

PATRICIA A. BLANCHETTE

## MODELS AND MODALITY

**ABSTRACT.** This paper examines the connection between model-theoretic truth and necessary truth. It is argued that though the model-theoretic truths of some standard languages are demonstrably “necessary” (in a precise sense), the widespread view of model-theoretic truth as providing a general guarantee of necessity is mistaken. Several arguments to the contrary are criticized.

Models, in the sense in which logicians use the term “model”, are often taken to be arbiters of a certain sort of modality. In particular, it is often held that *truth on every model implies necessary truth*. Where necessary truth can be taken, perhaps only metaphorically, to be truth across possible worlds, the principle in question can be stated succinctly:

(\*) Truth across models implies truth across possible worlds.

The purposes of this paper are to attempt to clarify what this principle might mean, and to investigate the grounds one might have for supposing it to be true.

### 1. PRELIMINARIES

#### 1.1. *Clarifying (\*)*

At first glance, (\*) looks as if it should license inferences from claims of the form “ $\varphi$  is true on every model” to the corresponding “ $\varphi$  is true in every possible world”. But this cannot be what is intended. The kinds of things which have truth-values on models are the well-formed formulas of formal languages: they are purely syntactic items, and hence not the kinds of things which have truth-values either at the actual world or at possible worlds. The formula “ $Fa$ ”, for example, is true on some models and false on others (typically, true on model  $m$  iff  $m(\langle a \rangle) \in m(\langle F \rangle)$ ), but it has no truth value at worlds. The kinds of things which have truth-values at worlds, on the other hand, are either the fully-interpreted sentences of (typically) natural languages, or the nonlinguistic entities expressed by these sentences. Let us use the term “proposition” for those entities which have truth-values at possible worlds. To say that the proposition *Alice is*



*French* is true at a world  $w$  is just to say that, at  $w$ , Alice is French. We need not take a stand here on the precise nature – linguistic or nonlinguistic – of propositions; the important point for our purposes is that because propositions are true or false at possible worlds, they are not the bare syntactic formulae of formal languages. They are therefore not the kinds of things directly assigned truth-values by models.

The claim (\*), then, if it is to make sense at all, cannot mean that the items true on every model are also true at every possible world. The intention behind (\*) is rather that the propositions expressed by formulas true on every model are themselves true at every possible world. We take (\*) to license, for instance, the inference from the model-theoretic truth of the formula

$$(2) \quad (Fa \vee \sim Fa)$$

to the necessary truth of the proposition

$$(3) \quad \textit{Either Alice is French or Alice is not French,}$$

and of any other proposition reasonably expressible by (1). As a first attempt at clarifying (\*), then, we have:

If a formula  $\varphi$  is true on every model, then each proposition which  $\varphi$  can reasonably be used to express is true at every possible world.

The “reasonably” in this gloss is important. For any of the usual formal languages, there are straightforward restrictions on the allowable propositions expressible by a formula. Thus for instance we usually take “ $(Fa \vee \sim Fa)$ ” to be the right kind of formula to express the proposition *either Jones is happy or Jones is not happy*, but not to express the proposition *snow is white*. And in general, we usually take a formal language to come along with a handful of (implicit) restrictions on the kinds of propositions expressible by its formulas. The precise nature of these constraints will be discussed below. Here we simply introduce some terminology: where  $L$  is a formal language, an assignment of propositions to  $L$ ’s closed formulas will be called a *reading* of  $L$ . When, as is usual,  $L$  comes along with various restrictions on its readings, a reading meeting those restrictions will be called an *acceptable* reading of  $L$ .

For example, let  $L$  be a familiar first-order language, of the kind found in standard introductory logic texts. There are acceptable readings of this language which assign the proposition *either Jones is happy or Jones is not happy* to “ $(Fa \vee \sim Fa)$ ”, and which assign the proposition *either snow*

is white or snow is not white to that formula, but no acceptable reading assigns to this formula the proposition *snow is white*. It is a result of these restrictions on its acceptable readings that, for example, “ $(Fa \ \& \ \sim Fa)$ ” expresses a false proposition on each acceptable reading of  $L$ , “ $(Fa \ \vee \ \sim Fa)$ ” expresses truths on each such reading, and “ $Fa$ ” expresses truths on some, and falsehoods on other acceptable readings of that language. Where  $R$  is a reading of a language  $L$  and  $\varphi$  a formula of  $L$ , let  $R(\varphi)$  be the proposition assigned by  $R$  to  $\varphi$ . The inferences in which we are interested, then, are inferences from a formula  $\varphi$ ’s truth on every model to a proposition  $R(\varphi)$ ’s truth at every possible world. For a particular formal language  $L$ , then, the principle in question is:

- (I) For every wff  $\varphi$  of  $L$ : If  $\models \varphi$ , then for every acceptable reading  $R$  of  $L$ ,  $R(\varphi)$  is true at every possible world.

Here and in what follows, “ $\models \varphi$ ” abbreviates “ $\varphi$  is true on each of  $L$ ’s models”. (I) is the clarified version of (\*), and is the principle with which this paper will be primarily concerned.

### 1.2. *Models, Readings, and Languages*

Before examining (I), a few remarks on models and readings are in order. Because both models and readings are often called “interpretations”, it is easy and often natural to overlook the differences between the two. But their differences will be important in what follows.

Typically, a model “interprets” a formal language by specifying a domain (or multiple domains) of quantification, and assigning individuals and sets to the simple individual-, relation-, and function-symbols of that language. A model also assigns truth-values to formulas, via the language’s defined *true-on* relation: to say that a formula is true on a model  $m$  is just to say that  $\varphi$  bears that defined relation to  $m$ . In short, a model interprets a language by assigning set-theoretic constructions to parts of formulas, and truth-values to formulas. A model does not, however, assign to formulas anything like propositions. The important point for our purposes is that a model does not assign to a formula anything which has truth-values at possible worlds. Though formulas are true-on and false-on models, there is no sense in which a formula is *necessarily* (or contingently) true (or false) on a model.<sup>1</sup>

A reading, on the other hand, “interprets” a language by assigning propositions to formulas. The acceptable readings of a language are typically given by the requirement that some symbols (the “logical constants”) be assigned fixed meanings, that members of other syntactic categories be

assigned meanings from a specific range of alternatives, and that the assignments of propositions to whole formulas be determined as a function of the assignments to their parts. The truth of a formula under a reading, unlike its truth on a model, has nothing to do with the defined *true-on* relation. To say that  $\varphi$  is true under reading  $R$  is just to say that the proposition  $R(\varphi)$  is, in the ordinary sense, true. And to say that  $R(\varphi)$  is true at a possible world  $w$  is just to say that the proposition  $R(\varphi)$  is such that, had  $w$  been actual, that proposition would have been true. Where  $\varphi$  is a formula and  $R$  a reading,  $R(\varphi)$  is either necessarily or contingently true or false.

The standard practice of ignoring the distinction between models and readings is to some extent justified by the fact that models and readings are typically very intimately connected with one another. Ordinarily, for instance, if there is a reading  $R$  of a language on which “ $Fa$ ” expresses the proposition *two is an even integer*, then there will be a model  $m$  of that language which assigns to “ $a$ ” the number two, and to “ $F$ ” the set of even integers. In general: say that a reading  $R$  and a model  $m$  for a language *correspond* just in case the domain of  $m$  is the same as the range of the first-order variables under  $R$ ,  $m$  and  $R$  assign the same individual to each individual constant, and  $m$  assigns to the function- and relation-terms the extensions of those functions and relations assigned them by  $R$ .<sup>2</sup> A model is, essentially, an extensional version of any reading to which it corresponds.

The parallel between models and readings is largely mediated by the *true-on* relation. Standardly, this relation is defined in such a way that a formula of a given language is *true on* a model iff that formula expresses a truth under any corresponding reading. That is, an important characteristic of typical languages is that they satisfy condition (T):<sup>3</sup>

- (T) For every formula  $\varphi$  of language  $L$ , every model  $m$  of  $L$ , and every acceptable reading  $R$  of  $L$ : If  $m$  and  $R$  correspond, then  $\varphi$  is true on  $m$  iff  $R(\varphi)$  is true.

Typically, the guarantee of (T) is given in part by defining *true-on* so that the semantic role of the logical constants with respect to models parallels their semantic role with respect to acceptable readings. Thus, e.g., if each acceptable reading assigns the meaning *and* to the symbol “&”, then the definition of *true-on* will include the stipulation that a formula of the form ‘ $(\varphi \ \& \ \psi)$ ’ is true-on a model  $m$  iff both of  $\varphi$  and  $\psi$  are as well. The upshot of this principle is that a formula is true on a model just in case it says something true when read as being “about” the entities, relations, etc. out of which the model is constructed. This is what makes it reasonable to call the defined relation “true-on”.

Every familiar language (at least, every familiar language with a coherent *true-on* relation) satisfies (T). There is nothing interesting, for our purposes, about those languages which fail to satisfy (T). Such languages typically fail (I) for the mundane reason that in these cases, truth on a model has nothing to do with truth. Henceforth, then, by a *language* we will mean a triple  $\langle F, M, A \rangle$  which satisfies (T), where  $F$  is a class of formulas,  $M$  a model-theoretic apparatus (consisting of a class of models and a binary *true-on* relation between formulas and models), and  $A$  is a class of acceptable readings.

## 2. FROM MODEL-THEORETIC TRUTH TO TRUTH

### 2.1. *Correspondence and Truth*

A crucial question to ask of a language is whether it has “enough models” to correspond with each of its readings. That is, the question is whether the language  $L$  satisfies (MOD):

(MOD) For every acceptable reading of  $L$ , there is a corresponding model for  $L$ .

(MOD) is simply the natural constraint that if it is acceptable to *read* the wffs of  $L$  as quantifying over a particular collection  $C$ , and in such a way that its constant terms stand for particular individuals and relations, then there is a model for  $L$  whose domain is  $C$ , and which assigns those individuals and the extensions of those relations to the same constant terms.

The importance of (MOD) is that if a language  $L$  satisfies it, then there is a straightforward sense in which each of its acceptable readings is given a precise, formal representation by (at least) one of its models. We are already guaranteed by (T) that a formula is true on a model iff it is true under a corresponding reading; further satisfaction of (MOD) entails that those formulas true on every model will express truths under each of the acceptable readings of the language. That is, if a language  $L$  satisfies (MOD), then:

(II) For every wff  $\varphi$  of  $L$ : if  $\models \varphi$ , then for every acceptable reading  $R$  of  $L$ ,  $R(\varphi)$  is true.

If a language  $L$  satisfies (II), we are straightforwardly justified in making the familiar inferences from “ $\varphi$  is true on every model” to “ $\varphi$  is true no matter how you read it”.

Because necessary truth implies truth, satisfaction of (II) is a necessary condition for satisfaction of (I).

## 2.2. (MOD) and (II)

Though (MOD) is a fairly naturally-motivated condition on formal languages, many familiar languages do not satisfy it. Consider the following first-order arguments and typical readings:

<p>(A1) <math>(\forall x)(x = x)</math></p> <hr style="width: 20%; margin-left: 0;"/> <p><math>\therefore a = a</math></p>	<p>Everything is self-identical</p> <hr style="width: 20%; margin-left: 0;"/> <p><math>\therefore</math> Jones is self-identical</p>
<p>(A2) <math>(\forall x)(x \notin x)</math></p> <hr style="width: 20%; margin-left: 0;"/> <p><math>\therefore \emptyset \notin \emptyset</math></p>	<p>No set is a member of itself</p> <hr style="width: 20%; margin-left: 0;"/> <p><math>\therefore</math> The empty set is not a member of itself.</p>
<p>(A3) <math>(\forall x)(Ox \rightarrow (\exists y)Syx)</math></p> <p><math>Oa</math></p> <hr style="width: 20%; margin-left: 0;"/> <p><math>\therefore (\exists y)Sya</math></p>	<p>Every ordinal has a successor</p> <p>3 is an ordinal</p> <hr style="width: 20%; margin-left: 0;"/> <p><math>\therefore</math> 3 has a successor.</p>

Typical first-order languages countenance these readings of premises and conclusions. But given the usual constraint that a model's domain is a set (and not a proper class), no models correspond to these readings. In this case, (MOD) fails: the class of models for the typical first-order language  $L$  does not contain extensional versions of each of  $L$ 's acceptable readings.

Call a reading of a language a "set-reading" if the range it assigns to the first-order variables is a set, and a "class-reading" otherwise. The failure of (MOD) in the ordinary first-order case is due to the existence of class-readings, to which no models correspond. Nevertheless, (II) holds in this case, despite the failure of (MOD). This is at least mildly interesting. Because of (T), a formula's truth on a model implies its truth under every corresponding reading. Thus a formula's truth on every model guarantees its truth under all of those readings which have corresponding models. One might think, then, that when (MOD) fails, the readings without corresponding models might occasionally falsify formulas which are true on every

model. But as it turns out, this does not happen in the ordinary first-order case, since:

- (X) If a formula  $\varphi$  of an ordinary first-order language  $L$  expresses a false proposition on some class-reading  $R$  of  $L$ , then it also expresses a false proposition on some set-reading  $R'$  of  $L$ <sup>4</sup>.

(By an “ordinary” first-order language, we mean one with the usual classes of models and of readings, counting only the usual quantifiers and sentential connectives as logical constants.) First-order formulas do not discriminate, in this sense, between sets and classes. As long as every set-reading has a corresponding model, as is the case with every ordinary first-order language, any formula true on every model is, via (X), true under every reading. Thus despite the fact that models do not precisely mirror readings, the differences are insufficient, in the first-order case, to derail the familiar inferences from “ $\varphi$  is true on every model” to “ $\varphi$  is true no matter how we read it”. First-order models are in this sense good stand-ins for first-order readings.

The failure of (MOD) becomes more problematic when we expand the expressive resources of the first-order language. To borrow an example from Vann McGee: consider the language obtained from an ordinary first-order language by adding the quantifier “There exist absolutely infinitely many”,  $(\exists^{\text{AI}}x)$ .<sup>5</sup> The formula “ $\sim(\exists^{\text{AI}}x)x = x$ ” is true on every first-order model, while its natural reading, namely the proposition *there aren't absolutely infinitely many things* is presumably false. Thus we have a counterexample to (II) (and hence to (I) as well). Truth across models in this case not only fails to guarantee necessary truth; it fails even to guarantee truth.

Another candidate for a counterexample concerns second-order set theory. Let “ZFC2” abbreviate the conjunction of the second-order ZFC axioms.<sup>6</sup> This sentence, under its ordinary reading, would be taken by most people to express a true proposition.<sup>7</sup> For it describes what we typically suppose to be the case about the (or a) set-theoretic hierarchy. But it is not at all clear that it has a model, since a model would require the existence of an inaccessible cardinal, something not required by either the plausibility or the truth of either ZFC2 or of any other widely-accepted principles of set theory. If there are no inaccessible cardinals, then ZFC2 expresses, under its “intended” reading, a truth, but has no model. In this case, the negation of ZFC2 is true on every model, but expresses a falsehood on its intended reading, *contra* (II). And of course, if there are no inaccessible cardinals, then we have a clear counterexample to (I) as well.

More mundane counterexamples to (MOD) arise with the usual second-order languages (without added quantifiers), even without invoking large

cardinals. These languages typically fail (MOD) straightforwardly in virtue of their inclusion of class-readings. The question of their satisfaction of (II) is less straightforward, for two reasons. First of all, there is no uniform, established set of constraints on the acceptable readings of second-order languages, in the way in which, for the most part, there is for first-order languages. The different options here in the second-order case concern primarily the appropriate way to read the ranges of the higher-order quantifiers: the existential quantifier “ $\exists X$ ”, for instance, is sometimes read as “there is a property”, sometimes as “there is a set”, and sometimes quite differently altogether. Secondly, there are two familiar choices of model-theoretic apparatus for second-order languages: the first invokes the standard, “full” models, while the second uses the alternative “Henkin” models.<sup>8</sup> The variety of different combinations of readings and of classes of models gives us a variety of importantly-different second-order languages, some of which satisfy (II), and some of which do not. Some of the details are as follows:

The sentences “ $\exists X \forall y Xy$ ” and “ $\exists X \forall y (Xy \leftrightarrow y \notin y)$ ” are true on every model of “full” second-order logic, but are false on some fairly natural readings, e.g., on readings on which these sentences say, respectively: “There is a universal set”, and “There is a collection of all the non-self-membered things”. Any second-order language  $L$  with full models which includes such readings will contain numerous counterexamples to (II), and hence to (I).

Second-order languages with full models, but with more-restricted classes of readings do, however, satisfy (II). One such restriction involves counting only “set-readings” as acceptable; in this case, (II) holds (typically) because (MOD) holds.<sup>9</sup> A different restriction comes from employing George Boolos’ “plural quantification” readings, under which no formulas express the objectionable existence-assertions just mentioned; arguably, all “plurally-quantified” readings of model-theoretic truths are indeed true.<sup>10</sup> In both of these cases, the potential counterexamples to (II) are avoided by simply ruling “unacceptable” all those readings of the problematic model-theoretic truths on which they express falsehoods.

Second-order languages with Henkin models satisfy (II) on any of the various standard choices of readings, even when they fail (MOD), for the same reasons that first-order languages do. In this case, the counterexamples to (II) are avoided because the problematic formulae of the kind noted above are not model-theoretic truths.

There are, then, two ways of modifying languages which fail to satisfy (I) and (II) so as to obtain new languages satisfying both. The first is to impose restrictions on acceptable readings in such a way that the false

readings of model-theoretic truths are ruled “unacceptable”. This strategy is illustrated above by the move to set-readings or to plural-quantification readings. The second method is to expand the class of models in such a way that those problematic formulae are falsified on some model. This is illustrated by the move from full models to Henkin models.

There remains a great deal to be said about the relative advantages and disadvantages of the several second-order languages discussed, and of further variants of them. The important point for our purposes is simply that the question of whether a language satisfies (II) will turn on the details of its model-theoretic apparatus and on its range of acceptable readings. The license for inferences from model-theoretic truth to truth will have to be established on a case-by-case basis.

In short, though (II) is clearly satisfied by ordinary first-order languages, its varied status elsewhere means that, as Vann McGee has put it, “it is by no means obvious that being true in every model is any guarantee that a sentence is true”.<sup>11</sup>

### 3. FROM MODEL-THEORETIC TRUTH TO NECESSARY TRUTH

Despite the failure of (II) and hence of (I) in various cases, we do have some straightforward guarantees that (I) will hold for perhaps the two most important languages, those of standard propositional and first-order logics.

#### 3.1. *Standard Propositional Languages*

A straightforward guarantee of (I) in the propositional case follows from a demonstration by Richard Cartwright of a similar result.<sup>12</sup> Put in terms of readings, the demonstration is as follows:

Let  $L$  be a propositional language whose formulas comprise an infinite collection of atomic formulas together with the closure of this collection under the two operations which generate formulas of the form ‘ $\sim \varphi$ ’ and of the form ‘ $(\varphi \rightarrow \psi)$ ’ from a formula  $\varphi$  and a pair  $\langle \varphi, \psi \rangle$  of formulas, respectively.

$L$ ’s “models” are simply the usual rows of truth-tables. Where a valuation of  $L$  is simply a function from  $L$ ’s formulas into  $\{T, F\}$ , we can put this as follows: the class of models of  $L$  is just the class of valuations  $v$  of  $L$  satisfying the following two constraints:

- (M1) For every formula  $\varphi$ ,  $v(\sim \varphi) \neq v(\varphi)$ .
- (M2) For all formulas  $\varphi$  and  $\psi$ ,  $v((\varphi \rightarrow \psi)) = F$  iff  $v(\varphi) = T$  and  $v(\psi) = F$ .

Consider now a particular reading  $R$  of  $L$  which satisfies the following two conditions:

- (C1) For every formula  $\varphi$ ,  $R(\sim \varphi)$  is the negation of  $R(\varphi)$ .
- (C2) For all formulas  $\varphi$  and  $\psi$ ,  $R((\varphi \rightarrow \psi))$  is the material conditional whose antecedent is  $R(\varphi)$  and whose consequent is  $R(\psi)$ .

Given  $R$ , possible worlds induce valuations, in the obvious way: where  $w$  is a possible world, the valuation  $V_w$  induced by  $w$  is defined as follows:

For every wff  $\varphi$ ,  $V_w(\varphi) = T$  iff  $R(\varphi)$  is true at  $w$ .

Note that because there is no possible world at which both a proposition and its negation are true, there is no possible world  $w$  such that  $V_w$  violates constraint (M1). Similar considerations show that for no possible world  $w$  does  $V_w$  violate (M2). So every valuation assigned to  $L$ 's formulas by a possible world (given reading  $R$ ) is also assigned by one of  $L$ 's models. This suffices to show that: *If  $\varphi$  is true on every model, then  $R(\varphi)$  is true at every possible world.*

Note that this demonstration holds for every reading which meets conditions (C1) and (C2). Define a “standard propositional language” to be one whose formulas are as described for  $L$ , whose models are the class of valuations meeting (M1) and (M2), and whose acceptable readings all meet conditions (C1) and (C2). We have shown, then, that (I) holds for standard propositional languages. QED.

### 3.2. *Standard First-Order Languages*

Because the valuations given by models of quantified languages cannot be characterized in terms of constraints which, like (M1) and (M2), correspond directly with constraints on acceptable readings, there is no straightforward way to extend Cartwright's proof to quantified languages. An argument by John Etchemendy, however, provides us a means of demonstrating (I) for a range of first-order languages.<sup>13</sup> Etchemendy points out that, as he would put it, since the theorems of a standard first-order deductive system are truths of logic, and (via the completeness theorem) all first-order model-theoretic truths are theorems, we can conclude that all first-order model-theoretic truths are truths of logic. And since truths of logic are presumably necessary truths, this gives us essentially (I).

This way of putting the issue is not as careful as one might like, since it presupposes both that the language comes along with an “intended interpretation” – i.e., an intended *reading* – and that this reading assigns a

necessarily-true proposition to each theorem. In the more general case, in which the language may not have a particular intended reading, the crucial point is that each acceptable reading assigns only necessary truths to theorems. Say that a deductive system  $D$  for a language  $L$  is *nice* if its theorems express only necessary truths – i.e., if it satisfies (N):

- (N) For every theorem  $\varphi$  of  $D$  and every acceptable reading  $R$  of  $L$ ,  $R(\varphi)$  is true at every possible world.

Virtually every familiar deductive system satisfies this requirement (or *almost* satisfies it; see below), as is easily verified by noting that axioms express necessary truths on acceptable readings, and rules of inference preserve this property. Say that a first-order language is *standard* if there is some complete and nice deductive system for that language. We can now put Etchemendy's point slightly more carefully, as follows: *Every standard first-order language satisfies (I).*

In fact, the most popular first-order languages are not quite “standard”. Those which count formulas of the form “ $\exists x(x = a)$ ” as model-theoretic truths, and which contain the usual range of acceptable readings, clearly fail this requirement. For any complete deductive system for such a language will count these existential formulas as theorems, as is usual; but these formulas express merely contingent truths on some of their acceptable readings. In order to have a genuine guarantee of (I) in the first-order case, we need to stick with first-order languages which are truly standard, i.e., which differ from the most popular first-order languages in including empty models, and which can therefore be completely axiomatized by a deductive system satisfying (N). We will return to a discussion of the problematic existential formulas in Section 4.1 below. For now, the important point is that, aside from a relatively-isolated collection of formulas, ordinary first-order languages do demonstrably satisfy (I).

The interesting general result here is that (I) is satisfied by any language with a nice, complete deductive system. Thus we are assured that (I) holds not just for standard first-order languages, but additionally, e.g., for those second-order languages with Henkin models and the usual choices of acceptable readings.

This result guarantees that we are justified in inferring necessary truth from model-theoretic truth in an important, useful range of cases. It doesn't of course help establish any *general* connection between model-theoretic truth and necessary truth, since it turns essentially on the completeness theorem. What it shows, rather, is that when the model-theoretic truths of a language are also theorems of its deductive system, they will inherit whatever modal character is had by those theorems.

## 4. THE GENERAL CASE

Our results so far show the following: There is an important and familiar range of languages each of which satisfies (I), and which does so because its model-theoretic truths form a subset of its deductive theorems. This range includes the familiar propositional and standard first-order languages. Expansion of the expressive resources beyond this range in various ways, for example by the addition either of new first-order quantifiers or of second-order quantifiers, yields some languages which clearly fail (I) and some whose satisfaction of (I) remains an open question, one which turns on the existence of large cardinals.

The cases so far discussed in which (I) fails are ones in which this failure can be traced to the failure of (II). That is, these have all been cases in which model-theoretic truths fail to express necessary truths under certain acceptable readings because, under those readings, they actually express falsehoods. These have all, of course, been cases in which (MOD) fails.

These are not the only conditions under which (I) fails. Because the *truth* of those propositions expressed by a formula  $\varphi$  under its acceptable readings is a necessary but not sufficient condition for the *necessary* truth of each of those propositions, satisfaction of (II) is a necessary but not a sufficient condition for satisfaction of (I). It is a simple matter to define languages which satisfy (II) but not the more demanding (I), and indeed, languages which satisfy the more stringent (MOD) and (T) without satisfying (I).<sup>14</sup> What, then, of the status of (I) in general? Clearly it cannot be claimed that (I) is true for all languages, nor that it is true for all languages standardly in use. It is, as just noted, true that (I) is satisfied by every language with a nice, complete deductive system. This is an important claim, but, as above, it does not underwrite the intuition we have taken to motivate (I), the intuition that model-theoretic truth itself has an essentially modal character.

One might of course argue that the languages discussed above which pose counterexamples to (I) are somehow beside the point when investigating the connection between model-theoretic truth and necessary truth, in that they trespass upon some standard of “normality” for formal languages. One might argue, that is, that there is some identifiable standard of normality for formal languages which rules out the kinds of counterexamples discussed here, which rules “in” the usual examples of formal languages, and which demonstrably implies (I). This is the argument which must be made if there is to be a defense of any scaled-down version of the view that model-theoretic truth implies necessary truth. The argument must consist of a clear characterization of the class of languages in question, a demon-

stration that (I) holds for this class, and some reason to view this class of languages as representative of the “genuine” character of model-theoretic truth.

There is no such defense in the literature. There are, however, various lines of reasoning which bear some responsibility for the pervasive intuition that there simply *must* be a general connection between model-theoretic truth and necessary truth. We turn next to consideration of several of these lines of reasoning.

#### 4.1. *Models as Representatives*

There is a widely-held intuition that models *represent* possible worlds or parts thereof (“possible situations”, “possible states of affairs”) in such a way that formulas true across models are guaranteed to express necessary truths. Gila Sher, for example, has argued that as long as the models for a language meet a quite general condition, those models will successfully represent possible states of affairs in just this way. Sher’s condition on the class of models for a language is essentially that for each set  $S$  of objects, the language have a model whose domain is  $S$ , and have models assigning every possible configuration of members, subsets etc. of  $S$  to the “non-logical” terms of the language, where the “logical” terms are those which stand for structural or mathematical properties or relations.<sup>15</sup> Sher’s claim is that as long as this constraint is met,

Since every set of objects is the universe of some model, any possible state of affairs – any possible configuration of individuals, properties, relations, and functions – vis-à-vis the extralogical terms of a given formalized language (possible, that is, with respect to their meaning prior to formalization) is represented by some model.<sup>16</sup>

In keeping with Sher’s way of putting the issue, let us take the language  $L$  in question to be fully-interpreted, i.e., to have, as we would put it, a single canonical reading  $R$ . We will also assume that  $R$  is by any ordinary measure “acceptable” – i.e., that it assigns the “right” meanings to the logical constants, and so on. The question is whether Sher is right to suppose that as long as the models of  $L$  meet the general condition just outlined, they will *represent* possible states of affairs in such a way that  $L$  satisfies (I<sub>R</sub>):

- (I<sub>R</sub>) For every formula  $\varphi$  of  $L$ : If  $\varphi$  is true on every model, then  $R(\varphi)$  is true at every possible world.

Let us clarify the sense of *representation* at issue. There are many different senses in which models resemble states of affairs or possible worlds: they can resemble these possibilities in virtue of having the same cardinality, in virtue of having the same population, perhaps in virtue of

bearing some abstract structural similarity to one another, and so on. In the end, though, the only “resemblance” between a model and a state of affairs which matters for the satisfaction of  $(I_R)$  is the agreement in assignments of truth-values to formulas. When  $R$  assigns to “ $Fa$ ” the proposition *Jones is wise*, a model  $m$  represents a state of affairs in which Jones isn’t wise simply by assigning the value *false* to “ $Fa$ ”. It makes no difference to  $m$ ’s representational capacity in this regard whether Jones is a member of  $m$ ’s domain. In general, say that a model  $m$  represents a possible state of affairs in which the proposition  $R(\varphi)$  is true (false) iff  $m$  assigns the value *true* (*false*) to  $\varphi$ . The language  $L$  will satisfy  $(I_R)$  if, and only if,  $L$ ’s models represent possibilities in this way. In particular, where  $\Phi$  is the class of propositions assigned by  $R$  to formulas of  $L$ ,  $L$  will satisfy  $(I_R)$  if and only if every possible situation in which some member of  $\Phi$  is false is represented by one of  $L$ ’s models.

The precise makeup of  $m$ ’s universe will typically be irrelevant to its representational capacity, and, as above, need have nothing in common with the objects involved in the “represented” state of affairs. The question of representation, then, does not – despite Sher’s opening remark above – have much to do with the particular objects found in a model’s domain; it turns only on the various models’ assignments of truth-values.

Sher’s condition on models ensures that, where  $L$  is any one of the usual quantified languages,  $L$ ’s class of models will be combinatorially quite rich: for each non-logical  $n$ -place predicate letter and  $n$ -tuple of terms, there will be models that assign to that  $n$ -tuple of terms an  $n$ -tuple of objects which is a member of the set assigned to the predicate-letter; there will be models that assign to the terms an  $n$ -tuple that is not a member of the set assigned to the predicate letter; there will be models that assign to the predicate letter the empty set, and others that assign it the  $n$ -fold cartesian product of the universe, and so on.

Satisfaction of this condition implies that  $(I_R)$  holds, trivially, for a language consisting entirely of atomic formulas of the form  $Pt_1 \dots t_n$ . For under this condition, none of these formulas will be a model-theoretic truth. Expanding the language by adding the usual propositional connectives also gives a language in which satisfaction of Sher’s condition implies  $(I_R)$ ; this is easily shown along the lines of Cartwright’s proof above.<sup>17</sup>

What about the general case? Is the combinatorial richness of the class of models sufficient to guarantee the existence of a model falsifying  $\varphi$  whenever  $R(\varphi)$  is false in some possible situation? Though this is the crucial claim, Sher simply supposes it without argument. The intuition here would seem to be that if  $R(\varphi)$  is false in some possible situation, then the combinatorial richness of the class of models *must* guarantee the

existence of a model on which  $\varphi$  is assigned the value *false*. But this rather vague intuition simply begs the question. What needs to be established is that the satisfaction of Sher's criterion by a class of models guarantees that that class contains representatives of each possible state of affairs. And this can only be established by demonstrating that satisfaction of this criterion ensures the existence of models which give the right distributions of truth-values.

This central claim, however, is clearly false. All of the languages discussed in Section 2.2 that falsify (I) and (II) meet Sher's criterion. In each case, the class of models is combinatorially rich, and the logical constants express "structural or mathematical" properties.<sup>18</sup> But in none of these cases do models "represent" possible states of affairs in a way sufficient to underwrite the inference from model-theoretic truth to necessary truth. Indeed, in those cases in which the language falsifies (II), the models fail to provide representatives even of the *actual* state of affairs, let alone of non-actual possibilities. This despite the combinatorial richness of the class of models.

The satisfaction of Sher's criterion by a language fails to guarantee that the models of that language represent possibilities in any sense relevant to (I). It is instructive, however, to see what the satisfaction of this criterion does imply. The combinatorial richness of the class of models guaranteed by this criterion entails that from model-theoretic truth in certain cases we can reliably infer "general" truth, in the following sense. Suppose the reading  $R$  corresponds with one of  $L$ 's models. Then, as long as  $L$  satisfies Sher's criterion, we are guaranteed that: If  $\varphi$  is true on every model, then not only is  $R(\varphi)$  true, but so too are all those propositions  $R(\psi)$  expressed by formulas  $\psi$  which are of the same syntactic form as  $\varphi$ .<sup>19</sup>

The richness of the class of models (together with reasonable requirements on acceptable readings) ensures that the model-theoretic truth of a formula entails the truth of the proposition expressed by that formula, and of a typically large class of structurally-similar propositions. That is, under these conditions, the model-theoretic truth of  $\varphi$  implies the actual truth of every member of the set

$$(\alpha) \quad \{R(\psi) \mid \psi \text{ is of the same syntactic form as } \varphi\}.$$

The gulf between the truth of each member of  $(\alpha)$  and the necessary truth of any of its members (including  $R(\varphi)$ ) is most easily seen with the simple example " $(\exists x)(x = a)$ ". On any standard reading of the language, the propositions expressed by formulas of the same syntactic form as  $(\exists x)(x = a)$ , i.e., by those formulas  $(\exists x)(x = b)$ ,  $(\exists x)(x = c)$ , and so on, are all true. That is, for example, each of the propositions *There is*

something identical with Smith, There is something identical with Jones, and so on, is true. But of course many of these propositions are only contingently, not necessarily, true. In the most favorable case, in which each acceptable reading of the language does correspond with some model, we can infer from the model-theoretic truth of a formula  $\varphi$  the actual truth of a wide range of propositions: i.e., of those propositions expressible (under acceptable readings) by formulas of the same syntactic form as  $\varphi$ . But this, of course, is no guarantee of the necessary truth of propositions expressible by  $\varphi$ .

The example “ $(\exists x)(x = a)$ ” also serves, incidentally, as a further counterexample to Sher’s claim, since it too is a model-theoretic truth of a language meeting Sher’s criterion, but which does not express a necessary truth. Sher’s response to the example is as follows: concerning the formula

$$(9) \quad (\exists x)(x = \text{Jean-Paul Sartre}),$$

Sher claims that we have no counterexample to the necessity of model-theoretic truths, because: “since ‘Jean-Paul Sartre’ is a strongly variable term [i.e., a non-logical term], what (9) says is ‘There is *a* Jean-Paul Sartre’, not ‘*The* (French philosopher) Jean-Paul Sartre exists’”.<sup>20</sup> I don’t know what it means to say “there is *a* Jean-Paul Sartre”. But the important point here is that the formula (9) must be taken to be fully-interpreted – i.e., to have a particular fixed reading – if we are to make any sense of the question whether its model-theoretic truth implies the necessary truth of what it expresses. And if the reading in question is any of the readings ordinarily associated with formulas of that kind, then the proposition it expresses is contingent. The model-theoretic truth of the formula does nothing to ensure the necessary truth of the proposition it expresses, despite the fact that the language satisfies Sher’s criterion.

#### 4.2. *Models and Consistency Proofs*

To show that a set  $\Gamma$  of formulas has a model is to give a relative consistency proof for  $\Gamma$ . This much is uncontroversial. It might well be supposed, however, that this use of models demonstrates a general modal feature of those models, one which suffices to underwrite (I). Thus in particular, one might hope to establish a connection between the use of models in consistency-proofs and some modal feature of model-theoretic truth via the following kind of argument: Since (i) a formula’s truth on a model is equivalent to its consistency, it must be that (ii) a formula’s truth on every model entails its necessity (since  $\varphi$ ’s truth on every model entails that  $\sim \varphi$  is true on no model, in which case  $\sim \varphi$  is inconsistent (by (i)), so  $\varphi$  is necessary). Thus one might conclude that as long as the language

in question is one of those for which model-theoretic consistency-proofs are unproblematic, it must be one in which model-theoretic truth implies necessity.

There are a number of problems with this reasoning. A model-theoretic consistency proof works as follows: Suppose a language  $L$  is associated with a deductive system  $D$  which satisfies the standard soundness constraint:

- (S) For every set  $\Gamma$  of formulas of  $L$ , and every formula  $\varphi$  of  $L$ : If  $\Gamma \vdash \varphi$ , then  $\Gamma \models_L \varphi$

That is, (S) is the condition that if  $\varphi$  is derivable in  $D$  from  $\Gamma$ , then  $\varphi$  is model-theoretically implied, in  $L$ , by  $\Gamma$ . Under this condition, the existence of a model  $m$  of a set  $\Gamma$  of formulas implies that *if* a contradiction is derivable in  $D$  from  $\Gamma$ , then that contradiction is “true on”  $m$  – in which case the “background theory” governing the attribution of truth on a model is itself inconsistent. Assuming the consistency of the background theory, then, a model of  $\Gamma$ , for a system satisfying (S), implies that  $\Gamma$  is *consistent* in the sense that no contradiction is derivable (in  $D$ ) from  $\Gamma$ . We will call this sense of consistency “deductive consistency”.

The first problem with the reasoning of steps (i)–(ii) above is that (i) is typically false: though the languages in question are always ones for which the existence of a model *implies* consistency, the reverse implication is not a general requirement, and often fails. The fact that no contradiction is derivable from  $\Gamma$  does not entail that  $\Gamma$  has a model – unless of course the language and deductive system together meet the much stronger requirement of *completeness*. Thus the first thing to note about this reasoning is that it must be restricted to systems satisfying the completeness theorem.

The second problem is that (ii) makes little sense, at least if “necessity” is interpreted to mean anything like “true across possible worlds”. Deductive consistency and inconsistency apply to formulas, and not to the kinds of things which have truth-values at possible worlds. If  $\sim \varphi$  has no model, then, assuming the existence of a complete deductive system  $D$ , we can conclude that  $\varphi$  is a theorem of  $D$ . But from this we cannot conclude that the propositions  $R(\varphi)$  expressible by  $\varphi$  under acceptable readings of the language are necessary truths, unless we add the further assumption that the deductive system is, as characterized above, *nice*. Once again, the existence of a *nice*, complete deductive system turns out to be crucial in establishing (I).

The only essential requirement on a language and deductive system that are to be used for relative consistency proofs is the soundness theorem (S). And satisfaction of (S) is neither necessary nor sufficient for satisfaction of

(I). The use of models in consistency proofs gives us no reason to suppose that truth across models implies necessary truth in the sense of (I).

#### 4.3. “Possible Models”

A quite different kind of argument for a connection between model-theoretic truth and necessary truth has recently been offered by Vann McGee. The sense of necessity involved here has to do with what might be called “possible models”. Given a language  $L$  with a particular class of models, we can enquire about the models which would have existed, had the world been different. Thus though, for example,  $L$  has no models with unicorns in their domains, it might well have had such models if there had existed any unicorns. The question McGee raises is this: would the model-theoretic truths of  $L$  have been any different had  $L$ 's models included not just the ones it actually has, but additionally models which, though structurally similar to the actual ones, contained such additional “merely possible” objects?<sup>21</sup> In particular, where  $M$  is  $L$ 's class of models, and  $M'$  is the “expanded” class of models which would have existed had the world been richer in inhabitants, we can ask the following question about the wffs of  $L$ :

- (Q) Is it the case that for each wff  $\varphi$  of  $L$ , if  $\varphi$  is true on every member of  $M$ , then  $\varphi$  would have been true on every member of the expanded  $M'$ ?

The answer, as McGee demonstrates, is “yes”, assuming that  $L$  and  $M$  meet some fairly standard constraints.<sup>22</sup> For: Suppose  $\varphi$  would have been false on some member  $m'$  of  $M'$ . Then, given the set-theoretic complexity of the “actual” models, there is a model  $m$  in  $M$  that is isomorphic to  $m'$ . And because isomorphic models satisfy the same formulas,  $\varphi$  is false on  $m$  as well.<sup>23</sup>

If we use the term “possible models” for the members of  $M'$ , then this result can be expressed as the claim that if  $\varphi$  is true on every model, then  $\varphi$  is true on every possible model. If, further, one uses the phrase “formula  $\varphi$  is necessary” to mean “formula  $\varphi$  is true on every possible model”, then the result is expressible, and indeed sometimes expressed, as the claim that all model-theoretic truths are necessary. It is important to note that despite the similarity in terminology, this claim has nothing to do with (I): that model-theoretic truths remain model-theoretically true when we expand the class of models as described above (or in any other way, for that matter) does not tell us anything about the propositions expressed by those model-theoretic truths. “Necessity”, in the sense just introduced, has nothing to do with the truth of a proposition across possible worlds.

4.4. *Cardinality*

Formulas which express (under acceptable readings) claims about the cardinality of the universe offer a particularly interesting class of counterexamples to (I) and to (II). Consider for example a language  $L$  all of whose models are finite, and which contains the sentence

$$(INF) \quad (\forall x)(\forall y)(fx = fy \rightarrow x = y) \& (\exists x)(\forall y)(fy \neq x).$$

Though INF expresses true propositions under various acceptable readings, it is false on every model of  $L$ . Thus its negation,  $\sim INF$ , is a model-theoretic truth of  $L$  which expresses falsehoods under acceptable readings. And as we have seen, sentences of ordinary second-order languages exhibit the same phenomenon: in a language whose models are all smaller than the least inaccessible cardinal, the sentence  $\sim ZFC2$  is true on every model, despite the fact that the proposition it expresses under its usual intended reading is presumably false. Similarly for larger cardinals: as Stewart Shapiro has pointed out, there are sentences that are satisfiable only in domains of Mahlo cardinality, sentences that are satisfiable only in domains of greater than measurable cardinality, and so on.<sup>24</sup> The negations of such sentences will be model-theoretic truths of any language whose models are all smaller than the relevant cardinals. But of course it does not follow from the model-theoretic truth of these negations that the propositions they express, under intended readings, are in fact true. In those cases in which there exists a set of sufficiently large cardinality, the proposition expressed by the negation in question (i.e., by  $\sim INF$ ,  $\sim ZFC2$ , etc.) will be false.

This suggests that a way to guard against such counterexamples is simply to “rule out” any languages in which  $\sim INF$ ,  $\sim ZFC2$ , etc., are model-theoretic truths, and to do this by paying attention only to languages with “large enough” models. As long as we restrict our attention to languages some of whose models are infinite, we will have restricted ourselves to languages in which  $\sim INF$  fails to be a model-theoretic truth, as desired. As long as we restrict ourselves to languages some of whose models are of an inaccessible cardinality, we will encounter only languages in which  $\sim ZFC2$  fails, as desired, to be a model-theoretic truth. And so on. We overcome potential counterexamples to (II) by, first, deciding which of the cardinality-sentences “ought” to be falsified by some model, and then restricting our attention to languages which include the relevant models – i.e., whose “metatheory” includes the relevant cardinality-axioms.<sup>25</sup>

Such a strategy has recently been proposed by Stewart Shapiro. As Shapiro puts it,

As far as cardinality goes, an axiom of infinity is enough for first-order languages. The downward Löwenheim–Skolem theorem entails that if a (countable) argument has a counter-model at all, it has one in a model whose domain is at most countable. The situation with second-order logic is more complex. There are sentences, with no nonlogical terminology, that are satisfiable only in domains of Mahlo cardinality, domains greater than measurable cardinality, etc. So, if these sentences are not to be declared logically false, the metatheory must have some rather strong axioms of infinity (see Shapiro 1991, chap. 6) - unless the possibility of such large domains can be ruled out on logical grounds. . . .

So far, the only constraints on the metatheory we have considered concern the cardinalities of the models. Is this the only type of constraint we need to consider in order to avoid substantive generalizations? With standard first-order or higher-order languages, the answer is “yes”. This is a consequence of what may be called the *isomorphism property* of the formalism: If two models  $M$ ,  $M'$  are isomorphic vis-à-vis the nonlogical items in a formula  $\Phi$ , then  $M$  satisfies  $\Phi$  if and only if  $M'$  satisfies  $\Phi$ .<sup>26</sup>

The connection with modality is as follows:

If the isomorphism property holds, then in evaluating sentences and arguments, the only ‘possibility’ we need to ‘vary’ is the size of the universe. If enough sizes are represented in the universe of models, then the modal nature of logical consequence will be registered.<sup>27</sup>

The strategy, then, is to rule out the problematic model-theoretic truths just discussed by dealing only with languages which have sufficiently-large models. This, together with the isomorphism property, is intended to ensure that model-theoretic consequences, and a fortiori model-theoretic truths, are necessary.

There are three things to note about this strategy. The first is that this method of avoiding the cardinality-counterexamples can succeed only if there are in fact sets of appropriately large cardinality. And we have no guarantee that there are in fact such sets. Consider for example the question of the existence of an inaccessible cardinal. If there is no such cardinal, then no language has models of “sufficiently large” cardinality, and  $\sim\text{ZFC2}$  will be a model-theoretic truth which expresses a falsehood, no matter how carefully we choose our class of models. Similarly for the larger cardinals. The interesting point here is that the positive cardinality-sentences in question (e.g., INF, ZFC2, etc.) will express truths under their ordinary intended readings as long as the set-theoretic *universe as a whole* has certain characteristics (e.g., having infinitely many members, satisfying the ZFC axioms, etc.). But these characteristics of the set-theoretic universe do not guarantee the existence of a *set* with the relevant characteristics, i.e., of a model on which the cardinality-sentence in question is true.

The second point to make about the strategy is that it is not a defense of (I), i.e., of a general connection between model-theoretic truth and logical truth for any non-circularly defined class of languages. It is

rather the claim that, once we have determined which sentences we would like to count as model-theoretic truths, we can choose languages which are guaranteed to have this model-theoretic output; and that we can do this simply by choosing languages whose models are sufficiently large. This is an extremely interesting claim. That simple cardinality constraints (together with the isomorphism property) can guarantee a connection between model-theoretic truth and necessary truth is a highly non-trivial claim, and if true promises to give us a straightforward way of constructing languages whose model-theoretic truths do indeed express necessary truths. Indeed, Shapiro's central claim appears to be not that there is an essentially "modal" characteristic of model-theoretic truth, but rather that, in appropriately-chosen languages, judicious selection of model-theoretic apparatus can bring about a connection between the model-theoretic truths of that language and the necessary truths expressible in it.<sup>28</sup>

But the third point to make about this strategy is that the counterexamples to (I) that it rules out are only those having to do explicitly with cardinality. Formulas not directly asserting the cardinality of the universe (under acceptable readings) remain as problematic as before. In the first-order case, for example, the addition of the quantifier " $\exists^{\text{AI}}x$ " still gives us counterexamples to (I), no matter how large our models are. Similarly for formulas of the form " $\exists x(x = a)$ ". In the second-order case, such formulas as " $\exists X\forall y(Xy \leftrightarrow y = y)$ ", " $\exists X\forall y(Xy \leftrightarrow y \notin y)$ ", and the like will pose counterexamples to (I) on various choices of acceptable readings for these formulas (see Section 2.2). And these formulas remain model-theoretically true independently of the cardinalities of models.

Moreover, the general strategy of dealing only with languages whose model-theoretic apparatus has been designed specifically to give the pre-determined, "correct" results threatens to become viciously circular if applied as a means of avoiding the kinds of counterexamples just noted. Given any collection of formulas whatsoever, it is possible to design a model-theoretic apparatus and class of readings for these formulas which together satisfy (I). The important question, if one is interested in the modal status of model-theoretic truth, however, is whether there is anything about model-theoretic truth itself, in some ordinary or canonical or natural range of languages (not just in those designed ad hoc to get the right result), which underwrites a general connection between model-theoretic truth and necessity. And of such a connection there seems, as yet, to be no evidence.

#### 4.5. *A Wealth of Evidence?*

A fundamental intuition underlying the use of model-theoretic truth to assess necessity seems to be the following: If  $\varphi$  expresses something that *could be* false, then  $\varphi$  itself must *be* false on some model. As we have seen, however, this is not in general true. It *is* true for any language with a nice deductive system and a completeness theorem, since here we can reason as follows: If  $\varphi$  expresses something that *could be* false, then  $\varphi$  is not a theorem, and, by completeness,  $\varphi$  is false on some model. It is also true for a variety of other specialized languages, some of which are noted above. But it is by no means a universal feature of formal languages.

Nevertheless, there is a pervasive intuition that (I) is simply and obviously true, especially when couched in its vague formulation (\*). Part of the explanation of this is, I think, that there is a good deal of anecdotal evidence for (I). Pick “at random”, as it were, a formula which is true on every model, and odds are you will have picked a formula which expresses a necessary truth. But this anecdotal evidence is due to the fact that the “random” model-theoretic truths we are likely to pick are first-order. The beautiful organization and balance of first-order logic, in which all model-theoretic truths are theorems, and (virtually) all theorems necessarily true (under every acceptable reading) leaves us with the easy impression that model-theoretic truth implies necessary truth. But once we recognize that it is the carefully-engineered balance of first-order logic, and not any modal character of model-theoretic truth which is responsible, this impression ought to fade.

A second source of intuitive support for (I) comes from the mathematical contexts in which we frequently employ formal languages. Consider for example the languages of set theory, number theory, algebra, etc., whose formulas express (under acceptable readings) only necessary truths or necessary falsehoods. In these cases, satisfaction of (II) implies satisfaction of (I): if every model-theoretic truth expresses only *truths*, then every model-theoretic truth will express only *necessary* truths, for the simple reason that the language expresses no contingent truths. If, further, as in the typical case with, e.g., languages of set theory and of number theory, there is only one acceptable reading (the “intended” reading), the guarantee of (II) itself requires very little: instead of the quite strong (MOD), all that’s required is the rather weak condition that the language have a model corresponding to the intended reading. The truth of a formula  $\varphi$  on this “intended” model implies the truth of the proposition expressed by  $\varphi$  under the intended reading (via (T)); and the truth of this proposition entails its necessary truth, simply because all true propositions expressed by formulas of such a language are necessarily true. In short, a mathematical

language with a single acceptable reading will satisfy (I) as long as it has a model corresponding with the intended reading.

As we have seen, it is not always a simple matter – and indeed perhaps it is not always possible – to provide one’s language with a model which corresponds with its intended reading. But in those mathematical cases in which there is such a model, the necessary truth of the propositions expressed by model-theoretic truths is straightforwardly guaranteed. This of course has nothing to do with a connection between model-theoretic truth and necessity: the only model which plays any role whatsoever here is the single “intended” model, and the necessity of the propositions expressed is due entirely to the subject-matter of the intended reading. But the straightforwardness of this connection, together with the ubiquity of mathematical languages of this type, doubtless helps to reinforce the pervasive impression that truth across models is in some sense an indicator of necessary truth.

## 5. CONCLUSION

Models and possible worlds are extremely different kinds of things. Models typically provide “interpretations” of the non-logical vocabulary of the language, and assign truth-values accordingly: a formula is true on a model iff that formula is in fact true when interpreted in the way given by the model. To say that a formula is true on every model is to say that under each of these different interpretations, the formula says something which is in fact true. Possible worlds, on the other hand, have nothing to do with alternative interpretations. To assess a formula’s truth-value at a possible world, we must take that formula already to have a determinate interpretation, or reading. A fully-interpreted formula (i.e., one paired with a reading) is true at every possible world iff the claim actually made by that formula, under that reading, is true and would have remained true no matter how the world turned out.

We can put the difference between truth across models and truth across possible worlds (somewhat crudely) as follows:  $\varphi$ ’s truth on every model is a matter of the actual truth of a range of propositions;  $\varphi$ ’s truth at every possible world is a matter of the necessary truth of a given proposition. As we have seen, there are important cases in which these two converge: cases in which if a formula  $\varphi$  expresses (under some acceptable reading of the language) something which is *not* a necessary truth, then  $\varphi$  also expresses (under the assignment made by some model) something which is actually false. That is, there are cases in which (I) holds. But the differences between the way in which models assign truth-values and the way in which

possible worlds do so should make it unsurprising that this convergence is not a universal feature of formal languages, not even of fairly typical ones.

We are now in a position to make a few general comments about the conditions under which (I) and (II) do hold. A fairly natural constraint to impose on a language is that its models serve as precise, tractable versions of its acceptable readings – i.e., as “corresponding to” those readings in the sense discussed above, and as assigning truth-values accordingly. This constraint is met by just those languages satisfying (MOD) and (T). As we have seen, satisfaction of (MOD) and (T) is sufficient for satisfaction of (II).

A variety of stronger conditions imply satisfaction of (I). One such condition, as we have seen, is the existence of a nice, complete deductive system. There are also “trivial” ways of satisfying (I): if  $L$  has no model-theoretic truths, then it satisfies (I); if each of  $L$ 's readings assigns to every formula a necessary truth, then it satisfies (I), and so on. More interesting are those cases in which the readings all assign either necessary truths or necessary falsehoods to each formula, as for example is the case with languages used for purely mathematical purposes. In these cases, the necessity of the propositions expressed by the model-theoretic truths is due not to the model theory, but to the subject-matter of the readings.

In short, there are at least two interesting, and a variety of uninteresting, constraints one might impose on a language  $L$  which, when met, ensure that  $L$ 's model-theoretic truths express necessary truths. There remains a great deal to be said about what kinds of constraints might be useful and important to impose on the trio of formulas, acceptable readings, and model-theoretic apparatus that make up a formal language, and about which of these together suffice to guarantee (I) or the weaker (II). What seems clear, however, is that the connection between model-theoretic truth and necessary truth, when it obtains, will turn on the details of the connections between formulas, models, and readings, and not on any general modal character of model-theoretic truth.

It is important to note that this last point has no bearing on the logical and mathematical purposes for which models, and model-theoretic relations, were originally introduced. Its importance concerns entirely the philosophical significance of those applications. The use of models in demonstrating (relative) consistency and independence results, and the use of model-theoretic techniques in characterizing and investigating similarities across abstract mathematical structures, are independent of the issues raised here concerning models and modality.<sup>29</sup> The use of truth across models to indicate necessary truth is a philosophers' invention. As we have seen, model-theoretic truth *is* a good indicator of necessary truth

in particular cases, most importantly in the case of standard first-order logic. But it is important not to mistake this important feature of first-order model-theoretic truth for a general characteristic of model-theoretic truth.<sup>30</sup>

## NOTES

<sup>1</sup> This point is easy to miss. Given a series of symbols which constitutes a well-formed formula, it might appear that that series, together with a chosen model, might be the kind of thing that is necessarily true. Consider for instance the sentence “ $Fa$ ” and a model which assigns to “ $a$ ” the number two and to “ $F$ ” the set of even integers. The sentence-on-that-model seems a good candidate for necessary truth, since two is, presumably, necessarily an even integer. But the association of this sentence-model pair with the proposition *two is an even integer* is misleading. Note that the sentence-model pair is equally accurately associated with the (presumably contingent) *Jones is thinking of two*, if Jones happens to be thinking just of the even integers. Because the kinds of “meanings” assigned by models are purely extensional, they do not associate unique propositions with formulas.

<sup>2</sup> This presupposes that  $R$  assigns a range to the first-order variables, and assigns individuals, functions and relations to the individual constants, function-terms and relation-terms respectively. Readings which do not do so (e.g., readings which simply assign propositions directly to wffs) will not be said to correspond to models.

<sup>3</sup> This constraint of course does not apply to all those things which are commonly called languages; it does not apply, for example, to languages whose model-theoretic apparatus includes Boolean-valued models. But it does apply to all those languages which have a coherent notion of model-theoretic truth, and for which the question of the truth of (I) can reasonably arise.

<sup>4</sup> Suppose  $\varphi$  is true on every set-reading. Then since every model corresponds with some set-reading,  $\varphi$  is true on every model. By completeness,  $\varphi$  is a standard first-order theorem, and all standard first-order theorems have only true readings. QED. Kreisel (1967, 152–5) has given a similar demonstration, also appealing to completeness, of the equivalence between a first-order formula’s truth on every “set-theoretic” model and its truth on every (possibly proper-class-sized) model. (X) can be obtained from Kreisel’s result by assuming (T), and by supposing (a) that every set-sized model corresponds with some acceptable reading, and (b) that the language can be expanded in such a way that for every reading there is a corresponding (possibly proper-class-sized) model. What Kreisel calls “intuitive validity” is model-theoretic truth in such an expanded language. See also Section 3.2 below.

<sup>5</sup> There are absolutely infinitely many  $\alpha$ ’s if there are “too many”  $\alpha$ ’s to form a set. See Vann McGee (1992a). For the origin of the quantifier, see Cantor (1899).

<sup>6</sup> The second-order axioms are those obtained from the first-order by replacing axiom-schemes (replacement, separation) by their universally-quantified second-order versions. See Shapiro (1991), esp. chaps. 4 and 5.

<sup>7</sup> See McGee op. cit. pp. 273, 292.

<sup>8</sup> See Henkin (1950). The usual “full” models of second-order logic are those in which the second-order variables of degree  $n$  range over all sets of ordered  $n$ -tuples of the model’s domain. In a Henkin model, the second-order variables need not contain within their range all such sets of ordered  $n$ -tuples; thus many formulas true on every full model are falsified

by a Henkin model. There is a complete, recursive axiomization for second-order logic with Henkin models (op. cit.).

<sup>9</sup> See Boolos (1975).

<sup>10</sup> See Boolos (1984) and (1985).

<sup>11</sup> McGee, op. cit. p. 273. The second occurrence of “true” here expresses what we have called “true under its intended reading”.

<sup>12</sup> See Cartwright (1987). Cartwright’s proof actually demonstrates not just (I) but the stronger result that every model-theoretically valid argument is necessarily truth-preserving.

<sup>13</sup> See Etchemendy (1990, esp. pp. 144–155).

<sup>14</sup> For example: Begin with an ordinary first-order language  $L$ . Let  $m$  be one of  $L$ ’s models and  $R$  a reading which corresponds with  $m$ , and on which some formula  $\varphi$  of  $L$  expresses a contingent truth. Let  $L^*$  be the language which is syntactically just like  $L$ , and which has  $m$  as its only model and  $R$  as its only reading.  $L^*$ , then, satisfies (MOD) [and (T)] but not (I).

<sup>15</sup> Sher (1991, 44–48) and Sher (1996).

<sup>16</sup> Sher (1991, 47).

<sup>17</sup> This requires additionally that none of the formulas of the form  $Pt_1 \dots t_n$  is a model-theoretic falsehood.

<sup>18</sup> That is, as long as the models of standard first-order logic meet Sher’s criterion, as Sher takes it they do, then the models of these languages do as well; in most of the cases discussed above, these are the same classes of models.

<sup>19</sup> A formula  $\varphi$  is of the same *syntactic form* as a formula  $\psi$  iff  $\psi$  is obtainable from  $\varphi$  by a function mapping the logical constants to themselves, and mapping non-logical primitive terms to non-logical primitive terms of the same syntactic category. (Thus which formulae count as “of the same syntactic form” as one another will turn in part on the choice of logical constants.) If  $R(\psi)$  is false, then  $\psi$  is false on that model  $m$  with which  $R$  corresponds, and because of the combinatorial richness of the class of models, there is a model  $m'$  which assigns to the non-logical constants of  $\varphi$  what  $m$  assigns to their counterparts in  $\psi$ , so that  $\varphi$  is false on  $m'$ .

<sup>20</sup> Sher (1991, 48).

<sup>21</sup> This is not a perfectly clear specification: we would need to say more about what counts as “structurally similar”, particularly if there are more “merely possible” objects than there are actual objects. But nothing will turn on this lack of specificity.

<sup>22</sup> Vann McGee, op. cit. and Vann McGee (1992b). Incidentally, though McGee characterizes Etchemendy’s concern, in *The Concept of Logical Consequence*, as being with the question (Q), this is a mistake. In asking whether model-theoretic truths are generally necessary, Etchemendy has no concern with “possible models”; his question is whether model-theoretic truths express, under their intended readings, claims which are necessarily true. See Etchemendy (1990, chaps. 6 and 7). That is, taking the language to have a single (“intended”) reading Int, Etchemendy’s concern, couched in our terminology, is a concern with (I’):

(I’) For every wff  $\varphi$  of  $L$ : If  $\varphi$  is true on every model, then Int( $\varphi$ ) is true at every possible world.

Each of the problems with (I) discussed in this paper is equally a problem for (I’); and of course cases in which (I) holds are a fortiori cases in which (I’) holds. Thus, contra

McGee, the very points McGee makes (op. cit.) concerning the “reliability problem” go towards showing that Etchemendy’s central claim is true, not false.

<sup>23</sup> The important “standard constraints” appealed to here are that the domains of  $M$ ’s models include all the usual sets, and that “truth-on” be defined in the usual way, so that, in particular, isomorphic models satisfy the same formulas.

<sup>24</sup> See Shapiro (1991, chap. 6) and also Shapiro (1998, 151).

<sup>25</sup> Here, the “metatheory” of a language  $L$  is the set theory governing  $L$ ’s collection of models.

<sup>26</sup> Ibid., 151.

<sup>27</sup> Ibid., 152.

<sup>28</sup> As Shapiro puts it, in terms of the more general notion of model-theoretic consequence: “My claim is that model-theoretic consequence *can be made into* a good mathematical model of” the modal relation of logical consequence (emphasis added). Ibid., 148.

<sup>29</sup> The claim that model-theoretic consistency (and a fortiori independence) results indeed have no “modal” flavor of the kind investigated here is argued at greater length in my Blanchette (1996).

<sup>30</sup> Versions of this paper have been presented to the 1997 meeting of the Society for Exact Philosophy, the Indiana University-Bloomington Logic Group, the University of Notre Dame Logic Group, and the University of Notre Dame Philosophy Department; many thanks to the members of these audiences for helpful comments. Thanks to Marian David, Mic Detlefsen, Charlie Donahue, Martin Jones, Vann McGee, and Alvin Plantinga; and particularly to Stewart Shapiro for detailed comments. It is also a pleasure to acknowledge the influence of the work of John Etchemendy, and to thank him as well for comments and criticism.

#### REFERENCES

- Blanchette, P.: 1996, ‘Frege and Hilbert on Consistency’, *Journal of Philosophy* **XCIII**, 317–36.
- Boolos, G.: 1975, ‘On Second-Order Logic’, *Journal of Philosophy* **LXXII**, 509–27
- Boolos, G.: 1984, ‘To Be is to Be a Value of a Variable (Or to be Some Values of Some Variables)’, *The Journal of Philosophy* **LXXXI**, 430–49.
- Boolos, G.: 1985, ‘Nominalist Platonism’, *The Philosophical Review* **XCIV**(3), 327–44
- Cantor: 1899, ‘Letter to Dedekind, trans by S. Bauer-Mengelberg’, in J. van Heijenoort (ed.), *From Frege to Gödel*, Harvard University Press, Cambridge, MA (1967), pp. 113–17.
- Cartwright, R.: 1987, ‘Implications and Entailments’, in R. Cartwright (ed.), *Philosophical Essays*, MIT Press, Cambridge.
- Etchemendy, J.: 1990, *The Concept of Logical Consequence*, Harvard University Press, Cambridge.
- Henkin, L.: 1950, ‘Completeness in the Theory of Types’, *Journal of Symbolic Logic* **15**, 81–91.
- Kreisel, G.: 1967, ‘Informal Rigour and Completeness Proofs’, in Lakatos (ed.), *Problems in the Philosophy of Mathematics*, North-Holland, Amsterdam, pp. 138–71.
- McGee, V.: 1992a, ‘Two Problems with Tarski’s Theory of Consequence’, *Proceedings of the Aristotelian Society* **92**, 273–92.

- McGee, V.: 1992b, 'Review of Etchemendy (1990)', *The Journal of Symbolic Logic* **57**, 254–55.
- Shapiro, S.: 1991, *Foundations without Foundationalism*, Clarendon Press, Oxford.
- Shapiro, S.: 1998, 'Logical Consequence: Models and Modality', in M. Schirn (ed.), *The Philosophy of Mathematics Today*, Clarendon Press, Oxford.
- Sher, G.: 1991, *The Bounds of Logic*, MIT Press, Cambridge.
- Sher, G.: 1996, "Did Tarski Commit 'Tarski's Fallacy?'" , *Journal of Symbolic Logic* **61**, 653–686.

Department of Philosophy  
University of Notre Dame  
Notre Dame, IN 46556  
E-mail: blanchette.1@nd.edu