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 $\mathscr{V}$ now must take into account not only predicates, but also worlds, in assigning extensions. Thus we write

$\mathscr{V}$('B', w_1) = {α, β}
$\mathscr{V}$('B', w_2) = { }
$\mathscr{V}$('B', w_3) = {α, δ, ε}

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, $\mathscr{V}$('∃xBx', w_1) = T, but $\mathscr{V}$('∃xBx', 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 $\mathscr{V}$('M', w_1) = T, but $\mathscr{V}$('M', w_2) = $\mathscr{V}$('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 $\mathscr{V}(\alpha)$, rather than as $\mathscr{V}(\alpha, w)$, 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 $\mathscr{V}(\text{'n'}) = \beta$. (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, $\mathcal{V}$('Bn', w_1) = T, but $\mathcal{V}$('Bn', w_2) = 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. $\mathcal{V}(\Box\Phi, w) = T$ iff for all worlds u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\Phi, u) = T$;
 $\mathcal{V}(\Box\Phi, w) = F$ iff for some world u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\Phi, u) \neq T$.
12. $\mathcal{V}(\Diamond\Phi, w) = T$ iff for some world u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\Phi, u) = T$;
 $\mathcal{V}(\Diamond\Phi, w) = F$ iff for all worlds u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\Phi, u) \neq T$.

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

Using the valuation rules, prove the following with respect to this model:

1. $\mathscr{V}$('P ∨ Q', 1) = T
2. $\mathscr{V}$('□R', 1) = T
3. For any world w in $\mathscr{W}_{\mathscr{V}}$, $\mathscr{V}$('□R', w) = T
4. There is no world w in $\mathscr{W}_{\mathscr{V}}$ such that $\mathscr{V}$('□P', w) = T
5. For any world w in $\mathscr{W}_{\mathscr{V}}$, $\mathscr{V}$('◇P', w) = T
6. For any world w in $\mathscr{W}_{\mathscr{V}}$, $\mathscr{V}$('~□R', w) = F
7. For any world w in $\mathscr{W}_{\mathscr{V}}$, $\mathscr{V}$('◇~R', w) = F
8. For any world w in $\mathscr{W}_{\mathscr{V}}$, $\mathscr{V}$('P ∨ ~P', w) = T
9. For any world w in $\mathscr{W}_{\mathscr{V}}$, $\mathscr{V}$('□(P ∨ ~P)', w) = T
10. For any world w in $\mathscr{W}_{\mathscr{V}}$, $\mathscr{V}$('~◇(P & ~P)', w) = T

Our semantics is democratic: It treats all possible worlds as equals; none is singled out as uniquely actual. This models another prominent idea in modal metaphysics: the thesis of the **indexicality of actuality.** According to this doctrine, no world is actual in an absolute sense; each is actual from a perspective within that world but not from any perspective external to it. For those whose perspective (consciousness?) is rooted in other possible worlds, our world is merely possible, just as their worlds are merely possible for us. Actuality, then, is indexed to worlds (world-relative) in just the way truth is.

The thesis of the indexicality of actuality is much disputed. Logicians who think that actuality is not indexical may incorporate this idea into their semantics by designating exactly one world of each model as actual. But this bifurcates their concept of truth. They have, on the one hand, a notion of nonrelative or actual truth—that is, truth in the actual world—and, on the other, the same relative notion of truth (truth-in-a-world) that we use in defining possibility and necessity. I use the indexical conception of actuality here because it requires only one sort of truth (world-relative) and hence yields a simpler semantics.

Those who find the indexicality of actuality dizzying may appreciate the following analogy. Imagine you are a transcendent God, perusing the actual universe from creation to apocalypse. As you contemplate this grand spectacle, ask yourself: Which moment is the present?

In your omniscience you should see at once that this question is nonsensical. There is a present moment only for creatures situated within time, not for a God who stands beyond it. The present moment for me at noon on my third birthday is different from the present moment for me as I write these words, which is different from the present moment for you as you read this. None of these is *the* present moment, for there is no absolute present.[8] Presentness is indexed to moments of time—that is, relative to temporal position. If I have lived or will live at a given moment, then that moment is present to the temporal part of me that intersects it but not present to other temporal parts of me.

[8] This is not idle speculation; the thesis that there is no absolute present is central to Einsteinian relativity theory, which is the source of the best understanding of time available at the moment.

Analogously, according to the understanding that grounds our semantics, there is an actual world only for creatures situated within worlds, not for a God—or a modal semanticist—standing beyond them. A world in which I become a farmer is just as actual for that farmer (i.e., for that possible "part" of me) as the world I am currently experiencing is for the professorial portion of me that inhabits it. Neither of these, nor any other, is *the* actual world in some absolute sense, because actuality is always relative to a perspective within some possible world.[9]

That, at any rate, is one way of understanding the "democratic" semantics presented here: Models do not single out an actual world, because our model theory operates from a perspective beyond worlds from which no world is uniquely actual.

Having relativized truth to worlds, we must make compensatory adjustments in those metatheoretic concepts that are defined in terms of truth. Consistency, for example, is no longer merely truth on some valuation (model), for formulas are no longer simply true or false on a valuation; they are true or false *at a world* on a valuation. Thus we must revise our definitions of metatheoretic concepts as follows:

DEFINITION A formula is valid iff it is true in all worlds on all of its valuations.

DEFINITION A formula is consistent iff it is true in at least one world on at least one valuation.

DEFINITION A formula is inconsistent iff it is not true in any world on any of its valuations.

DEFINITION A formula is contingent iff there is a valuation on which it is true in some world and a valuation on which it is not true in some world.

DEFINITION A *set of formulas* is consistent iff there is at least one valuation containing a world in which all the formulas in the set are true.

DEFINITION A *set of formulas* is inconsistent iff there is no valuation containing a world in which all the formulas in the set are true.

DEFINITION Two formulas are equivalent iff they have the same truth value at every world on every valuation of both.

DEFINITION A counterexample to a sequent is a valuation containing a world at which its premises are true and its conclusion is false.

[9] Here we contradict Leibniz, who thought that actuality *was* something absolute—namely, whatever it was that God added to our possible world in order to create it (ours was, according to Leibniz, the only world God created). For a fuller discussion of the indexicality of actuality, see David Lewis, *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986), sec. 1.9, pp. 92–96.

DEFINITION A sequent is valid iff there is no world in any valuation on which its premises are true and its conclusion is not true.

DEFINITION A sequent is invalid iff there is at least one valuation containing a world at which its premises are true and its conclusion is not true.

Using these concepts, we now embark upon a metatheoretic exploration of Leibnizian modal semantics. Our first metatheorem confirms the truism that what is necessary is the case.

METATHEOREM: Any sequent of the form $\Box\Phi \vdash \Phi$ is valid.

PROOF: Suppose for reductio that some sequent of this form is not valid—that is, that there is some formula Φ, some valuation $\mathcal{V}$, and some world w of $\mathcal{V}$ such that $\mathcal{V}(\Box\Phi, w) = T$ but $\mathcal{V}(\Phi, w) \neq T$. Since $\mathcal{V}(\Box\Phi, w) = T$, it follows by valuation rule 11 that $\mathcal{V}(\Phi, u) = T$ for all worlds u in $\mathcal{W}_{\mathcal{V}}$. Hence in particular $\mathcal{V}(\Phi, w) = T$, which contradicts our supposition that $\mathcal{V}(\Phi, w) \neq T$.

Thus we have shown that any sequent of the form $\Box\Phi \vdash \Phi$ is valid. QED

The converse, of course, does not hold. What is need not be necessary. The Earth is populated; but this is not necessarily the case. (It might cease to be the case through any of a variety of catastrophic events, and indeed it might never have happened at all.) To vivify the next metatheorem, think of 'P' as meaning "the Earth is populated." and think of world 1 as the actual world and world 2 as a world in which the Earth is barren.

METATHEOREM: The sequent 'P $\vdash \Box$P' is invalid.

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

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

Now since $\mathcal{V}$('P', 2) $\neq$ T, there is some world u in $\mathcal{W}_{\mathcal{V}}$ (namely, world 2) such that $\mathcal{V}$('P', u) $\neq$ T. Hence by rule 11, $\mathcal{V}$('$\Box$P', 1) $\neq$ T. Therefore we have both $\mathcal{V}$('P', 1) = T and $\mathcal{V}$('$\Box$P', 1) $\neq$ T, which constitutes a counterexample. Thus the sequent is invalid. QED

On Leibnizian semantics what is necessary at one world is necessary at all; therefore, what is necessary is necessarily necessary. This is because necessity itself is truth in all worlds, and if something is true in all worlds, then it is true in all worlds that it is true in all worlds. The following metatheorem gives the details:

METATHEOREM: Any sequent of the form $\Box\Phi \vdash \Box\Box\Phi$ is valid.

PROOF: Suppose for reductio that some sequent of this form is not valid—that is, that there is some formula Φ, some valuation $\mathcal{V}$, and some world w of $\mathcal{V}$ such that $\mathcal{V}(\Box\Phi, w) = \mathrm{T}$ but $\mathcal{V}(\Box\Box\Phi, w) \neq \mathrm{T}$. Since $\mathcal{V}(\Box\Box\Phi, w) \neq \mathrm{T}$, it follows by valuation rule 11 that $\mathcal{V}(\Box\Phi, u) \neq \mathrm{T}$ for some world u in $\mathcal{W}_{\mathcal{V}}$. But then again by rule 11, for some world x in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\Phi, x) \neq \mathrm{T}$. However, since $\mathcal{V}(\Box\Phi, w) = \mathrm{T}$, by rule 11, $\mathcal{V}(\Phi, y) = \mathrm{T}$ for all worlds y in $\mathcal{W}_{\mathcal{V}}$ (in particular for world x); and so we have a contradiction.

Consequently, contrary to our supposition, any sequent of the form $\Box\Phi \vdash \Box\Box\Phi$ is valid. QED

World variables ('*w*', '*u*', '*x*', and '*y*', for example, in the previous metatheorem) are a pervasive feature of modal metalogic. Each such variable should be introduced with a metalinguistic quantifier to indicate whether it stands for all worlds or just some. Variables standing for a particular world may be repeated later in the proof if there is need to refer to that world again. Early in the previous metatheorem, for example, '*w*' is introduced (via existential quantification: "there is a valuation $\mathcal{V}$ containing a world *w*") to stand for a particular world; later it is employed several times to refer to that same world. To avoid ambiguity, it is best to choose a typographically new variable for each quantification. Thus, for example, in the same proof, '*y*' is used to make a universally quantified statement, and '*u*' and '*x*' to make separate existentially quantified statements.

Our next metatheorem proves one of the two biconditionals expressing the idea that '$\Box$' and '$\Diamond$' are duals. (The other is left as an exercise below.) In some systems one of these two operators is taken as primitive and the other is defined in terms of it using one of these biconditionals.

METATHEOREM: Any formula of the form $\Diamond\Phi \leftrightarrow \sim\Box\sim\Phi$ is valid.

PROOF: Suppose for reductio that some formula of this form is not valid. That is, for some formula Φ there exists a valuation $\mathcal{V}$ and world w of $\mathcal{V}$ such that $\mathcal{V}(\Diamond\Phi \leftrightarrow \sim\Box\sim\Phi, w) \neq \mathrm{T}$. It follows by valuation rule 8 that either $\mathcal{V}(\Diamond\Phi, w) = \mathrm{T}$ and $\mathcal{V}(\sim\Box\sim\Phi, w) \neq \mathrm{T}$ or $\mathcal{V}(\Diamond\Phi, w) \neq \mathrm{T}$ and $\mathcal{V}(\sim\Box\sim\Phi, w) = \mathrm{T}$. We show that either case leads to contradiction.

Suppose, first, that $\mathcal{V}(\Diamond\Phi, w) = \mathrm{T}$ and $\mathcal{V}(\sim\Box\sim\Phi, w) \neq \mathrm{T}$. Since $\mathcal{V}(\Diamond\Phi, w) = \mathrm{T}$, by rule 12, $\mathcal{V}(\Phi, u) = \mathrm{T}$ for some world u in $\mathcal{W}_{\mathcal{V}}$. Hence by rule 4 there is some world u in

$\mathcal{W}_{\mathcal{V}}$ at which $\mathcal{V}(\sim\Phi, u) \neq T$. So by rule 11, $\mathcal{V}(\Box\sim\Phi, w) \neq T$. But we had assumed that $\mathcal{V}(\sim\Box\sim\Phi, w) \neq T$, whence it follows by rule 4 that $\mathcal{V}(\Box\sim\Phi, w) = T$; and so we have a contradiction.

Hence it is not the case that both $\mathcal{V}(\Diamond\Phi, w) = T$ and $\mathcal{V}(\sim\Box\sim\Phi, w) \neq T$.

Suppose now that $\mathcal{V}(\Diamond\Phi, w) \neq T$ and $\mathcal{V}(\sim\Box\sim\Phi, w) = T$. Since $\mathcal{V}(\Diamond\Phi, w) \neq T$, by rule 12, $\mathcal{V}(\Phi, u) \neq T$ for all worlds u in $\mathcal{W}_{\mathcal{V}}$. Hence by rule 4 for all worlds u in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}(\sim\Phi, u) = T$. So by rule 11, $\mathcal{V}(\Box\sim\Phi, w) = T$. But we had assumed that $\mathcal{V}(\sim\Box\sim\Phi, w) = T$, whence it follows by rule 4 that $\mathcal{V}(\Box\sim\Phi, w) \neq T$; and so we have a contradiction.

Therefore it is not the case that both $\mathcal{V}(\Diamond\Phi, w) \neq T$ and $\mathcal{V}(\sim\Box\sim\Phi, w) = T$. Thus, since, as we saw above, it is also not the case that both $\mathcal{V}(\Diamond\Phi, w) = T$ and $\mathcal{V}(\sim\Box\sim\Phi, w) \neq T$, then contrary to what we had concluded above,
$\mathcal{V}(\Diamond\Phi \leftrightarrow \sim\Box\sim\Phi, w) = T$.

Thus we have shown that every formula of the form $\Box\Phi \leftrightarrow \sim\Diamond\sim\Phi$ is valid. QED

One of the most important consequences of the doctrine that names are rigid designators is the thesis expressed in the next metatheorem: the neccessity of identity. Kripke, who popularized this thesis in its contemporary form,[10] illustrates it with the following example. 'Phosphorus' is a Latin name for the morning star; 'Hesperus' is the corresponding name for the evening star. But the morning star and the evening star are in fact the same object, the planet we now call Venus. Hence the statement

Hesperus = Phosphorus

is true. Now if names are rigid designators, then since this statement is true, the object designated by the name 'Hesperus' in the actual world is the very same object designated by 'Hesperus' in any other world, and the object designated by the name 'Phosphorus' in the actual world is the same as the object designated by that name in any other world. Thus in every world both names designate the same object they designate in the actual world: the planet Venus. So 'Hesperus = Phosphorus' is not only true in the actual world but necessarily true.

Yet this conclusion is disturbing. So far as the ancients knew, Hesperus and Phosphorus could have been separate bodies; it would seem, then, that it is not necessary that Hesperus = Phosphorus.

But this reasoning is fallacious. The sense in which it was possible that Hesperus was not Phosphorus is the *epistemic* sense; it was possible *so far as the*

[10] *Naming and Necessity* (Cambridge: Harvard University Press, 1972).

ancients knew—that is, compatible with their knowledge—that Hesperus was not Phosphorus. It was not genuinely (i.e., alethically) possible. The planet Venus is necessarily itself; that is, it is itself in any possible world in which it occurs. And if names are rigid designators, then the names 'Hesperus' and 'Phosphorus' both denote Venus in every world in which Venus exists. Hence, given the dual doctrines of transworld identity and rigid designation (both of which are incorporated in our semantics), it is alethically necessary that Hesperus is Phosphorus, despite the fact that it is not epistemically necessary. Keep this example in mind while considering the metatheorem below.

METATHEOREM: Every sequent of the form $\alpha = \beta \vdash \Box\alpha = \beta$ is valid.

PROOF: Suppose for reductio that some sequent of this form is not valid. Then for some names α and β there is a valuation $\mathcal{V}$ and world w of $\mathcal{V}$ such that $\mathcal{V}(\alpha = \beta, w) = T$ and $\mathcal{V}(\Box\alpha = \beta, w) \neq T$. Since $\mathcal{V}(\Box\alpha = \beta, w) \neq T$, by rule 11 there is some world u in $\mathcal{W}_{\mathcal{V}}$ such that $\mathcal{V}(\alpha = \beta, u) \neq T$. Hence by rule 3, $\mathcal{V}(\alpha) \neq \mathcal{V}(\beta)$. But then again by rule 3, $\mathcal{V}(\alpha = \beta, w) \neq T$, which contradicts what we had said above.

Thus, contrary to our supposition, every sequent of the form $\alpha = \beta \vdash \Box\alpha = \beta$ is valid. QED

Modal operators interact with the quantifiers of predicate logic in logically interesting ways. The last two metatheorems of this section illustrate this interaction.

METATHEOREM: The sequent '$\exists x Fx \vdash \exists x \Diamond Fx$' is valid.

PROOF: Suppose for reductio that '$\exists x Fx \vdash \exists x \Diamond Fx$' is not valid. Then there is some valuation $\mathcal{V}$ and world w of $\mathcal{V}$ such that $\mathcal{V}$('$\exists x Fx$', w) = T and $\mathcal{V}$('$\exists x \Diamond Fx$', w) $\neq$ T. Since $\mathcal{V}$('$\exists x Fx$', w) = T, it follows by rule 10 that for some potential name α of some object d in $\mathcal{D}_w$, $\mathcal{V}_{(\alpha,d)}(F\alpha, w) = T$. So for some world u (namely w) in $\mathcal{W}_{\mathcal{V}}$, $\mathcal{V}_{(\alpha,d)}(F\alpha, u) = T$. But then by rule 12, $\mathcal{V}_{(\alpha,d)}(\Diamond F\alpha, w) = T$. Hence, since d is in $\mathcal{D}_w$, by rule 10, $\mathcal{V}$('$\exists x \Diamond Fx$', w) = T, contrary to what we had supposed above.

Thus we have established that '$\exists x Fx \vdash \exists x \Diamond Fx$' is valid. QED

The sequent says that given that something is F, it follows that something (that very same thing, if nothing else) is *possibly* F. This is a consequence of the fact that the actual world, which we may think of as w— and also u—in the proof, is also a possible world, so that whatever actually has a property also possibly has 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 '$\Diamond\exists x Fx \vdash \exists x \Diamond Fx$' is invalid.

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

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

where

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

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

However, $\mathcal{V}$('$\exists x \Diamond Fx$', 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 $\mathcal{W}_{\mathcal{V}}$, name ν and object ω in the domain of world 1 such that $\mathcal{V}_{(\nu,\omega)}$(F$\nu$, u) = T. That is, by rule 12 there is no name ν and object ω in the domain of world 1 such that $\mathcal{V}_{(\nu,\omega)}$($\Diamond$F$\nu$, 1) = T. So by rule 10, $\mathcal{V}$('$\exists x \Diamond Fx$', 1) ≠ T.

Thus, since $\mathcal{V}$('$\Diamond\exists x Fx$', 1) = T but $\mathcal{V}$('$\exists x \Diamond Fx$', 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 '$\Diamond\exists x Fx \vdash \exists x \Diamond Fx$' 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 '$P \vdash \Diamond P$' is valid.
2. The sequent '$\Diamond P \vdash P$' is invalid.
3. The sequent '$\Diamond(P \,\&\, Q) \vdash \Diamond P \,\&\, \Diamond Q$' is valid.
4. The sequent '$\Diamond P \,\&\, \Diamond Q \vdash \Diamond(P \,\&\, Q)$' is invalid.
5. Every sequent of the form $\Phi \vdash \Box\Diamond\Phi$ is valid.
6. Every sequent of the form $\Diamond\Box\Phi \vdash \Box\Phi$ is valid.
7. For any formula Φ, if Φ is a valid formula, then so is $\Box\Phi$.
8. Every formula of the form $\Box\Phi \leftrightarrow \sim\Diamond\sim\Phi$ is valid.
9. Every sequent of the form $\Box\Phi \vdash \Diamond\Phi$ is valid.
10. Every sequent of the form $\Box(\Phi \rightarrow \Psi) \vdash (\Box\Phi \rightarrow \Box\Psi)$ is valid.
11. Every sequent of the form $\Box(\Phi \rightarrow \Psi) \vdash \sim\Diamond(\Phi \,\&\, \sim\Psi)$ is valid.
12. Every sequent of the form $\sim\Diamond(\Phi \,\&\, \sim\Psi) \vdash \Box(\Phi \rightarrow \Psi)$ is valid.
13. The sequent '$\Box P, P \rightarrow Q \vdash \Box Q$' is invalid.
14. Every formula of the form $\Box\alpha = \alpha$ is valid.
15. Every sequent of the form $\sim\alpha = \beta \vdash \Box\sim\alpha = \beta$ is valid.
16. Every sequent of the form $\Diamond\alpha = \beta \vdash \alpha = \beta$ is valid.
17. The sequent '$\forall x \Box Fx \vdash \Box\forall x Fx$' is invalid.
18. Every sequent of the form $\forall\beta\Box\Phi \vdash \forall\beta\Phi$ 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.

11.4 INFERENCE IN LEIBNIZIAN LOGIC

Leibnizian propositional logic retains all the inference rules of classical propositional logic but adds new rules to handle the modal operators. Though we shall examine inferences involving identity, we shall not deal with quantifiers in this section, since the quantifier rules depend on how we resolve the question of predication for nonexisting objects. One reasonable way of resolving this question is to adopt a **free logic**—that is, a logic free of the presupposition that every name always names some existing thing. We shall consider free logics in Section 15.1, and we defer the treatment of modal inferences involving quantification to that section.

The nonquantificational Leibnizian logic that we will explore in this section adds to the rules of classical propositional logic and the classical rules for identity seven new inference rules (the names of most are traditional and of various origins):

Duality (DUAL) From either of $\Diamond\Phi$ and $\sim\Box\sim\Phi$, infer the other; from either of $\Box\Phi$ and $\sim\Diamond\sim\Phi$, infer the other.

K rule (K) From $\Box(\Phi \rightarrow \Psi)$, infer $(\Box\Phi \rightarrow \Box\Psi)$.

T rule (T) From $\Box\Phi$, infer Φ.

S4 rule (S4) From $\Box\Phi$, infer $\Box\Box\Phi$.

Brouwer rule (B) From Φ, infer $\Box\Diamond\Phi$.

Necessitation (N) If Φ has previously been proved as a theorem, then any formula of the form $\Box\Phi$ may be introduced at any line of a proof.

Necessity of identity ($\Box$=) From $\alpha = \beta$, infer $\Box\alpha = \beta$.

It is not difficult to show that every instance of each of these rules is valid on a Leibnizian semantics—and indeed we did this for some of them in Section 11.2 (the rest were left as exercises).

The necessitation rule differs from the others in that it uses no premises but refers, rather, to theorems established by previous proofs. A theorem is a valid formula, a formula true in all worlds on all valuations. Therefore, if Φ is a theorem, $\Box\Phi$ and any formula of the form $\Box\Phi$ may be asserted anywhere in a proof without further assumptions. When we use the rule of necessitation, we annotate it by writing its abbreviation 'N' to the right of the introduced formula, followed by the previously proved theorem or axiom schema employed.

These seven inference rules, together with the rules of classical propositional logic and the identity rules =I and =E, constitute a system of inference that is sound and complete with respect to a Leibnizian semantics for the modal propositional logic with identity—but to show this is beyond our scope. The purely propositional rules (i.e., the ones other than =I, =E, and □=) comprise a logic known as S5.[12] This section is largely an exploration of the valid inferential patterns of S5.

We begin by proving the sequent 'P ⊢ ◇P':

1.	P	A
2.	│ □~P	H (for ~I)
3.	│ ~P	2 T
4.	│ P & ~P	1, 3 &I
5.	~□~P	2–4 ~I
6.	◇P	5 DUAL

The strategy is an indirect proof. Recognizing initially that '◇P' is interchangeable with '~□~P', we hypothesize '□~P' for reductio. Using the T rule, the contradiction is obtained almost immediately. This yields '~□~P', which is converted into '◇P' by DUAL at line 6.

The rules N and K are often used together to obtain modalized versions of various theorems and rules. The sequent '□(P & Q) ⊢ □P', for example, which is a modalized version of &E, is proved by using N and then K:

1.	□(P & Q)	A
2.	□((P & Q) → P)	N ((P & Q) → P)[13]
3.	□(P & Q) → □P	2 K
4.	□P	1, 3 →E

A similar but more sophisticated strategy utilizing N and K yields sequents involving possibility. Our next example is a proof of '◇P ⊢ ◇(P ∨ Q)', a modalized version of ∨I. Here we apply N to the theorem '~(P ∨ Q) → ~P', the contrapositive of 'P → (P ∨ Q)', which in effect expresses ∨I. (This strategy of applying N to contraposed nonmodal versions of the modal sequent we want to prove is typical when the modality involved is possibility.)

1.	◇P	A
2.	□(~(P ∨ Q) → ~P)	N (~(P ∨ Q) → ~P)[14]
3.	□~(P ∨ Q) → □~P	2 K
4.	~□~P	1 DUAL
5.	~□~(P ∨ Q)	3, 4 MT
6.	◇(P ∨ Q)	5 DUAL

Note the use of the derived rule modus tollens at line 5. Derived rules for classical propositional logic (see Section 4.4) are all available in Leibnizian modal logic.

N and K are used together once again in this derivation of the theorem '⊢ ◇~P → ~□P':

[12] The name originates with the logician C. I. Lewis, whose pioneering work on modal logic dates from the first few decades of the twentieth century. Lewis explored a number of modal systems, which he christened with such unmemorable labels. Inexplicably, the labels stuck.

[13] This theorem is problem 2 of Exercise 4.4.2.

[14] See problem 6 of Exercise 4.4.2.

1.	◇~P	H (for →I)
2.	~□~~P	1 DUAL
3.	□(P → ~~P)	N (P → ~~P)[15]
4.	□P → □~~P	3 K
5.	~□P	2, 4 MT
6.	◇~P → ~□P	1–5 →I

However, a very different strategy may be used to prove the related theorem '⊢ □~P → ~◇P':

1.	□~P	H (for →I)
2.	◇P	H (for ~I)
3.	~□~P	2 DUAL
4.	□~P & ~□~P	1, 3 &I
5.	~◇P	2–4 ~I
6.	□~P → ~◇P	1–5 →I

Here, after hypothesizing the theorem's antecedent for conditional proof, we employ an indirect proof, hypothesizing '◇P' for reductio at line 2. The use of DUAL at line 3 immediately provides a contradiction, which is recorded at line 4, and the conclusion follows by easy steps of ~I and →I at lines 5 and 6.

The following proof of the sequent '□(P → Q) ⊢ □~Q → □~P', which is a kind of modalized version of modus tollens, displays further uses of N and K:

1.	□(P → Q)	A
2.	□((P → Q) → (~Q → ~P))	N ((P → Q) → (~Q → ~P))
3.	□(P → Q) → □(~Q → ~P)	2 K
4.	□(~Q → ~P)	1, 3 →E
5.	□~Q → □~P	4 K

The necessitation rule N is used at line 2 with the theorem '⊢ (P → Q) → (~Q → ~P)', which was proved in Section 4.4. In the use of K at line 3, Φ is 'P → Q' and Ψ is '~Q → ~P', but at line 5 Φ is '~Q' and Ψ is '~P'.

The B rule is used in the following proof of '◇□P ⊢ P':

1.	◇□P	A
2.	~P	H (for ~I)
3.	□◇~P	2 B
4.	□(◇~P → ~□P)	N (◇~P → ~□P)
5.	□◇~P → □~□P	4 K
6.	~□~□P	1 DUAL
7.	~□◇~P	5, 6 MT
8.	□◇~P & ~□◇~P	3, 7 &I
9.	~~P	2–8 ~I
10.	P	9 ~E

[15] See problem 3 of Exercise 4.4.2.

Note the use of N with the previously proved modal theorem '$\Diamond{\sim}P \rightarrow {\sim}\Box P$' at line 4.

We next prove the theorem '$\vdash \Diamond\Diamond P \rightarrow \Diamond P$', using the S4 rule:

1.		$\Diamond\Diamond P$		H (for →I)
2.			$\Box{\sim}P$	H (for ~I)
3.			$\Box\Box{\sim}P$	2 S4
4.			$\Box(\Box{\sim}P \rightarrow {\sim}\Diamond P)$	N ($\Box{\sim}P \rightarrow {\sim}\Diamond P$)
5.			$\Box\Box{\sim}P \rightarrow \Box{\sim}\Diamond P$	4 K
6.			$\Box{\sim}\Diamond P$	3, 5 →E
7.			${\sim}\Box{\sim}\Diamond P$	1 DUAL
8.			$\Box{\sim}\Diamond P \; \& \; {\sim}\Box{\sim}\Diamond P$	6, 7 &I
9.		${\sim}\Box{\sim}P$		2–8 ~I
10.		$\Diamond P$		9 DUAL
11.	$\Diamond\Diamond P \rightarrow \Diamond P$			1–10 →I

This theorem can easily be strengthened to the biconditional '$\Diamond P \leftrightarrow \Diamond\Diamond P$', using the previously proved sequent '$P \vdash \Diamond P$' as a derived rule. This biconditional shows that repetition of possibility operators is in effect redundant in Leibnizian logic. The same can be shown for necessity operators—that is, '$\vdash \Box P \leftrightarrow \Box\Box P$', but the proof is left as an exercise.

As in propositional and predicate logic, we may use derived rules. We will not, however, bother to name them, since few have widely used names. Instead, we simply list the previously proved sequent to the right, together with the line numbers of the premises (if any) that are instances of the previously proved sequent's premises. (Rules derived from theorems have no premises, and we cite no lines for them.) This proof of '$\Box(P \rightarrow Q) \vdash \Diamond P \rightarrow \Diamond Q$' uses the previously proved sequent '$\Box(P \rightarrow Q) \vdash \Box{\sim}Q \rightarrow \Box{\sim}P$' as a derived rule at line 4:

1.	$\Box(P \rightarrow Q)$			A
2.		$\Diamond P$		H (for →I)
3.			$\Box{\sim}Q$	H (for ~I)
4.			$\Box{\sim}Q \rightarrow \Box{\sim}P$	1 $\Box(P \rightarrow Q) \vdash \Box{\sim}Q \rightarrow \Box{\sim}P$
5.			$\Box{\sim}P$	3, 4 →E
6.			${\sim}\Box{\sim}P$	2 DUAL
7.			$\Box{\sim}P \; \& \; {\sim}\Box{\sim}P$	5, 6 &I
8.		${\sim}\Box{\sim}Q$		3–7 ~I
9.		$\Diamond Q$		8 DUAL
10.	$\Diamond P \rightarrow \Diamond Q$			2–9 →I

Notice the use of indirect proof with the duality rule to obtain '$\Diamond Q$'.

As I pointed out in Section 4.4, proof of a sequent establishes the validity of any formula that shares that sequent's form. Thus, when we use a sequent as a derived rule, we may use any instance of it. The following proof of the sequent '${\sim}a = b \vdash \Box{\sim}a = b$' utilizes the previously proved sequent '$\Box(P \rightarrow Q) \vdash \Diamond P \rightarrow \Diamond Q$' as a derived rule at line 5. This sequent is used, however, in the form '$\Box(a = b \rightarrow \Box a = b) \vdash \Diamond a = b \rightarrow \Diamond\Box a = b$', where '$a = b$' replaces '$P$' and '$\Box a = b$' replaces '$Q$'.

Similarly, the previously proved sequent '$\Diamond\Box P \vdash P$' is used in the form '$\Diamond\Box a = b \vdash a = b$' at line 7.

1.	$\sim a = b$	A
2.	$\sim\Box\sim a = b$	H (for ~I)
3.	$\Diamond a = b$	2 DUAL
4.	$\Box(a = b \rightarrow \Box a = b)$	N $(a = b \rightarrow \Box a = b)$
5.	$\Diamond a = b \rightarrow \Diamond\Box a = b$	4 $\Box(P \rightarrow Q) \vdash \Diamond P \rightarrow \Diamond Q$
6.	$\Diamond\Box a = b$	3, 5 →E
7.	$a = b$	6 $\Diamond\Box P \vdash P$
8.	$a = b \ \& \sim a = b$	1, 7 &I
9.	$\sim\sim\Box\sim a = b$	2–8 ~I
10.	$\Box\sim a = b$	9 ~E

At line 4 we use the necessitation rule with the theorem '$\vdash a = b \rightarrow \Box a = b$'. We didn't actually prove this theorem, but its proof is trivial, given the $\Box =$ rule, and is left for an exercise below.

This proof shows that not only is identity necessary as the $\Box =$ axiom schema asserts, but also nonidentity is necessary—a result fully appropriate in light of the semantics of rigid designation.

Our next result establishes that whatever is possible is necessarily possible—that is, (on Leibnizian semantics) possible with respect to any world. The sequent expressing this idea is '$\Diamond P \vdash \Box\Diamond P$':

1.	$\Diamond P$	A
2.	$\Box\Diamond\Diamond P$	1 B
3.	$\Box(\Diamond\Diamond P \rightarrow \Diamond P)$	N $(\Diamond\Diamond P \rightarrow \Diamond P)$
4.	$\Box\Diamond\Diamond P \rightarrow \Box\Diamond P$	3 K
5.	$\Box\Diamond P$	2, 4 →E

And finally we show that whatever is even possibly necessary is necessary. That is, the sequent '$\Diamond\Box P \vdash \Box P$' is provable:

1.	$\Diamond\Box P$	A
2.	$\Box(\Diamond\Box P \rightarrow P)$	N $(\Diamond\Box P \rightarrow P)$
3.	$\Box\Diamond\Box P \rightarrow \Box P$	2 K
4.	$\Box\Diamond\Box P$	1 $\Diamond P \vdash \Box\Diamond P$
5.	$\Box P$	3, 4 →E

Exercise 11.4

Prove the following sequents:

1. $\vdash a = b \rightarrow \Box a = b$
2. $\Box P \vdash \Box(P \vee Q)$
3. $\vdash \Box\sim\sim P \rightarrow \Box P$
4. $\Diamond(P \ \& \ Q) \vdash \Diamond P$
5. $\Box Q \vdash \Box(P \rightarrow Q)$

6. $\sim\Diamond P \vdash \Box(P \rightarrow Q)$
7. $\vdash \Box P \leftrightarrow \Box\Box P$
8. $\vdash \sim\Diamond(P \,\&\, \sim P)$
9. $\Box P, \Box Q \vdash \Box(P \,\&\, Q)$
10. $\Box P \vdash \Box\Box\Box P$

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.

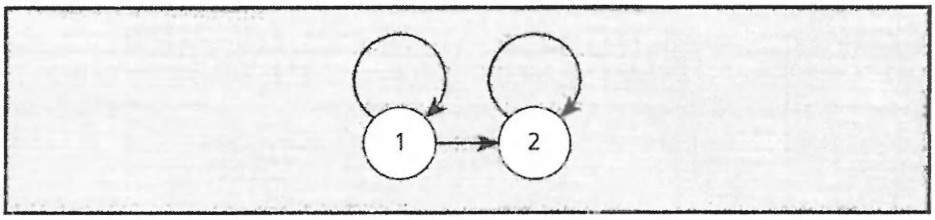
The fundamental notion of Kripkean semantics is the concept of **relative possibility** (which is also called **alternativeness** or **accessibility**). Relative possibility is the relation which holds between worlds x and y just in case y is possible relative to x. The letter '$\mathscr{R}$' is customarily 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 '$\Box$' and '$\Diamond$'. 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' $\mathcal{V}(\Box\Phi, w) = T$ iff for all worlds u such that $w\mathcal{R}u$, $\mathcal{V}(\Phi, u) = T$;
$\mathcal{V}(\Box\Phi, w) = F$ iff for some world u, $w\mathcal{R}u$ and $\mathcal{V}(\Phi, u) \neq T$.
12' $\mathcal{V}(\Diamond\Phi, w) = T$ iff for some world u, $w\mathcal{R}u$ and $\mathcal{V}(\Phi, u) = T$;
$\mathcal{V}(\Diamond\Phi, w) = F$ iff for all worlds u such that $w\mathcal{R}u$, $\mathcal{V}(\Phi, u) \neq T$.

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 $\vdash \Box\Diamond$P', which was valid on Leibnizian semantics, is invalid on this Kripkean model. For $\mathcal{V}$('P', 1) = T, but $\mathcal{V}$('$\Box\Diamond$P', 1) $\neq$ T. That is, world 1 provides a counterexample.

We can see that $\mathcal{V}$('$\Box\Diamond$P', 1) $\neq$ 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) $\neq$ T. Hence for all worlds u in $\mathcal{W}_{\mathcal{V}}$ such that $2\mathcal{R}u$, $\mathcal{V}$('P', u) $\neq$ T. So by rule 12', $\mathcal{V}$('$\Diamond$P', 2) $\neq$ 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 $\mathcal{V}$('$\Diamond$P', x) $\neq$ T. It follows by rule 11' that $\mathcal{V}$('$\Box\Diamond$P', 1) $\neq$ T. We restate this finding as a formal metatheorem:

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

PROOF: As given above.

Moreover, neither of the other sequents mentioned in this section—'$\Box$P $\vdash$ P' and '$\Box$P $\vdash \Box\Box$P'—is valid, either. Let's take '$\Box$P $\vdash$ P' first.

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

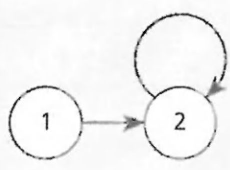
PROOF: Consider the following Kripkean model for propositional logic. Let the set $\mathcal{W}$ of worlds be {1, 2} and the accessibility relation $\mathcal{R}$ be the set {<1, 2>, <2, 2>}, and let

$\mathcal{V}$('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 'OP ⊢ 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 $\mathcal{V}$ such that $\mathcal{V}(\Box\Phi, w) = \text{T}$ and $\mathcal{V}(\Phi, w) \neq \text{T}$. Now since $\mathcal{V}(\Box\Phi, w) = \text{T}$, by rule 11' $\mathcal{V}(\Phi, u) = \text{T}$, for every world u such that $w\mathscr{R}u$. But since $\mathscr{R}$ is reflexive, $w\mathscr{R}w$. Therefore $\mathcal{V}(\Phi, w) = \text{T}$, 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 intepretations 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 $\mathcal{W}_{\mathcal{V}}$ of worlds be {1, 2, 3} and the accessibility relation $\mathscr{R}$ be the set {<1, 2>, <2, 3>}, and let

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

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 $1\mathscr{R}u$. Hence for all worlds u such that $1\mathscr{R}u$, $\mathcal{V}$('P', u) = T. Therefore by rule 11', $\mathcal{V}$('□P', 1) = T. However, since $2\mathscr{R}3$ and $\mathcal{V}$('P', 3) ≠ T, by rule 11', $\mathcal{V}$('□P', 2) ≠ T. And since $1\mathscr{R}2$ and $\mathcal{V}$('□P', 2) ≠ T, again by rule 11', $\mathcal{V}$('□□P', 1) ≠ T. Therefore, since $\mathcal{V}$('□P', 1) = T and $\mathcal{V}$('□□P', 1) ≠ T, we have a counterexample, and so '□P ⊢ □□P' is not valid on Kripkean semantics. QED

Yet the S4 rule *is* valid relative to models whose accessibility relation is transitive. The relation $\mathscr{R}$ is transitive if and only if for any worlds x, y, and z, if $x\mathscr{R}y$ and $y\mathscr{R}z$, then $x\mathscr{R}z$. Think of this in relation to physical possibility. We said that a world y is physically possible relative to a world x if and only if y obeys the same physical laws (and perhaps some additional physical laws as well). That is,

$x\mathscr{R}y$ if and only if y obeys all the physical laws that hold in x.

Now clearly if y obeys all the laws that hold in x and z obeys all the laws that hold in y, then z obeys all the laws that hold in x. That is, if $x\mathscr{R}y$ and $y\mathscr{R}z$, then $x\mathscr{R}z$. So the accessibility relation for physical possibility is transitive. The next metatheorem shows how we get from this fact about the accessibility relation to the conclusion that all sequents of the form □Φ ⊢ □□Φ are valid, where '□' is interpreted as physical necessity.

METATHEOREM: All sequents of the form □Φ ⊢ □□Φ are valid relative to the set of models whose accessibility relation is transitive.

PROOF: Suppose for reductio that this is not the case—that is, for some formula Φ there exists a valuation $\mathcal{V}$ whose accessibility relation $\mathscr{R}$ is transitive and some world x of $\mathcal{V}$ such that $\mathcal{V}$(□Φ, x) = T and $\mathcal{V}$(□□Φ, x) ≠ T. Now since $\mathcal{V}$(□Φ, x) = T, by rule 11', $\mathcal{V}$(Φ, u) = T for every world u such that $x\mathscr{R}u$. But since $\mathcal{V}$(□□Φ, x) ≠ T, by rule 11' there is a world y such that $x\mathscr{R}y$ and $\mathcal{V}$(□Φ, y) ≠ T. And since $\mathcal{V}$(□Φ, y) ≠ T, again by rule 11' there is a world z such that $y\mathscr{R}z$ and $\mathcal{V}$(Φ, z) ≠ T. Now since $x\mathscr{R}y$ and $y\mathscr{R}z$, and $\mathscr{R}$ is transitive, it follows that $x\mathscr{R}z$. But we saw above that $\mathcal{V}$(Φ, u) = T for every world u such that $x\mathscr{R}u$. So in particular $\mathcal{V}$(Φ, z) = T, contrary to what we just concluded.

Thus all sequents of the form □Φ ⊢ □□Φ are valid relative to the set of models whose accessibility relation is transitive. QED

For our last example, we return to the principle Φ ⊢ □◇Φ, which was valid on Leibnizian semantics (indeed, it is just the B rule introduced in Section 11.4) but seemed invalid for physical possibility. (The fact that planets move in elliptical orbits does not mean that it is *necessarily possible* that planets move in elliptical orbits, for there are physically possible worlds in which planetary orbits are necessarily circular and hence in which elliptical orbits are impossible.) The property of $\mathscr{R}$ that would make this sequent valid is symmetry. $\mathscr{R}$ is **symmetric** if and only if for any worlds x and y, if $x\mathscr{R}y$, then $y\mathscr{R}x$. The accessibility relation for physical possibility is not symmetric, since a world with our physical laws plus some "extra" laws would be physically possible relative to our world, but ours would not be physically possible relative to it (since our world violates its "extra" laws). Logical possibility, however, presumably does have a symmetric accessibility relation—assuming (as is traditional) that the laws of logic are the same for all worlds. The final metatheorem in this section shows why symmetry guarantees the validity of $\Phi \vdash \Box\Diamond\Phi$.

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

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 symmetric and some world x of $\mathscr{V}$ such that $\mathscr{V}(\Phi, x) = T$ and $\mathscr{V}(\Box\Diamond\Phi, x) \neq T$. Now since $\mathscr{V}(\Box\Diamond\Phi, x) \neq T$, by rule 11' there is a world y such that $x\mathscr{R}y$ and $\mathscr{V}(\Diamond\Phi, y) \neq T$. And since $\mathscr{V}(\Diamond\Phi, y) \neq T$, by rule 12' for all worlds u such that $y\mathscr{R}u$, $\mathscr{V}(\Phi, u) \neq T$. But $\mathscr{R}$ is symmetric; and so since $x\mathscr{R}y$ it follows that $y\mathscr{R}x$. Thus since for all worlds u such that $y\mathscr{R}u$, $\mathscr{V}(\Phi, u) \neq T$, it follows in particular that $\mathscr{V}(\Phi, x) \neq T$. But we concluded above that $\mathscr{V}(\Phi, x) = T$, which is a contradiction.

So, contrary to our hypothesis, $\Phi \vdash \Box\Diamond\Phi$ is valid relative to the set of models whose accessibility relation is symmetric. QED

We have said so far that the accessibility relation for all forms of alethic possibility is reflexive. For physical possibility, I have argued that it is transitive as well. And for logical possibility it seems also to be symmetric. Thus the accessibility relation for logical possibility is apparently reflexive, transitive, *and* symmetric. It can be proved, though we shall not do so here, that these three characteristics together define the logic S5, which is characterized by Leibnizian semantics. That is, making the accessibility relation reflexive, transitive, and symmetric has the same effect on the logic as making each world possible relative to each.

Leibnizian semantics can in fact be viewed as a special case of Kripkean semantics—the case in which we restrict admissible models to those whose accessibility relation is universal, that is, those in which each world is accessible from each. Universal relations are, of course, automatically reflexive, transitive, and symmetric. Thus, for example, any sequent which is valid in all reflexive models is also valid in all universal models. Sequents valid on Leibnizian semantics can from the Kripkean perspective be regarded as sequents valid relative to the special class of models with universal accessibility relations. Since Leibnizian semantics seems appropriate for logical possibility, from a Kripkean point of view logical possibility is characterized by the class of Kripkean models with universal accessibility relations.

If we drop the requirement of symmetry, we lose the law $\Phi \vdash \Box\Diamond\Phi$ (the inference rule B of Section 11.4), and principles derivable from it, and obtain a weaker logic, S4, which is a good candidate for being the logic of physical possibility.

Logics for the other modalities involve other principles and other properties of $\mathscr{R}$, many of which are disputed. The chief merit of Kripkean semantics is that it opens up new ways of conceiving and interrelating issues of time, possibility, knowledge, obligation, and so on. For each we can imagine a relevant set of worlds (or moments) and a variety of ways an accessibility relation could structure this set and define an appropriate logic. This raises intriguing questions that, were it not for Kripke's work, we never would have dreamed of asking.

Exercise 12.1

Prove the following metatheorems. [Note that saying that a form is valid relative to the set of all Kripkean models is just another way of saying that it is (unqualifiedly) valid on Kripkean semantics.]

1. $\Phi \vdash \Diamond\Phi$ is valid relative to the set of models whose accessibility relation is reflexive.
2. $\Diamond\Diamond\Phi \vdash \Diamond\Phi$ is valid relative to the set of models whose accessibility relation is transitive.
3. $\Diamond\Box\Phi \vdash \Phi$ is valid relative to the set of models whose accessibility relation is symmetric.
4. '$P \vdash \Diamond P$' is not valid relative to the set of all Kripkean models.
5. '$\Diamond\Diamond P \vdash \Diamond P$' is not valid relative to the set of all Kripkean models.
6. '$\Diamond\Box P \vdash P$' is not valid relative to the set of all Kripkean models.
7. $\Box(\Phi \rightarrow \Psi) \vdash \Box\Phi \rightarrow \Box\Psi$ is valid relative to the set of all Kripkean models.
8. $\Box\Phi \vdash \Diamond\Phi$ is valid relative to the set of models whose accessibility relation is reflexive.
9. $\Diamond\Phi \vee \Diamond{\sim}\Phi$ is valid relative to the set of models whose accessibility relation is reflexive.
10. ${\sim}\Diamond(\Phi \,\&\, {\sim}\Phi)$ is valid relative to the set of all Kripkean models.

12.2 INFERENCE IN KRIPKEAN LOGICS

In Section 11.4 we introduced the full Leibnizian logic S5. Since then we have seen that some of the rules of S5 are inappropriate for certain forms of modality. The T rule (from $\Box\Phi$ infer Φ), for example, is plainly invalid when '$\Box$' is taken to express obligation, as it is in deontic logics. We have now seen that this rule was validated by the reflexivity of the accessibility relation. Likewise, the B rule (from Φ infer

$\Box\Diamond\Phi$), which is validated by the symmetry of the accessibility relation, seems invalid for physical possibility. And again the S4 rule (from $\Box\Phi$ infer $\Box\Box\Phi$), which is validated by the transitivity of the accessibility relation, is of questionable validity for several modalities.

Just as Kripkean *semantics* permits nonreflexive, nonsymmetric, or nontransitive accessibility relations, which are fragments, as it were, of the full universal accessibility relation of Leibnizian semantics, so Kripkean *logics* may be fragments of the full Leibnizian logic S5. Less metaphorically, Kripkean logics may lack some of the rules of inference (either basic or derived) that are available in S5.

There are, however, certain rules that are valid relative to the set of all Kripkean models. These rules, in other words, have no counterexamples no matter how severely we diminish the accessibility relation. Three rules in particular are fundamental in this way:

Duality (DUAL) From either of $\Diamond\Phi$ and $\sim\Box\sim\Phi$, infer the other; from either of $\Box\Phi$ and $\sim\Diamond\sim\Phi$, infer the other.

K rule (K) From $\Box(\Phi \rightarrow \Psi)$, infer $(\Box\Phi \rightarrow \Box\Psi)$.

Necessitation (N) If Φ has previously been proved as a theorem, then any formula of the form $\Box\Phi$ may be introduced at any line of a proof.

These rules are common to all Kripkean modal logics. Together with the ten basic rules of classical propositional logic they constitute a logic that is sound and complete relative to the set of all Kripkean models. This logic is known as the system K (for Kripke!). In other words, a sequent of propositional modal logic (modal logic without the identity predicate or quantifiers) is provable in the system K iff it has no counterexample in any Kripkean model.[4]

K itself is not very interesting. But by adding various rules to K we may obtain differing logics that are useful for different purposes. Each rule corresponds to a particular structural requirement on the accessibility relation. Imposing new structural requirements diminishes the range of admissible models—models that may serve as counterexamples. Thus imposing new structural requirements on $\mathscr{R}$ increases the number of valid rules. Among systems we have considered, the one with the most structural requirements is S5, for whose admissible models $\mathscr{R}$ must be reflexive, transitive, and symmetric. In a sense S5 is the maximal Kripkean logic, since it is sound and complete for the most restrictive class of models, the class of models whose accessibility relation is universal. (Though reflexivity, transitivity, and symmetry don't entail universality, the class of all universal models determines the same logic, S5, as the class of reflexive, transitive, and symmetric models does.)

[4] Proofs of the soundness and completeness of a great variety of Kripkean systems may be found in Brian F. Chellas, *Modal Logic: An Introduction* (Cambridge: Cambridge University Press, 1980), chap. 3.

Table 12.1 summarizes some characteristics of five important Kripkean logics. But there are, in fact, infinitely many Kripkean logics, dozens if not hundreds of which have received detailed treatment. Table 12.1, then, presents only a small sample.

Exercise 12.2

Note that in the problems below it is not safe to use the sequents proved in Section 11.4 as derived rules, since these were proved using the full logic of S5 and the systems in which we are working are fragments of S5 in which certain rules are unavailable. Nevertheless, some of the strategies illustrated in that section may be useful here.

I. Construct proofs for the following sequents in the system K:
   1. ~◇P ⊢ □~P
   2. ~□P ⊢ ◇~P
   3. ⊢□(P → P)

II. Construct proofs for the following sequents in the system D:
   1. ⊢ ◇(P → P)
   2. ⊢ ~□(P & ~P)
   3. ~◇P ⊢ ~□P
   4. □□P ⊢ □◇P

III. Construct proofs for the following sequents in the system T:
   1. □P ⊢ ◇P
   2. ~◇P ⊢ ~P
   3. ~P ⊢ ~□P

IV. Construct proofs for the following sequents in the system S4:
   1. ⊢ ◇P ↔ ◇◇P
   2. ◇~□P ⊢ ◇~P
   3. ~◇P ⊢ □~◇P

12.3 STRICT CONDITIONALS

We have until now been using the material conditional, symbolized by →, to render the English operator 'if . . . then' into formal logic. This practice, as we noted in Section 3.1, is, strictly speaking, illegitimate. The material conditional is at best only a loose approximation to 'if . . . then'. Many inferences which are valid for the material conditional are invalid for English conditionals. Consider, for example:

    Socrates grew to manhood.
∴ If Socrates died as a child, then Socrates grew to manhood.

    Socrates did not die as a child.
∴ If Socrates died as a child, then Socrates grew to manhood.

TABLE 12.1
Some Important Kripkean Propositional Modal Logics

Logic	Basic Rules*	Accessibility Relation	Application
K	DUAL From either of $\Diamond\Phi$ and $\sim\Box\sim\Phi$, infer the other; from either of $\Box\Phi$ and $\sim\Diamond\sim\Phi$, infer the other. K From $\Box(\Phi \rightarrow \Psi)$, infer $(\Box\Phi \rightarrow \Box\Psi)$. N If Φ has previously been proved as a theorem, then any formula of the form $\Box\Phi$ may be introduced at any line of a proof.	No restrictions	Minimal Kripkean logic
D	DUAL, K, and N, together with: D From $\Box\Phi$ infer $\Diamond\Phi$.	Serial (see Section 13.1)	Good candidate for minimal deontic logic
T	DUAL, K, and N, together with: T From $\Box\Phi$ infer Φ.	Reflexive	Minimal alethic logic
S4	DUAL, K, N, and T, together with: S4 From $\Box\Phi$ infer $\Box\Box\Phi$.	Reflexive, transitive	Good candidate for logic of physical possibility; closely related to intuitionistic logic (see Section 16.2)
S5	DUAL, K, N, T, and S4, together with: B From Φ infer $\Box\Diamond\Phi$.	Reflexive, transitive, symmetric	Logic of logical possibility, perhaps other kinds of possibility as well (see Section 13.2)

*In addition to the ten rules of classical propositional logic.

It is not the case that if Socrates was a rock then Socrates was a man.
∴ Socrates was a rock, but not a man.

If we eliminate auto accidents, then we save thousands of lives.
If we nuke the entire planet, then we eliminate auto accidents.
∴ If we nuke the entire planet, then we save thousands of lives.

If the Atlantic is an ocean, then it's a polluted ocean.
∴ If the Atlantic is not a polluted ocean, then it's not an ocean.

In each case, the premises are true and the conclusion is false in the actual world, using our ordinary understanding of the conditional. Yet in each case, the argument is valid if we interpret 'if . . . then' as the material conditional. The last two arguments have forms that at first glance appear to be paradigms of good reasoning: hypothetical syllogism,

$$A \rightarrow B, B \rightarrow C \vdash A \rightarrow C$$

and contraposition,

$$A \rightarrow B \vdash \sim B \rightarrow \sim A$$

(sometimes called "transposition"). Yet these forms are apparently invalid for 'if . . . then'.

C. I. Lewis, the inventor of S4, S5, and other modern modal systems, was one of the first formal logicians to investigate the disparity between English and material conditionals. Lewis noticed that ordinary English conditionals seemed to express, not just a truth function, but a necessary connection between antecedent and consequent. Defying skepticism about the intelligibility of the concept of necessary connection, Lewis introduced in 1918 a new conditional, represented by the symbol '$⥽$', which incorporated this idea. $\Phi ⥽ \Psi$ is true if and only if it is impossible for both Φ to be true and Ψ false. $\Phi ⥽ \Psi$ is true, in other words, if and only if it is necessarily the case that if Φ then Ψ, where 'if . . . then' signifies the material conditional. Thus '$⥽$' is often introduced as a defined operator into modal systems using the definition

$$\Phi ⥽ \Psi \text{ iff } \Box(\Phi \rightarrow \Psi)$$

An equivalent definition in terms of the possibility operator is

$$\Phi ⥽ \Psi \text{ iff} \sim \Diamond (\Phi \,\&\, \sim\Psi)$$

Translated into Kripkean semantics, the truth conditions for the strict conditional are as follows:

$\mathscr{V}(\Phi ⥽ \Psi, w) = \text{T}$ iff for all worlds u such that $w\mathscr{R}u$ and $\mathscr{V}(\Phi, u) = \text{T}$, $\mathscr{V}(\Psi, u) = \text{T}$
$\mathscr{V}(\Phi ⥽ \Psi, w) = \text{F}$ iff for some world u such that $w\mathscr{R}u$, $\mathscr{V}(\Phi, u) = \text{T}$ and $\mathscr{V}(\Psi, u) \neq \text{T}$

The strict conditional is in some respects a better approximation to English conditionals than is the material conditional. But the closeness of the approximation depends in part upon which brand of alethic necessity we intend the strict conditional to express. Usually, the necessity built into the connection expressed by English conditionals seems to be something more like practical than physical, metaphysical, or logical necessity. So, though $\mathscr{R}$, like all alethic accessibility relations, should be reflexive, it is doubtful that it need also be transitive and symmetric (as the accessibility relation for logical possibility probably is). Accordingly, we adopt as **admissible for strict conditionals** all and only those Kripke models in which $\mathscr{R}$ is reflexive.

We began this section with five arguments, the forms of the first three of which were as follows:

B ⊢ A → B
~A ⊢ A → B
~(A → B) ⊢ A & ~B

All three arguments are valid, reading '→' as the material conditional, but all are outrageous reading '→' as an English conditional. (Indeed, the first two have often been called the "paradoxes of material implication.") Yet, if we replace '→' by '⥽', we get the reasonable result that none of the three arguments is valid.

Let's consider the sequent 'B ⊢ A ⥽ B' first. To facilitate comparison with the first argument above, think of 'B' as meaning "Socrates grew to manhood" and 'A' as meaning "Socrates died as a child." Socrates did, of course, grow to manhood; yet it is (or was) possible for him to have died as a child and not grown to be a man. So the premise is true and the conclusion false. To represent this counterexample formally we need two worlds: world 1, representing the actual world, a world in which Socrates did grow to be a man, and a merely possible world, world 2, in which he died as a child:

METATHEOREM: The sequent 'B ⊢ A ⥽ B' is invalid relative to the admissible models for strict conditionals.

PROOF: Consider the Kripkean model $\mathcal{V}$ in which

$\mathcal{W}_{\mathcal{V}} = \{1, 2\}$ $\quad$ $\mathcal{V}$('A', 1) = F
$\mathcal{R} = \{<1, 1>, <1, 2>, <2, 2>\}$ $\quad$ $\mathcal{V}$('B', 1) = T
$\mathcal{V}$('A', 2) = T
$\mathcal{V}$('B', 2) = F

This model is admissible for strict conditionals, because $\mathcal{R}$ is reflexive. Since $1\mathcal{R}2$, $\mathcal{V}$('A', 2) = T, and $\mathcal{V}$('B', 2) ≠ T, it follows that $\mathcal{V}$('A ⥽ B', 1) ≠ T. Thus, since $\mathcal{V}$('B', 1) = T, the sequent is invalid. QED

The same counterexample establishes the invalidity of the sequent '~A ⊢ A ⥽ B'. In the actual world, Socrates did not die as a child, which makes '~A' true; but, since it was possible (relative to the actual world) that he did and never grew to manhood, 'A ⥽ B' is false. Proof of the invalidity of this sequent is left to the reader (see the exercise at the end of this section).

The sequent '~(A ⥽ B) ⊢ A & ~B' is also invalid. The fact that A does not necessitate B tells us nothing about the truth values of either A or B. The formal treatment of this problem is left entirely to the reader.

Though reasoning in these counterintuitive patterns is not valid for the strict conditional, many natural and familiar patterns—modus ponens and modus tollens, for example—are valid. So far, then, the strict conditional seems to answer accurately to our understanding of 'if . . . then' in English.

But the situation is not as tidy as it seems. The last two of our five arguments have the forms hypothetical syllogism and contraposition, respectively. These forms, as we saw, seem invalid for English conditionals, but they are valid for the strict conditional. We shall prove this for contraposition only, leaving the proof for hypothetical syllogism as an exercise:

METATHEOREM: The sequent 'A ⥽ B ⊢ ~B ⥽ ~A' is valid relative to the admissible models for strict conditionals.

PROOF: Suppose for reductio that this sequent is invalid relative to the admissible models. Then there exists some admissible model containing a world w such that $\mathcal{V}$('A ⥽ B', w) = T and $\mathcal{V}$('~B ⥽ ~A', w) ≠ T. Since $\mathcal{V}$('~B ⥽ ~A', w) ≠ T, there exists a world u such that $w\mathcal{R}u$, $\mathcal{V}$('~B', u) = T and $\mathcal{V}$('~A', u) ≠ T. Hence by the valuation rule for negation $\mathcal{V}$('A', u) = T and $\mathcal{V}$('B', u) ≠ T. But since $w\mathcal{R}u$, this implies that $\mathcal{V}$('A ⥽ B', w) ≠ T, and so we have a contradiction.

Therefore the sequent 'A ⥽ B ⊢ ~B ⥽ ~A' is valid relative to the admissible models for strict conditionals. QED

The fact that it makes hypothetical syllogism and contraposition valid might be seen as an advantage, rather than a disadvantage of the strict conditional. These are, after all, common and persuasive forms of reasoning. But since they are apparently invalid for at least some English conditionals, their validity for the strict conditional is in fact a disadvantage, insofar as the strict conditional is supposed to accurately analyze the English.

The disparity beween strict and English conditionals also crops up in "paradoxes" reminiscent of the paradoxes of material implication. These concern the sequents '□B ⊢ A ⥽ B' and '~◇A ⊢ A ⥽ B', both of which are "paradoxically" valid. Reading '⥽' as an English conditional, we can produce preposterously invalid instances. For example:

It is necessarily the case that humans are mortal.
∴ If humans are immortal, then humans are mortal.

and

It is impossible for Socrates to be a rock.
∴ If Socrates is a rock, then Socrates is a chihuahua.

In both cases (thinking of the necessity or possibility invoked in the premise as practical rather than, say, logical), the premise is true and the conclusion (understood as an English conditional) is false. Thus it is rash to identify even strict conditionals with their English counterparts.

Exercise 12.3

Prove the following metatheorems for the logic of strict conditionals—whose admissible models are all Kripkean models in which $\mathscr{R}$ is reflexive.

1. The sequent '~A ⊢ A ⥽ B' is invalid.
2. The sequent '~(A ⥽ B) ⊢ A & ~B' is invalid.
3. The sequent '□B ⊢ A ⥽ B' is valid.
4. The sequent '~◇A ⊢ A ⥽ B' is valid.
5. The formula 'A ⥽ A' is valid.
6. The sequent 'A ⥽ B, A ⊢ B' is valid.
7. The sequent 'A ⥽ B, ~B ⊢ ~A' is valid.
8. The sequent 'A ⥽ B, B ⥽ C ⊢ A ⥽ C' is valid.
9. The sequent 'A, B ⊢ A ⥽ B' is invalid.
10. The sequent 'A, ~B ⊢ ~(A ⥽ B)' is valid.

12.4 LEWIS CONDITIONALS

What, then, does the English 'if . . . then' mean? Logicians are divided on this question, and it is presumptuous even to assume that English conditionals all mean the same thing. But for a good many English conditionals, the best answer I know of is this picturesque morsel from David Lewis:

> '*If kangaroos had no tails, they would topple over*' seems to me to mean something like this: In any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over.[5]

More generally, we may say:

> *If* Φ *then* Ψ is true in a world w iff in all the worlds most like w in which Φ is true, Ψ is also true.

Contrast this with similarly stated truth conditions for the strict conditional:

> Φ ⥽ Ψ is true in a world w iff in all the worlds possible relative to w in which Φ is true, Ψ is also true.

Here, of course, we have to specify the relevant sense of possibility; that is, we have to know which form of alethic modality we are dealing with.

Lewis's truth conditions, however, do not require us to specify the sort of possibility we intend. The antecedent of the conditional does that automatically. We are to consider, not all practically, or physically, or logically possible worlds, but rather all the worlds most like ours in which the antecedent is true.

[5] *Counterfactuals* (Cambridge: Harvard University Press, 1973), p. 1. The following analysis uses the truth conditions given on p. 25, which prevent vacuous truth.

As a result, Lewis's truth conditions do not flounder, as those for the strict conditional do, when the antecedent is impossible. With the strict conditional, if there are no possible worlds in which the antecedent is true, then, trivially, the consequent is true in all such worlds—no matter what that consequent may say. Thus, as we saw, given that it is impossible for Socrates to be a rock and reading 'if . . . then' as a strict conditional using the practical sense of possibility, we must concede that the absurd sentence 'If Socrates is a rock, then Socrates is a chihuahua' is true.

Lewis's semantics avoids this consequence. Having not found any practically possible worlds in which the antecedent is true, we do not simply punt and declare the conditional true; rather, rising to the challenge, we consider more and more remote possibilities. In our example, since it seems impossible, even in the metaphysical sense, for Socrates to be a rock, we must extend our consideration all the way out to mere logical possibilities before finding worlds in which he is. When we come to the first of these (i.e., those most like the actual world—so that, despite the fact that in them Socrates is a rock, as much as possible of the rest of the world is as it actually is), we stop. Then we ask: Is Socrates a chihuahua in all of these worlds? The answer, pretty clearly, is no. And so the sentence 'If Socrates is a rock, then Socrates is a chihuahua' is false.

Though this example is artificial, the general procedure is not. When considering whether or not a statement of the form *if* Φ *then* Ψ is true, we do in fact imagine things rearranged so that Φ is true and then try to determine whether under these new conditions Ψ would also be true. But we do this conservatively, excluding ways of making Φ true that are wilder than necessary. That is, we try to keep as much as possible of our world unchanged. Most of us would assent to the conditional 'if kangaroos had no tails, they would topple over', even though we can envision worlds in which kangaroos have no tails but do not topple over because, for example, there is no gravity. But the conditional asks us only to entertain the possibility of depriving kangaroos of tails. Depriving them of gravity too is impertinent; it changes the world in ways not called for by the conditional's antecedent. Hence, depriving kangaroos of gravity is not relevant to determining the truth value of the conditional.

Yet there may be more than one equally conservative way of changing the world to make the antecedent true. Consider the conditional 'if forests were not green, then they would not be so beautiful.' Now there are many worlds equally minimally different from ours in which the antecedent is true: worlds in which forests are brown or blue or yellow, and so on. Only if we regarded the consequent as true in all of these worlds would we assent to the conditional. If we regarded brown forests, but not blue, as more beautiful than green, we would judge the conditional false. That's why Lewis stipulates that *if* Φ *then* Ψ is true iff among *all* the worlds (plural) most like w in which Φ is true, Ψ is also true.

The one element required by Lewis's semantics that has not appeared in any model we have considered so far is a measure of "closeness" or similarity among worlds. While Lewis uses these terms, I prefer to think in terms of degree of possibility; where Lewis would speak of worlds as being more or less similar to a given world, I regard them as being more or less possible relative to that world.

There are two reasons for this. First, it allows us to make the transition to Lewis's semantics without introducing the entirely new concept of similarity; the only change we need make is to think of $\mathscr{R}$ as having degrees, rather than being an all-or-nothing affair. Second, similarity is symmetric; A is precisely as similar to B as B is to A. But, as we have seen, $\mathscr{R}$ should not, in general, be assumed to be symmetric.

How might a model treat $\mathscr{R}$ as a matter of degree? The simplest way would be to set up some arbitrary scale (say, from 0 to 1), where 0 represents complete lack of relative possibility and 1 the highest degree of relative possibility. Presumably, then, each world is maximally possible relative to itself, that is, has degree 1 of $\mathscr{R}$ to itself, and all other worlds are less possible relative to it.

Such a numerical scale is, however, not quite satisfactory. There is no a priori reason to suppose that degrees of relative possibility can be ordered like the real numbers from 0 to 1. A more abstract mathematical treatment of order could address this problem but would introduce complexities beyond the scope of this book. We shall, then, at the risk of slight (and not very significant) oversimplification, suppose degrees of $\mathscr{R}$ can be ranked along a 0 to 1 scale.

Accordingly, instead of treating $\mathscr{R}$ as a set of pairs, as we did before, we may treat it as a set of triples, in which the third member is a number from 0 to 1, indicating the degree to which the second member is possible relative to the third. Thus for a model consisting of worlds 1 and 2, we might have, for example:

$$\mathscr{R} = \{\langle 1, 1, 1\rangle, \langle 1, 2, 0.7\rangle, \langle 2, 1, 0\rangle, \langle 2, 2, 1\rangle\}$$

This means that worlds 1 and 2 are each fully possible relative to themselves, world 2 is possible relative to world 1 with a degree of 0.7, and world 1 is not at all possible relative to world 2. Rather than writing this all out in English, let's use the notation $\mathscr{R}(1, 2) = 0.7$ to mean that the degree to which world 2 is possible relative to world 1 is 0.7. We shall stipulate that

(1) each pair of worlds in the model must be assigned a number from 0 to 1

and that

(2) no pair of worlds may be assigned more than one number

so that for any worlds x and y in the model, $\mathscr{R}(x, y)$ will exist and will be unique. (Where in a Kripkean model we would say that it is not the case that $x\mathscr{R}y$, now we shall say $\mathscr{R}(x, y) = 0$.) We further stipulate that

(3) for any worlds x and y, $\mathscr{R}(x, y) = 1$ iff $x = y$.

This implies that no world is as possible relative to a world x as x itself is. A Lewis model, then, will be exactly like a Kripkean model except for these differences in $\mathscr{R}$.

Lewis represents his conditional formally as the binary operator '$\Box\!\!\rightarrow$', and we shall do likewise. But we shall differ from Lewis in reading this operator simply as "if . . . then." Lewis reads $\Phi \Box\!\!\rightarrow \Psi$ as "if Φ were the case, Ψ would be the case," confining his analysis to so-called subjunctive or counterfactual conditionals. But

I am persuaded that this analysis is more broadly applicable.[6] Its formal truth clause is as follows:

> $\mathcal{V}(\Phi \,\Box\!\!\rightarrow \Psi, w) = \text{T}$ iff there is some world u such that $\mathcal{V}(\Phi, u) = \text{T}$, and there is no world z such that $\mathcal{R}(w, z) \geq \mathcal{R}(w, u)$, $\mathcal{V}(\Phi, z) = \text{T}$, and $\mathcal{V}(\Psi, z) \neq \text{T}$.

This is just a transcription in our new terminology of the informal truth conditions given above. The world u is some arbitrary one of the worlds most possible relative to the actual world in which the antecedent Φ is true. We are saying, in other words, that $\Phi \,\Box\!\!\rightarrow \Psi$ is true at w iff

1. Φ is true in some world u, which is such that
2. there is no world at least as possible relative to w as u is in which Φ is true and Ψ is not.

Clause 2 implies that Ψ is true in u, as well as in any worlds more possible relative to w in which Φ is true. Putting both clauses together, this is to say that in all the worlds most possible relative to w in which Φ is true, Ψ is also true. The corresponding falsity clause is

> $\mathcal{V}(\Phi \,\Box\!\!\rightarrow \Psi, w) = \text{F}$ iff for all worlds u such that $\mathcal{V}(\Phi, u) = \text{T}$ there is some world z such that $\mathcal{R}(w, z) \geq \mathcal{R}(w, u)$, $\mathcal{V}(\Phi, z) = \text{T}$, and $\mathcal{V}(\Psi, z) \neq \text{T}$.

If we wish to retain the operators '$\Diamond$' and '$\Box$', we can do so in the Kripkean fashion, by stipulating that for any worlds x and y, $x\mathcal{R}y$ iff $\mathcal{R}(x, y) \neq 0$. That is, y counts as accessible from x if and only if y is accessible to even the slightest degree from x. This allows the standard Kripkean clauses to be used for these operators.

We shall illustrate the use of Lewis semantics first by proving that modus ponens is valid for a Lewis conditional:

METATHEOREM: The sequent 'A $\Box\!\!\rightarrow$ B, A $\vdash$ B' is valid for Lewis models.

PROOF: Assume for reductio that this sequent is invalid; that is, there is a Lewis model containing a world w such that $\mathcal{V}$('A $\Box\!\!\rightarrow$ B', w) = T, $\mathcal{V}$('A', w) = T, and $\mathcal{V}$('B', w) $\neq$ T. Now by the definition of a Lewis model, for any world u, $\mathcal{R}(w, w) \geq \mathcal{R}(w, u)$. Hence for all worlds u, there is some world z, namely w, such that $\mathcal{R}(w, z) \geq \mathcal{R}(w, u)$, $\mathcal{V}$('A', z) = T, and $\mathcal{V}$('B', z) $\neq$ T. Hence, in particular, for all worlds u such that $\mathcal{V}$('A', u) = T, there is some world z such that $\mathcal{R}(w, z) \geq \mathcal{R}(w, u)$, $\mathcal{V}$('A', z) = T, and $\mathcal{V}$('B', z) $\neq$ T. But this is to say that $\mathcal{V}$('A $\Box\!\!\rightarrow$ B', w) $\neq$ T, and so we have a contradiction.

Hence the sequent 'A $\Box\!\!\rightarrow$ B, A $\vdash$ B' is valid for Lewis models. QED

[6] See Michael Kremer, "'If' Is Unambiguous," *Nous* 21 (1987): 199–217, for a fuller discussion of this point.

It turns out, however, that contraposition and hypothetical syllogism, which we saw were invalid for English conditionals, are both also invalid for Lewis's conditionals. We shall prove this for contraposition, leaving hypothetical syllogism as an exercise. To set the stage, consider the invalid instance of contraposition mentioned above:

If the Atlantic is an ocean, then it's a polluted ocean.
∴ If the Atlantic is not a polluted ocean, then it's not an ocean.

According to Lewis, we evaluate a conditional by considering the worlds most similar to (or, in my terms, most *possible* relative to) the actual world in which the antecedent is true. In the case of this argument's premise, there is only one such world—the actual world itself—for the Atlantic is in fact an ocean. We now check to see if the consequent is true among all members of this (one-membered) class of worlds. And indeed it is, for the Atlantic *is* a polluted ocean. Therefore the premise is true in the actual world.

We then subject the conclusion to the same procedure. The conclusion's antecedent is not true in the actual world, so we must move in imagination out to those worlds most like the actual world (or most *possible* relative to the actual world) in which the Atlantic is pristine. Presumably there are many approximately equally possible ways in which this could have happened. The Industrial Revolution might never have occurred; or we might have developed an ecological conscience before it did; or we might have developed technology for cleaning up oceans. The details matter little; for, whatever we imagine here, it will not include the Atlantic's being transmuted into something other than an ocean. That possibility is much wilder than these others. It seems not even to be a metaphysical possibility, but merely a logical one. The others are all physical, if not practical, possibilities. In none of these more "homey" possibilities is the Atlantic not an ocean. Therefore the conditional's consequent is false in all the worlds most like the actual world in which its antecedent is true. And so the conditional is false.

We can model this counterexample with a domain of two worlds: world 1, representing the actual world, and world 2, representing one of the "homey" worlds in which the Atlantic is pristine but remains an ocean. Read 'A' as "the Atlantic is an ocean" and 'B' as "the Atlantic is a polluted ocean":

METATHEOREM: The sequent 'A □→ B ⊢ ~B □→ ~A' is invalid for Lewis models.

PROOF: Consider the model $\mathcal{V}$ defined as follows:

$\mathcal{W}_{\mathcal{V}} = \{1, 2\}$	$\mathcal{V}(\text{'A'}, 1) = \text{T}$
$\mathcal{R} = \{<1, 1, 1>, <1, 2, 0.7>, <2, 1, 0.5>,$	$\mathcal{V}(\text{'B'}, 1) = \text{T}$
$<2, 2, 1>\}$	$\mathcal{V}(\text{'A'}, 2) = \text{T}$
	$\mathcal{V}(\text{'B'}, 2) = \text{F}$

This meets conditions 1–3 and so is a Lewis model. Clearly there is no world z such that $\mathcal{R}(1, z) > \mathcal{R}(1, 1)$, $\mathcal{V}(\text{'A'}, z) = \text{T}$, and $\mathcal{V}(\text{'B'}, z) \neq \text{T}$. But

$\mathscr{V}$('A', 1) = T. Hence there is a world u, namely world 1, such that $\mathscr{V}$('A', u) = T and there is no world z such that $\mathscr{R}$(1, z) ≥ $\mathscr{R}$(1, u), $\mathscr{V}$('A', z) = T, and $\mathscr{V}$('B', z) ≠ T. Therefore $\mathscr{V}$('A □→ B', 1) = T. Now there is only one world u such that $\mathscr{V}$('~B', u) = T: This is world 2. Yet $\mathscr{R}$(1, 2) ≥ $\mathscr{R}$(1, 2), $\mathscr{V}$('~B', 2) = T, and $\mathscr{V}$('~A', 2) ≠ T. Thus for all worlds u such that $\mathscr{V}$('~B', u) = T, there is some world z, namely world 2, such that $\mathscr{R}$(1, z) ≥ $\mathscr{R}$(1, u), $\mathscr{V}$('~B', z) = T, and $\mathscr{V}$('~A', z) ≠ T. But this is to say that $\mathscr{V}$('~B □→ ~A', 1) = F. Thus since, as we saw above, $\mathscr{V}$('A □→ B', 1) = T, it follows that the sequent 'A □→ B ⊢ ~B □→ ~A' is invalid. QED

Further investigation of Lewis's semantics reveals many more respects in which his conditionals fit our intuitions about English conditionals (see the exercise below). Of the conditionals we have examined, Lewis's is surely the best approximation to the English. But whether it is uniquely *correct* as a formal semantics for the English conditional remains a disputed question.

Exercise 12.4

Prove the following metatheorems for '□→' using Lewis models.

1. The sequent 'A □→ B, ~B ⊢ ~A' is valid.
2. The sequent 'A □→ B, B □→ C ⊢ A □→ C' is invalid.
3. The sequent '~A ⊢ A □→ B' is invalid.
4. The sequent 'B ⊢ A □→ B' is invalid.
5. The sequent '~(A □→ B) ⊢ A & ~B' is invalid.
6. The sequent 'A, B ⊢ A □→ B' is valid.
7. The sequent 'A, ~B ⊢ ~(A □→ B)' is valid.
8. The sequent 'A □→ C ⊢ (A & B)□→ C' is invalid.
9. The sequent '□B ⊢ A □→ B' is invalid.
10. The sequent '~◇A ⊢ A □→ B' is invalid.