettled.

Valuation rules 1 and 2 below give truth conditions for atomic formulas at a world only on the condition that the extensions of the names contained in those formulas are in the domain of that world. The truth conditions at w for atomic formulas (other than identities) that contain names which denote no existing thing at w are left unspecified. (Identity statements are an exception, since their truth conditions are not world-relative.) Our semantics, then, will be deficient in this respect, though still usable in other respects. The deficiency will be remedied in Chapter 15.

A valuation, or model, then, consists of a set of things called **worlds**, each with its own domain of objects. In addition, it assigns to each name an object from at least one of those domains, and it assigns to each predicate and world an appropriate extension for that predicate in that world. An object may belong to the domain of more than one world, but it need not belong to domains of all worlds. Two different worlds may have the same domain. The full definition is as follows:

DEFINITION A **Leibnizian valuation** or **Leibnizian model** $\mathcal{V}$ for a formula or set of formulas of modal predicate logic consists of the following:

1. A nonempty set $\mathcal{W}_{\mathcal{V}}$ of objects, called the **worlds** of $\mathcal{V}$.
2. For each world w in $\mathcal{W}_{\mathcal{V}}$ a nonempty set $\mathcal{D}_w$ of objects, called the **domain** of w.
3. For each name or nonidentity predicate σ of that formula or set of formulas, an extension $\mathcal{V}(\sigma)$ (if σ is a name) or $\mathcal{V}(\sigma, w)$ (if σ is a predicate and w a world in $\mathcal{W}_{\mathcal{V}}$) as follows:
   i. If σ is a name, then $\mathcal{V}(\sigma)$ is a member of the domain of at least one world.
   ii. If σ is a zero-place predicate (sentence letter), $\mathcal{V}(\sigma, w)$ is one (but not both) of the values T or F.
   iii. If σ is a one-place predicate, $\mathcal{V}(\sigma, w)$ is a set of members of $\mathcal{D}_w$.
   iv. If σ is an n-place predicate ($n > 1$), $\mathcal{V}(\sigma, w)$ is a set of ordered n-tuples of members of $\mathcal{D}_w$.

Given any valuation, the following valuation rules describe how truth and falsity are assigned to complex formulas:

Valuation Rules for Leibnizian Modal Predicate Logic

Given any Leibnizian valuation $\mathcal{V}$, for any world w in $\mathcal{W}_{\mathcal{V}}$:

1. If Φ is a one-place predicate and α is a name whose extension $\mathcal{V}(\alpha)$ is in $\mathcal{D}_w$, then
   $\mathcal{V}(\Phi\alpha, w) = \text{T}$ iff $\mathcal{V}(\alpha) \in \mathcal{V}(\Phi, w)$;
   $\mathcal{V}(\Phi\alpha, w) = \text{F}$ iff $\mathcal{V}(\alpha) \notin \mathcal{V}(\Phi, w)$.
2. If Φ is an n-place predicate ($n > 1$) and $\alpha_1, \ldots, \alpha_n$ are names whose extensions are all in $\mathcal{D}_w$, then
   $\mathcal{V}(\Phi\alpha_1, \ldots, \alpha_n, w) = \text{T}$ iff $<\mathcal{V}(\alpha_1), \ldots, \mathcal{V}(\alpha_n)> \in \mathcal{V}(\Phi, w)$;
   $\mathcal{V}(\Phi\alpha_1, \ldots, \alpha_n, w) = \text{F}$ iff $<\mathcal{V}(\alpha_1), \ldots, \mathcal{V}(\alpha_n)> \notin \mathcal{V}(\Phi, w)$.
3. If α and β are names, then
   $\mathcal{V}(\alpha = \beta, w) = \text{T}$ iff $\mathcal{V}(\alpha) = \mathcal{V}(\beta)$;
   $\mathcal{V}(\alpha = \beta, w) = \text{F}$ iff $\mathcal{V}(\alpha) \neq \mathcal{V}(\beta)$.

For the next five rules, Φ and Ψ are any formulas:

4. $\mathcal{V}(\sim\Phi, w) = \text{T}$ iff $\mathcal{V}(\Phi, w) \neq \text{T}$;
   $\mathcal{V}(\sim\Phi, w) = \text{F}$ iff $\mathcal{V}(\Phi, w) = \text{T}$.
5. $\mathcal{V}(\Phi \,\&\, \Psi, w) = \text{T}$ iff both $\mathcal{V}(\Phi, w) = \text{T}$ and $\mathcal{V}(\Psi, w) = \text{T}$;
   $\mathcal{V}(\Phi \,\&\, \Psi, w) = \text{F}$ iff either $\mathcal{V}(\Phi, w) \neq \text{T}$ or $\mathcal{V}(\Psi, w) \neq \text{T}$, or both.
6. $\mathcal{V}(\Phi \vee \Psi, w) = \text{T}$ iff either $\mathcal{V}(\Phi, w) = \text{T}$ or $\mathcal{V}(\Psi, w) = \text{T}$, or both;
   $\mathcal{V}(\Phi \vee \Psi, w) = \text{F}$ iff both $\mathcal{V}(\Phi, w) \neq \text{T}$ and $\mathcal{V}(\Psi, w) \neq \text{T}$.
7. $\mathcal{V}(\Phi \rightarrow \Psi, w) = \text{T}$ iff either $\mathcal{V}(\Phi, w) \neq \text{T}$ or $\mathcal{V}(\Psi, w) = \text{T}$, or both;
   $\mathcal{V}(\Phi \rightarrow \Psi, w) = \text{F}$ iff both $\mathcal{V}(\Phi, w) = \text{T}$ and $\mathcal{V}(\Psi, w) \neq \text{T}$.
8. $\mathcal{V}(\Phi \leftrightarrow \Psi, w) = \text{T}$ iff either $\mathcal{V}(\Phi, w) = \text{T}$ and $\mathcal{V}(\Psi, w) = \text{T}$, or $\mathcal{V}(\Phi, w) \neq \text{T}$ and $\mathcal{V}(\Psi, w) \neq \text{T}$;
   $\mathcal{V}(\Phi \leftrightarrow \Psi, w) = \text{F}$ iff either $\mathcal{V}(\Phi, w) = \text{T}$ and $\mathcal{V}(\Psi, w) \neq \text{T}$, or $\mathcal{V}(\Phi, w) \neq \text{T}$ and $\mathcal{V}(\Psi, w) = \text{T}$.

For the next two rules, $\Phi^{\alpha}/_{\beta}$ stands for the result of replacing each occurrence of the variable β in Φ by α, and $\mathcal{D}_w$ is the domain that $\mathcal{V}$ assigns to world w.

9. $\mathcal{V}(\forall_{\beta}\Phi, w) = \text{T}$ iff for all potential names α of all objects d in $\mathcal{D}_w$, $\mathcal{V}_{(\alpha,d)}(\Phi^{\alpha}/_{\beta}, w) = \text{T}$;
   $\mathcal{V}(\forall_{\beta}\Phi, w) = \text{F}$ iff for some potential name α of some object d in $\mathcal{D}_w$, $\mathcal{V}_{(\alpha,d)}(\Phi^{\alpha}/_{\beta}, w) \neq \text{T}$.
10. $\mathcal{V}(\exists_{\beta}\Phi, w) = \text{T}$ iff for some potential name α of some object d in $\mathcal{D}_w$, $\mathcal{V}_{(\alpha,d)}(\Phi^{\alpha}/_{\beta}, w) = \text{T}$;
    $\mathcal{V}(\exists_{\beta}\Phi, w) = \text{F}$ iff for all potential names α of all objects d in $\mathcal{D}_w$, $\mathcal{V}_{(\alpha,d)}(\Phi^{\alpha}/_{\beta}, w) \neq \text{T}$.

11. $\mathcal{V}(\Box\Phi, w) = \text{T}$ iff for all worlds u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\Phi, u) = \text{T}$;
$\mathcal{V}(\Box\Phi, w) = \text{F}$ iff for some world u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\Phi, u) \neq \text{T}$.
12. $\mathcal{V}(\Diamond\Phi, w) = \text{T}$ iff for some world u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\Phi, u) = \text{T}$;
$\mathcal{V}(\Diamond\Phi, w) = \text{F}$ iff for all worlds u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\Phi, u) \neq \text{T}$.

Since the valuation rules are a lot to swallow in one bite, we'll take the propositional fragment of the semantics by itself first and come back to the full modal predicate logic later. This simplifies the definition of a valuation considerably:

DEFINITION A **Leibnizian valuation** or **Leibnizian model** $\mathcal{V}$ for a formula or set of formulas of modal propositional logic consists of

1. A nonempty set $\mathcal{W}_{\mathcal{V}}$ of objects, called the **worlds** of $\mathcal{V}$.
2. For each sentence letter σ of that formula or set of formulas and each world w in $\mathcal{W}_{\mathcal{V}}$, an extension $\mathcal{V}(\sigma, w)$ consisting of one (but not both) of the values T or F.

Here worlds are like the (horizontal) lines on a truth table, in that each is distinguished by a truth-value assignment to atomic formulas—though not all lines of a truth table need be represented in a single model.

Consider, for example, the following valuation of the formula '(V ∨ W)' which we may suppose means "Sam is virtuous or Sam is wicked":

$\mathcal{W}_{\mathcal{V}} = \{1, 2, 3, 4\}$

$\mathcal{V}$('V', 1) = T	$\mathcal{V}$('W', 1) = F
$\mathcal{V}$('V', 2) = F	$\mathcal{V}$('W', 2) = F
$\mathcal{V}$('V', 3) = F	$\mathcal{V}$('W', 3) = T
$\mathcal{V}$('V', 4) = F	$\mathcal{V}$('W', 4) = T

The "worlds" here are the numbers 1, 2, 3, and 4. (In a model, it doesn't matter what sorts of objects do the modeling.) In world 1, 'V' is true and 'W' is false—that is, Sam is virtuous, not wicked. In world 2, Sam is neither virtuous nor wicked. And in worlds 3 and 4, Sam is wicked, not virtuous.[7] Our model represents the situation in which Sam is both virtuous and wicked as impossible, since this situation occurs in none of the four possible worlds. In other words, only three of the four lines of the truth table for 'V ∨ W' are regarded as possible. This is arguably appropriate, given the meanings we have attached to 'V' and 'W'.

[7] In a sense, world 4 is redundant, since from the point of view of our model it differs in no way from world 3. But this sort of redundancy is both permissible and realistic. It may, for example, represent the idea that world 4 differs from world 3 in ways not relevant to the problem at hand; for example, Sam may be a sailor in world 3 but not in world 4. Of course, if the model were truly realistic, it would contain many more worlds representing many such irrelevant differences, but we are simplifying.

To understand more about how this model works, we must consider the valuation rules for propositional modal logic (rules 4–8 and 11–12 above). According to rule 6, for example, the statement 'V ∨ W' has the value T in a world w if and only if either 'V' or 'W' has the value T in that world, and it is false otherwise. Thus this statement is true in worlds 1, 3, and 4, but false in world 2. The rules for the other truth-functional propositional operators ('~', '&', '→', and '↔') are all similarly relativized to worlds.

The real novelty, though, and the heart of Leibniz's insight, lies in rules 11 and 12. Consider, for example, the statement '□~(V & W)', which according to our interpretation means "it is necessarily the case that Sam is not both virtuous and wicked." According to rule 11, this formula is true at a given world w if and only if the statement '~(V & W)' is true in all worlds. Now in our model '~(V & W)' is in fact true in all worlds. For there is no world in which both 'V' and 'W' are true; hence by rule 5, 'V & W' is not true in any world, and so by rule 4, '~(V & W)' is true in each world. This means by rule 11 that '□~(V & W)' is true in every world.

Similarly, the statement '◇V' ("it is possible that Sam is virtuous") is true in all worlds. For consider any given world w. Whichever world w is, there is some world u (namely, world 1) in which 'V' is true. Hence by rule 12, '◇V' is true in w.

Notice also that since '◇V' is true in all worlds, it follows by another application of rule 11 that '□◇V' ("it is necessarily possible that Sam is virtuous") is true in all worlds. In fact, repeated application of rule 11 establishes that '□□◇V', '□□□◇V', and so on are all true at all worlds in this model. The following metatheorem exemplifies the formal use of modal semantics; use it as a model for Exercise 11.2.1:

METATHEOREM: For any world w of the model just described, $\mathcal{V}$('□◇V', w) = T.

PROOF: Let u be any world of this model, that is, $u \in \mathcal{W}_{\mathcal{V}}$. Since $\mathcal{V}$('V', 1) = T, it follows by rule 12 that $\mathcal{V}$('◇V', u) = T. Thus, for all u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}$('◇V', u) = T. Now let w be any world in $\mathcal{W}_{\mathcal{V}}$. It follows by rule 11 that $\mathcal{V}$('□◇V', w) = T. QED

Exercise 11.2.1

Consider the following model:

$\mathcal{W}_{\mathcal{V}} = \{1, 2, 3\}$

$\mathcal{V}$('P', 1) = T	$\mathcal{V}$('Q', 1) = F	$\mathcal{V}$('R', 1) = T
$\mathcal{V}$('P', 2) = F	$\mathcal{V}$('Q', 2) = F	$\mathcal{V}$('R', 2) = T
$\mathcal{V}$('P', 3) = T	$\mathcal{V}$('Q', 3) = T	$\mathcal{V}$('R', 3) = T

it. In the proof, the object which actually has the property F is object *d*. Since *d* has F in *w*, *d* has F in some possible world, i.e., possibly has F. It follows, then, that something possibly has F. This enables us to contradict the reductio hypothesis.

Our final metatheorem shows that from the fact that it is possible something is F, it does not follow that the world contains anything which itself is possibly F. Suppose, for example, that we admit that it is (alethically) possible that there are such things as fairies. (That is, there is a possible world containing fairies.) From that it does not follow that there is in the actual world anything which itself is possibly a fairy. The counterexample presented in the following metatheorem is a formal counterpart of this idea. Think of world 1 as the actual world, which (we assume) contains no fairies and world 2 as a world in which fairies exist. (The fairies are objects δ and ε.) Read the predicate 'F' as "is a fairy."

METATHEOREM: The sequent '◇∃*x*F*x* ⊢ ∃*x*◇F*x*' is invalid.

PROOF: Consider the following valuation $\mathscr{V}$ whose set $\mathscr{W}_{\mathscr{V}}$ of worlds is {1, 2}:

World	Domain
1	{α, β, χ}
2	{α, β, χ, δ, ε}

where

$\mathscr{V}$('F', 1) = { } $\mathscr{V}$('F', 2) = {δ, ε}

Now $\mathscr{V}$('◇∃*x*F*x*', 1) = T. For $\mathscr{V}_{(`a',\delta)}$('*a*')—that is, δ—is in the domain of world 2 and $\mathscr{V}_{(`a',\delta)} \in \mathscr{V}_{(`a',\delta)}$('F', 2) so that by rule 1, $\mathscr{V}_{(`a',\delta)}$('F*a*', 2) = T. Thus by rule 10, $\mathscr{V}$('∃*x*F*x*', 2) = T. And from this it follows by rule 12 that $\mathscr{V}$('◇∃*x*F*x*', 1) = T.

However, $\mathscr{V}$('∃*x*◇F*x*', 1) ≠ T, for there is no member ω of the domain of world 1 which is in the extension of the predicate 'F' in any world. Hence by rule 1 there is no world *u* in $\mathscr{W}_{\mathscr{V}}$, name *v* and object ω in the domain of world 1 such that $\mathscr{V}_{(v,\omega)}$(F*v*, *u*) = T. That is, by rule 12 there is no name *v* and object ω in the domain of world 1 such that $\mathscr{V}_{(v,\omega)}$(◇F*v*, 1) = T. So by rule 10, $\mathscr{V}$('∃*x*◇F*x*', 1) ≠ T.

Thus, since $\mathscr{V}$('◇∃*x*F*x*', 1) = T but $\mathscr{V}$('∃*x*◇F*x*', 1) ≠ T, we have a counterexample and the sequent is invalid. QED

Notice that in the proof of this theorem we avoided the question of predication for nonexisting objects (which we have left unsettled). In this case it is the question whether the objects δ and ε, which are fairies in world 2, are also fairies in world 1, where they do not exist. Our valuation rules do not answer this question, but the sequent '◇∃*x*F*x* ⊢ ∃*x*◇F*x*' is invalid regardless of how it is answered.

Exercise 11.2.2

Prove the following metatheorems using Leibnizian semantics for modal predicate logic:

1. The sequent 'P ⊢ ◇P' is valid.
2. The sequent '◇P ⊢ P' is invalid.
3. The sequent '◇(P & Q) ⊢ ◇P & ◇Q' is valid.
4. The sequent '◇P & ◇Q ⊢ ◇(P & Q)' is invalid.
5. Every sequent of the form Φ ⊢ □◇Φ is valid.
6. Every sequent of the form ◇□Φ ⊢ □Φ is valid.
7. For any formula Φ, if Φ is a valid formula, then so is □Φ.
8. Every formula of the form □Φ ↔ ~◇~Φ is valid.
9. Every sequent of the form □Φ ⊢ ◇Φ is valid.
10. Every sequent of the form □(Φ → Ψ) ⊢ (□Φ → □Ψ) is valid.
11. Every sequent of the form □(Φ → Ψ) ⊢ ~◇(Φ & ~Ψ) is valid.
12. Every sequent of the form ~◇(Φ & ~Ψ) ⊢ □(Φ → Ψ) is valid.
13. The sequent '□P, P→Q ⊢ □Q' is invalid.
14. Every formula of the form □α = α is valid.
15. Every sequent of the form ~α = β ⊢ □~α = β is valid.
16. Every sequent of the form ◇α = β ⊢ α = β is valid.
17. The sequent '∀*x*□F*x* ⊢ □∀*x*F*x*' is invalid.
18. Every sequent of the form ∀β□Φ ⊢ ∀βΦ is valid.

11.3 A NATURAL MODEL?

Our model theory (semantics) deepens our understanding of the alethic modal operators, though to get interesting results we have had to make a metaphysical assumption or two along the way. Still we have not learned much about possibility per se. The models we have so far considered are all wildly unrealistic—because they contain too few worlds; because these "worlds" are not really worlds at all, but numbers; because their domains are too small; and because we never really said what the objects in the domains were. In this section we seek a more realistic understanding of possibility by correcting these oversimplifications.

In Section 7.2 we noted that, although most of the models we encounter even in predicate logic are likewise unrealistic (being composed of numbers with artificially constructed properties and relations) we can, by giving appropriate meanings to predicates and names, produce a *natural model*. A natural model is a model whose domain consists of the very objects we mean to talk about and whose predicates and names denote exactly the objects of which they are true on their intended meanings. A natural model for geometry, for example, might have a domain of points, lines, and planes. A natural model for subatomic physics might have a domain of particles and fields.[11]

A natural model for modal discourse will consist of a set of possible worlds—genuine worlds, not numbers—each with its own domain of possible objects. And that set of worlds will be infinite, since there is no end to possibilities.

But what *is* a possible world?

Leibniz thought of possible worlds as universes, more or less like our own. But how much like our own? Can a universe contain just one object? There is no obvious reason why not. Can it contain infinitely many? It seems so; in fact, for the century or two preceding Einstein, many astronomers thought that the actual universe really did. We have already said that there is a possible world in which I am a farmer. Is there one in which I am a tree?

This is a question concerning my **essence,** that set of properties which a thing must have in order to be me. What belongs to my essence? Being a professor is pretty clearly *not* essential to me. What about being (biologically) human? There are fairy tales in which people are turned into trees and survive. Do these tales express genuine possibilities? Such questions have no easy answers. Perhaps they have no answers at all.

Philosophers who think that the nature of things determines the answers are **realists** about essence. Realists believe that essences independent of human thought and language exist "out there" awaiting discovery. (Whether or not we *can* discover them is another matter.) Opposed to the realists are **nominalists,** who think that essences—if talk about such things is even intelligible—are not discovered, but created by linguistic practices. Where linguistic practices draw no sharp lines, there are no sharp lines; so if we say increasingly outrageous things about me (I am a farmer, I am a woman, I am a horse, I am a tree, I am a prime number . . .), there may be no definite point at which our talk no longer expresses possibilities. For nominalists, then, it is not to be expected that all questions about possibility have definite answers. (Extreme nominalists deny that talk about possibility is even intelligible.) The realist-nominalist debate has been going on since the Middle Ages; and, though lately the nominalists have seemed to have the edge, the issue is not likely to be settled soon.

To avoid an impasse at this point, we shall invoke a distinction that enables us to sidestep the problem of essence. Whether or not it is **metaphysically possible** (i.e., possible with respect to considerations of essence) for me to be a tree, it *does* seem **logically possible** (i.e., possible in the sense that the idea itself—in this case the idea of my being a tree—embodies no contradiction). Contradiction is perhaps a clearer notion than essence; so let us at least begin by thinking of our natural model as modeling logical, not metaphysical, possibility.

In confining ourselves to logical possibility, we attempt to think of objects as essenceless. What sorts of worlds are possible now? It would seem that a possible

[11] These would be models for theories expressed in predicate logic, not necessarily in modal logic.

world could consist of any set of objects possessing any combination of properties and relations whatsoever.

But new issues arise. Some properties or relations are mutually contradictory. It is a kind of contradiction, for example, to think of a thing as both red and colorless. Similarly, it seems to be a contradiction to think of one thing as being larger than a second while the second is also larger than the first. But these contradictions are dependent upon the meanings of certain predicates: 'is red' and 'is colorless' in the first example; 'is larger than' in the second. They do not count as contradictions in predicate logic, which ignores these meanings (see Section 9.4).

If we count them as genuine contradictions, then we must deny, for example, that there are logically possible worlds containing objects that are both red and colorless. If we refuse to count them as genuine contradictions, then we must condone such worlds. In the former case, our notion of logical possibility will be the **informal** concept introduced in Chapter 1. In the latter, we shall say that we are concerned with purely **formal** logical possibility.

Only if we accept the purely formal notion of logical possibility will we count as a logically possible world any set of objects with any assignment whatsoever of extensions to predicates. If we accept the informal notion, we shall be more judicious—rejecting valuations which assign informally contradictory properties or relations to things. We shall still face tough questions, however, about what counts as contradictory. Can a thing be both a tree and identical to me? That is, are the predicates 'is a tree' and 'is identical to John Nolt' contradictory? The problem of essence, in a new guise, looms once again. Only by insisting upon the purely formal notion of logical possibility can we evade it altogether.

In the next chapter the lovely simplicity of Leibnizian semantics will be shattered, so we might as well allow ourselves a brief moment of logical purity now. Let's adopt, then, at least for the remainder of this section, the formal notion of logical possibility.

Now, take any set of sentences you like and formalize them in modal predicate logic. The natural model for these sentences is an infinite array of worlds. Any set whatsoever of actual and/or merely possible objects is a domain for some world in this array. The predicates of the formalization are assigned extensions in each such set in all possible combinations (so that each domain is the domain of many worlds). Among these domains is one consisting of all the objects that actually exist and nothing more. And among the various assignments of extensions to predicates in this domain is one which assigns to them the extensions they actually do have. This assignment on this domain corresponds to the actual world. (Other assignments over the same domain correspond to worlds consisting of the same objects as the actual world does, but differing in the properties those objects have or the ways they are interrelated.) If our discourse contains any names, on the intended interpretation these names name whatever objects they name in the actual world; but they track their objects (i.e., continue to name them) through all the possibilities in which they occur.

CHAPTER 12

KRIPKEAN MODAL LOGIC

12.1 KRIPKEAN SEMANTICS

There is among modal logicians a modest consensus that Leibnizian semantics accurately characterizes logical possibility, in both its formal and informal variants. As we saw in Section 11.3, however, this does not tell us all we would like to know about informal logical possibility, because Leibnizian semantics does not specify which worlds to rule out as embodying informal contradictions. (Is the concept of a dimensionless blue point, for example, contradictory? What about the concept of a God-fearing atheist? The concept of a largest number?) Still, the semantic rules of Leibnizian logic as laid out in Section 11.2 and the inference rules of Section 11.4 do arguably express correct principles of both formal and informal logical possibility.

But logical possibility, whether formal or informal, is wildly permissive. Things that are logically possible need not be metaphysically possible (i.e., possible when we take essence into account). And things that are metaphysically possible need not be physically possible (i.e., possible when we take the laws of physics into account). It seems both logically and metaphysically possible, for example, to accelerate an object to speeds greater than the speed of light. But this is not physically possible. Moreover, what is physically possible need not be practically possible (i.e., possible when we take actual constraints into account). It is physically

possible to destroy all weapons of war, but it may not (unfortunately) be practically possible. Logical, metaphysical, physical, and practical possibility are all forms or degrees of alethic possibility. And there are, no doubt, other forms of alethic possibility as well. Furthermore there are, as we saw earlier, various non-alethic forms of "possibility": epistemic possibility, moral permissibility, temporal possibility, and so on. Does Leibnizian semantics accurately characterize them all—or do some modalities require a different semantics?

Consider the metatheorem, proved in Section 11.2, that any sequent of the form $\Box\Phi \vdash \Phi$ is valid. This seems right for all forms of alethic possibility. What is logically or metaphysically or physically or practically necessary is in fact the case. There are corresponding principles in epistemic, temporal, and deontic logic:

Modality	Principle	Meaning
Epistemic	$sK\Phi \vdash \Phi$	*s* knows that Φ; so Φ
Temporal	$H\Phi \vdash \Phi$	It has always been the case that Φ; so Φ
Deontic	$O\Phi \vdash \Phi$	It is obligatory that Φ; so Φ

The first is likewise valid. But the temporal and deontic principles are invalid. What was may be no longer, and what ought to be often isn't. Both temporal logic and deontic logic, then, have non-Leibnizian semantics.

Or, to take a more subtle example, consider sequents of the form $\Box\Phi \vdash \Box\Box\Phi$, which are also valid on a Leibnizian semantics. Some variants of this principle in different modalities are given below:

Modality	Principle	Meaning
Alethic	$\Box\Phi \vdash \Box\Box\Phi$	It is necessary that Φ; so it is necessarily necessary that Φ[1]
Epistemic	$sK\Phi \vdash sKsK\Phi$	s knows that Φ; so s knows that s knows that Φ
Temporal	$H\Phi \vdash HH\Phi$	It has always been the case that Φ; so it has always been the case that it has always been the case that Φ
Deontic	$O\Phi \vdash OO\Phi$	It is obligatory that Φ; so it is obligatory that it is obligatory that Φ

[1] Necessity can be understood here in any of the various alethic senses—logical, metaphysical, physical, practical, and so on.

The temporal and alethic versions are plausible, perhaps; but the epistemic and deontic versions are dubious. The epistemic version expresses a long-disputed principle in epistemology; it seems, for example, to rule out unconscious knowledge. And the deontic version expresses a kind of moral absolutism: The fact that something ought to be the case is not simply a (morally) contingent product of individual choice or cultural norms, but is itself morally necessary. These are controversial theses. We should suspect a semantics that validates them.

In fact, Leibnizian semantics seems inadequate even for some forms of alethic modality. Consider the sequent '$P \vdash \Box\Diamond P$' with respect to physical possibility. (This sequent is valid given a Leibnizian semantics; see problem 5 of Exercise 11.2.2.)

What does it mean for something to be physically possible or physically necessary? Presumably, a thing is physically possible if it obeys the laws of physics and physically necessary if it is required by those laws. But are the laws of physics the same in all worlds? Many philosophers of science believe that they are just the regularities that happen to hold in a given world. Thus in a more regular world there would be more laws of physics, in a less regular world fewer. If so, then the laws of physics—and physical possibility—are world-relative.[2] Leibnizian semantics treats possibility as absolute; all worlds are possible from the point of view of each. But our present reflections suggest that physical possibility, at least, is world-relative.

To illustrate, imagine a world, world 2, in which there are more physical laws than in the actual world, which we shall call world 1. In world 2, not only do all of *our* physical laws hold, but in addition it is a law that all planets travel in circular orbits. (Perhaps some novel force accounts for this.) Now in our universe, planets move in either elliptical or circular orbits. Thus in world 1 it is physically possible for planets to move in elliptical orbits (since some do), but in world 2 planets can move only in circular orbits. Since world 2 obeys all the physical laws of world 1, what happens in world 2, and indeed world 2 itself, is physically possible relative to world 1. But the converse is not true. Because what happens in world 1 violates a physical law of world 2 (namely, that planets move only in circles), world 1 is not possible relative to world 2. Thus the very possibility of worlds themselves seems to be a world-relative matter!

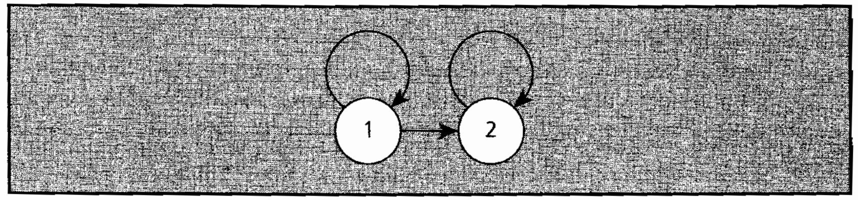
Kripkean semantics takes the world-relativity of possibility seriously. Within Kripkean semantics, various patterns of world-relativity correspond to different logics, and this variability enables the semantics to model a surprising variety of modal conceptions.

The fundamental notion of Kripkean semantics is the concept of **relative possibility** (which is also called **alternativeness** or **accessibility**). Relative possibility is the relation which holds between worlds x and y just in case y is possible relative to x. The letter '$\mathscr{R}$' is customarily used to express this relation in the metatheory. Thus we write

[2] I should confess that virtually everything I am saying here is controversial. But I have suppressed objections, not because I am confident that what I am saying here is true, but because I am trying to trace a line of thought that makes the transition from Leibnizian to Kripkean semantics intelligible. The metaphysics I spin out in the process should be regarded as illustration, not as gospel.

$$x\mathscr{R}y$$

to mean "y is possible relative to x" or "y is an alternative to x" or "y is accessible from x." (These are all different ways of saying the same thing.) So in the example just discussed it is true that $1\mathscr{R}2$ ("world 2 is possible relative to world 1"), but it is not true that $2\mathscr{R}1$. Each world is also possible relative to itself, since each obeys the laws which hold within it. Hence we have $1\mathscr{R}1$ and $2\mathscr{R}2$. The structure of this two-world model is represented in the following diagram, where each circle stands for a world and an arrow indicates that the world it points to is possible relative to the world it leaves:



A Kripkean model is in most respects like a Leibnizian model, but it contains in addition a specification of the relation $\mathscr{R}$—that is, of which worlds are possible relative to which. This is given by defining the set of pairs of the form $<x, y>$ where y is possible relative to x. In the example above, for instance, $\mathscr{R}$ is the set

$$\{<1, 2>, <1, 1>, <2, 2>\}$$

The definition of a Kripkean model mimics that of a Leibnizian model, with the addition of the requirement that $\mathscr{R}$ be defined (item 2 below):

DEFINITION A **Kripkean valuation** or **Kripkean model** $\mathscr{V}$ for a formula or set of formulas of modal predicate logic consists of the following:

1. A nonempty set $\mathscr{W}_{\mathscr{V}}$ of objects, called the **worlds** of $\mathscr{V}$.
2. A relation $\mathscr{R}$, consisting of a set of pairs of worlds from $\mathscr{W}_{\mathscr{V}}$.
3. For each world w in $\mathscr{W}_{\mathscr{V}}$ a nonempty set $\mathscr{D}_w$ of objects, called the **domain** of w.
4. For each name or nonidentity predicate σ of that formula or set of formulas, an extension $\mathscr{V}(\sigma)$ (if σ is a name) or $\mathscr{V}(\sigma, w)$ (if σ is a predicate and w a world in $\mathscr{W}_{\mathscr{V}}$) as follows:
   - i. If σ is a name, then $\mathscr{V}(\sigma)$ is a member of the domain of at least one world.
   - ii. If σ is a zero-place predicate (sentence letter), $\mathscr{V}(\sigma, w)$ is one (but not both) of the values T or F.
   - iii. If σ is a one-place predicate, $\mathscr{V}(\sigma, w)$ is a set of members of $\mathscr{D}_w$.
   - iv. If σ is an n-place predicate ($n > 1$), $\mathscr{V}(\sigma, w)$ is a set of ordered n-tuples of members of $\mathscr{D}_w$.

The addition of $\mathcal{R}$ brings with it a slight but significant change in the valuation rules for '□' and '◇'. Necessity at a world w is no longer simply truth in all worlds, but truth in all worlds that are possible *relative to w.* Likewise, possibility in w is truth in at least one world that is possible *relative to w.* Thus, instead of the valuation rules 11 and 12 for Leibnizian semantics (Section 11.2), Kripkean semantics has the modified rules:

11' $\mathcal{V}(\Box\Phi, w) = \text{T}$ iff for all worlds u such that $w\mathcal{R}u$, $\mathcal{V}(\Phi, u) = \text{T}$;
$\mathcal{V}(\Box\Phi, w) = \text{F}$ iff for some world u, $w\mathcal{R}u$ and $\mathcal{V}(\Phi, u) \neq \text{T}$.
12' $\mathcal{V}(\Diamond\Phi, w) = \text{T}$ iff for some world u, $w\mathcal{R}u$ and $\mathcal{V}(\Phi, u) = \text{T}$;
$\mathcal{V}(\Diamond\Phi, w) = \text{F}$ iff for all worlds u such that $w\mathcal{R}u$, $\mathcal{V}(\Phi, u) \neq \text{T}$.

No other valuation rules are changed.

Consider now a Kripkean model for propositional logic (which allows us to ignore the domains of the worlds), using the sentence letter 'P', which we interpret to mean "Planets move in elliptical orbits." Let $\mathcal{W}_\mathcal{V}$ be the set {1, 2} and $\mathcal{R}$ be the set

{<1, 2>, <1, 1>, <2, 2>}

as mentioned and diagramed in the example recently discussed. Suppose further that

$\mathcal{V}$('P', 1) = T
$\mathcal{V}$('P', 2) = F

as in that example. (That is, planets move in elliptical orbits in world 1 but not in world 2.) Now the sequent 'P ⊢ □◇P', which was valid on Leibnizian semantics, is invalid on this Kripkean model. For $\mathcal{V}$('P', 1) = T, but $\mathcal{V}$('□◇P', 1) ≠ T. That is, world 1 provides a counterexample.

We can see that $\mathcal{V}$('□◇P', 1) ≠ T as follows. Note first that the only world in $\mathcal{W}_\mathcal{V}$ accessible from world 2 is 2 itself; in other words, the only world u in $\mathcal{W}_\mathcal{V}$ such that $2\mathcal{R}u$ is world 2. Moreover, $\mathcal{V}$('P', 2) ≠ T. Hence for all worlds u in $\mathcal{W}_\mathcal{V}$ such that $2\mathcal{R}u$, $\mathcal{V}$('P', u) ≠ T. So by rule 12', $\mathcal{V}$('◇P', 2) ≠ T. Therefore, since $1\mathcal{R}2$, there is some world x in $\mathcal{W}_\mathcal{V}$ (namely, world 2) such that $1\mathcal{R}x$ and $\mathcal{V}$('◇P', x) ≠ T. It follows by rule 11' that $\mathcal{V}$('□◇P', 1) ≠ T. We restate this finding as a formal metatheorem:

METATHEOREM: The sequent 'P ⊢ □◇P' is not valid on Kripkean semantics.

PROOF: As given above.

Moreover, neither of the other sequents mentioned in this section—'□P ⊢ P' and '□P ⊢ □□P'—is valid, either. Let's take '□P ⊢ P' first.

METATHEOREM: The sequent '□P ⊢ P' is not valid on Kripkean semantics.

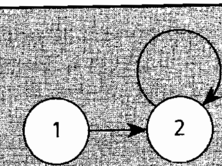
PROOF: Consider the following Kripkean model for propositional logic. Let the set $\mathcal{W}_\mathcal{V}$ of worlds be {1, 2} and the accessibility relation $\mathcal{R}$ be the set {<1, 2>, <2, 2>}, and let

$\mathcal{V}$('P', 1) = F
$\mathcal{V}$('P', 2) = T

Now $\mathcal{V}$('P', 2) = T and 2 is the only world possible relative to 1; that is, 2 is the only world u such that $1\mathcal{R}u$. Hence for all worlds u such that $1\mathcal{R}u$, $\mathcal{V}$('P', u) = T. Therefore by rule 11', $\mathcal{V}$('□P', 1) = T. But $\mathcal{V}$('P', 1) ≠ T. Therefore '□P ⊢ P' is not valid on Kripkean semantics. QED

This result poses a problem. Intuitively, '□P ⊢ P' is (or ought to be) valid on the alethic and epistemic interpretations. But it should not come out valid on the deontic interpretation (which, to distinguish it from the other interpretations, we usually write as '**O**P ⊢ P') or on the temporal interpretation discussed above.

The reasoning for the deontic interpretation is straightforward. Think of world 1 as the actual world, world 2 as a morally perfect world, and 'P' as expressing the proposition "Everything is morally perfect." Then, of course, 'P' is true in world 2 but not in world 1. Moreover, think of $\mathcal{R}$ as expressing the relation of permissibility or moral possibility. Now world 2 is morally permissible, both relative to itself and relative to world 1 (because what is morally perfect is surely morally permissible!). But world 1 is not morally permissible, either relative to itself or relative to world 2, because all kinds of bad (i.e., morally impermissible) things go on in it. Our model, then, looks like this:



Now since in this model every world that is morally permissible relative to the actual world is morally perfect (since there is, in the model, just one such world, world 2), it follows (by the semantics for '□', i.e., formally, rule 11') that it ought to be the case in world 1 that everything is morally perfect, even though that is not the case in world 1. Thus, when we interpret '□' as "it ought to be the case that,"[3] we can see how '□P ⊢ P' can be invalid. Kripkean semantics, then, seems

[3] We could, of course, have used the symbol '**O**' instead of '□' to express the deontic reading, but we are considering several different readings simultaneously here.

right for the deontic interpretation, but wrong for the epistemic, temporal, and alethic interpretations.

But in fact Kripkean semantics is applicable to the other interpretations, as well, provided that we are willing to relativize our concept of validity. The key to this new conception can be found by reexamining the proof from an alethic viewpoint. From this viewpoint the proof is just wrong. Surely, if it is alethically necessary that *P*, then *P*. But where is the mistake?

It lies, from the alethic point of view, in the specification of $\mathscr{R}$. The alethic sense of possibility requires that *every* world be possible relative to itself, for what is true in a world is certainly alethically possible in that same world. But the relation $\mathscr{R}$ used in the proof does not hold between world 1 and itself. The model is therefore defective from an alethic point of view.

To represent the alethic intepretation, we must insist that $\mathscr{R}$ be **reflexive**—that each world in the set $\mathscr{W}_{\mathscr{V}}$ of worlds be possible relative to itself. Thus the model given above as a counterexample is not legitimate for the alethic interpretation. *The only admissible models—the only models that count—for the alethic interpretation are models whose accessibility relation is reflexive.* This is also true for the epistemic modalities, but not for the deontic or temporal ones.

This suggests the following strategy: Each of the various modalities is to be associated with a particular set of admissible models, that set being defined by certain restrictions on the relation $\mathscr{R}$. Validity, then, for a sequent expressing a given modality is the lack of a counterexample among *admissible* models for the particular sorts of modal operators it contains. Other semantic notions (consistency, equivalence, and the like) will likewise be defined relative to this set of admissible models, not the full range of Kripkean models. In this way we can custom-craft a different semantics for each of the various modalities.

Let us, then, require admissible models for alethic or epistemic modalities, but not for the deontic or temporal ones, to be reflexive. Then we must redefine the notion of a valid sequent as follows:

> A sequent is **valid relative to a given set of models (valuations)** iff there is no model in that set containing a world in which the sequent's premises are true and its conclusion is not true.

To say that a sequent is valid relative to Kripkean semantics in general is to say that it has no counterexample in any Kripkean model, regardless of how $\mathscr{R}$ is structured.

With this new relativized notion of validity, we can now prove that all sequents of the form $\Box\Phi \vdash \Phi$ *are* valid—relative to the class of reflexive models:

METATHEOREM: All sequents of the form $\Box\Phi \vdash \Phi$ are valid relative to the set of models whose accessibility relation is reflexive.

PROOF: Suppose for reductio that this is not the case—that is, for some formula Φ there exists a valuation $\mathscr{V}$ whose accessibility relation $\mathscr{R}$ is reflexive and some world w of $\mathscr{V}$ such that $\mathscr{V}(\Box\Phi, w) = T$ and $\mathscr{V}(\Phi, w) \neq T$. Now since $\mathscr{V}(\Box\Phi, w) = T$, by rule 11 $\mathscr{V}(\Phi, u) = T$, for every world u such that $w\mathscr{R}u$. But since $\mathscr{R}$ is reflexive, $w\mathscr{R}w$. Therefore $\mathscr{V}(\Phi, w) = T$, which contradicts what we had concluded above.

Thus we have shown that all sequents of the form $\Box\Phi \vdash \Phi$ are valid relative to the set of models whose accessibility relation is reflexive. QED

We may say, then, that all sequents of the form $\Box\Phi \vdash \Phi$ are valid when '$\Box$' is interpreted as an alethic or epistemic operator, but not if we interpret it as a deontic or temporal operator of the sort indicated earlier. But the validity of all sequents of this form is the same thing as the validity of the T rule introduced in Section 11.4. Thus we may conclude that the T rule is valid for some modalities but not for others.

It is the reflexivity of the accessibility relation that guarantees that sequents of the form $\Box\Phi \vdash \Phi$ are valid. Such sequents were valid as a matter of course on Leibnizian semantics, where it is assumed that each world is possible relative to each, and hence that each world is possible relative to itself. Accessibility in Leibnizian semantics is therefore automatically reflexive. But Kripkean semantics licenses accessibility relations that do not link each world to each, thus grounding the construction of logics weaker in various respects than Leibnizian logic.

Just as the reflexivity of $\mathscr{R}$ guarantees the validity of $\Box\Phi \vdash \Phi$, so other requirements on $\mathscr{R}$ correspond to other modal principles. Principles which hold for all Kripkean models apply to all the logics encompassed by Kripkean semantics. Those which hold only in restricted classes of Kripkean models (such as models in which $\mathscr{R}$ is reflexive) are applicable to some interpretations of the modal operators but not to others.

We noted above that the principle $\Box\Phi \vdash \Box\Box\Phi$ seems plausible for temporal and alethic modalities, but questionable for deontic and epistemic ones. This principle is in fact just the S4 rule discussed in Section 11.4. It is valid on Leibnizian semantics, as we saw in the previous chapter, but it is invalid on Kripkean semantics, since, for example, the instance '$\Box P \vdash \Box\Box P$' is invalid:

METATHEOREM: The sequent '$\Box P \vdash \Box\Box P$' is not valid on Kripkean semantics.

PROOF: Consider the following Kripkean model for propositional logic. Let the set $\mathscr{W}_{\mathscr{V}}$ of worlds be {1, 2, 3} and the accessibility relation $\mathscr{R}$ be the set {<1, 2>, <2, 3>}, and let

$\mathscr{V}$('P', 1) = T
$\mathscr{V}$('P', 2) = T
$\mathscr{V}$('P', 3) = F