Adrian Miroiu

WORLDS WITHIN WORLDS

It is sometimes argued that possible worlds are too rich. Indeed, for every sentence ϕ and every world w, either ϕ is true at w, or $\neg\phi$ is true at w. Possible worlds leave nothing unsettled: for any question one may ask, there is always an answer, yes or no. My view goes along the opposite side. I take possible worlds to be incredibly poor, and I think that a good deal can be added to them. The claim I want to defend is that this can be done by taking into account the extremely large class of world-indexed sentences.

Let ϕ be some sentence. Then a hoard of world-indexed new sentences can be produced. From "Quine is a distinguished philosopher" we get the sentence: "Actually Quine is a distinguished philosopher", and also the sentence: "At world w Quine is a distinguished philosopher", for any w. On the standard account, a sentence is true or false at every possible world. Thus, the sentence "Quine is a distinguished philosopher" is true at a world w if at w it holds that Quine is a distinguished philosopher, and false otherwise. Now, what about the sentence "Actually Quine is a distinguished philosopher"? For any world w, it is true at w if it holds at the actual world that Quine is a distinguished philosopher, and false otherwise.

Next, how can we establish the truth value at world w' of the sentence "At world w Quine is a distinguished philosopher"? The issue we face is how to handle the condition that *at w' it holds that Quine is a distinguished philosopher at w*. The standard move looks to be something like this. Suppose that at w it is indeed the case that Quine is a distinguished philosopher, i.e. at w the truth-value of the sentence "Quine is a distinguished philosopher" is truth. Now, this is a condition that cannot fail to obtain: while it is of course possible to hold that at some world w'' Quine is not a distinguished philosopher, there is no way to deny that he is so at w. Hence at every world w the sen-

Nordic Journal of Philosophical Logic, Vol. 2, No. 1, pp. 25–40.

tence "At w Quine is a distinguished philosopher" should have value truth, i.e. at every w' it should be the case that Quine is a distinguished philosopher at w.[1]

(The argument goes on similar lines with time-indexed sentences. Starting with the sentence "It snows in Ithaca, NY" one gets an eternal sentence like "It snows in Ithaca, NY on January 27, 1996". While of course it is contingent that it actually snowed in Ithaca, NY on January 27, 1996, if it did snow on that day, then for all days after January 27, 1996 it should be the case that it snowed in Ithaca, NY on January 27, 1996.)

This standard approach involves two distinct claims. First, that world-indexing a sentence yields another sentence: starting with any sentence ϕ we get: *at* w, ϕ, and this is a sentence too. Second, that there is a fixed procedure for determining the truth-value of world-indexed sentences. While accepting the former claim, I reject the latter. On my view, there are more logically legitimate ways to determine the truth-values of world-indexed sentences, and they help us grasp a quite new class of modal logics.

I.

In order to get a grip on the logical standing of world-indexed sentences, let us start by considering a modal propositional calculus. Our language $\mathcal{L}$ will contain countably many sentence letters S, S', S'', etc., and the logical symbols $\wedge$, $\vee$, $\rightarrow$, $\neg$, $\equiv$, and $\Box$. In addition, $\mathcal{L}$ will contain countably many world symbols w, w', w'', etc., and let W be the set of all these world symbols. Consequently, the collection SEN($\mathcal{L}$) of the sentences of $\mathcal{L}$ is defined with only a bit more sophistication than usual: SEN($\mathcal{L}$) is the smallest collection such that

(i) every sentence letter is in SEN($\mathcal{L}$);

(ii) if ϕ and ψ are in SEN($\mathcal{L}$), then
$\phi \wedge \psi, \phi \vee \psi, \phi \rightarrow \psi, \phi \equiv \psi$ are in SEN($\mathcal{L}$);

(iii) if ϕ is in SEN($\mathcal{L}$), then $\neg\phi, \Box\phi$ and $w\phi$ are in SEN($\mathcal{L}$).

This definition intends to capture the intuitive requirements that the negation of a sentence, the conjunction, disjunction, implication and equivalence of two sentences are also sentences; that for each sentence ϕ, that ϕ is necessary is a sentence; and again, that world-indexing a

[1]The above argument is analogous to that of A. Plantinga (1976) on world-indexed properties.

sentence yields a sentence. For each sentence ϕ and each world symbol w, we get the sentence $w\phi$, intuitively: ϕ is the case at w.

For the sake of simplicity, I shall assume for the moment that the underlying modal logic is the standard system K. The theorems of K are the tautologies, all expressions of the form:

(1.1) $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$

and all expressions deducible from them by detachment and necessitation: the rule that from $\vdash \phi$ to infer $\vdash \Box\phi$.

The requirements on the behaviour of world-indexed sentences I shall advance below are intuitively motivated. As familiar from standard possible worlds semantics, $\neg\phi$ is the case at a world w if and only if ϕ is not the case at w; $\phi \wedge \psi$ is the case at w if and only if ϕ is the case at w and ψ is the case at w, and so on. Now we can mimic them in the language $\mathcal{L}$, and get:

(1.2) If ϕ is a sentence letter, then $\vdash w\phi \vee w\neg\phi$, for every w

(1.3) $\vdash w\neg\phi \equiv \neg w\phi$, for every w

(1.4) $\vdash w(\phi \wedge \psi) \equiv w\phi \wedge w\psi$, for every w

(1.5) $\vdash w(\phi \vee \psi) \equiv w\phi \vee w\psi$, for every w

(1.6) $\vdash w(\phi \to \psi) \equiv w\phi \to w\psi$, for every w,

etc. To these we need to add:

(1.7) If $\vdash \phi$, then $\vdash w\phi$ for every w.[2]

For reasons that will become apparent in what follows, I shall call the resulting logic LK (local K). A straight consequence of requirements (1.2)–(1.7) is that for each sentence ϕ, either $w\phi$ or $w\neg\phi$. Indeed, since $\phi \vee \neg\phi$ is a tautology, by (1.7) we get $w(\phi \vee \neg\phi)$, and with (1.5) it follows that

---

[2]It is possible to treat w as a sort of necessity operator $\Box_w$. If the following two conditions:

(1.01) $\Box_w(\phi \vee \psi) \to (\Box_w\phi \vee \Box_w\psi)$

(1.02) $\Box_w\phi \to \neg\Box_w\neg\phi$

are added to the axioms of system K (which we assume to hold for $\Box_w$), then expressions (1.2)–(1.7) are all accepted. (Semantically, the two conditions amount to the requirement that each WORLD has exactly one alternative.) It follows that we might view all the logics discussed in this paper simply as logics with many necessity operators like $\Box_w$. However, I do not think that such an account illuminates the intuitions behind the construction of these logics.

(1.8) $\vdash w\phi \vee w\neg\phi$.

The interesting point comes when we take ϕ to be a world-indexed sentence, and in particular a w'-indexed one, for some w'. Observe that for each ϕ, $w'\phi$ is a sentence, and it makes sense to ask if $w'\phi$ is the case at w. According to (1.8), for each $w'\phi$ we have either $ww'\phi$, or $w\neg w'\phi$. Further, by (1.3) we get: either $ww'\phi$, or $ww'\neg\phi$. Let us pause a moment to reflect on this claim. It asserts that for each ϕ either it holds at world w that ϕ is the case at w', or it holds at w that $\neg\phi$ is the case at w'. The world w says how w' is related to every sentence ϕ.

Suppose, for example, that according to w, ϕ is the case at w' if and only if really ϕ is the case at w'; or, to put it more formally: $ww'\phi$ if and only if $w'\phi$. The world w says that something is going on at w' if and only if that something really is going on at w'. Hence w reflects (or mirrors) w' adequately if and only if for each ϕ, $ww'\phi \equiv w'\phi$.

Now suppose that there is some world w'' (different from w') such that for each ϕ w says that ϕ is the case at w' if and only if ϕ is the case at w''; to put it more formally, for each ϕ: $ww'\phi$ if and only if $w''\phi$. I shall say that w reflects or mirrors w' as w'' (and consequently that it does not mirror w' adequately). Also, say that a world is self-conscious if it adequately mirrors itself, i.e. it holds that

(1.9) For each ϕ, $ww\phi$ if and only if $w\phi$
(or, a bit more formal: $w(w\phi \equiv \phi)$, for each ϕ).

Further, the world w is self-conscious in w' when for each ϕ it is the case at w' that $ww\phi$ if and only if $w\phi$; or, in a more formal wording, $w'(ww\phi \equiv w\phi)$, for each ϕ.

In general, there is no guarantee that for any two worlds w and w' there is some (other) world w'' such that w would mirror w' as w''. But the reflection of w' in w is world-like. Let me explain this. Consider the collection $\mathrm{SEN}_w(w')$ of all sentences ϕ such that according to w ϕ is the case at w'. One can easily prove that ϕ is in $\mathrm{SEN}_w(w')$ iff $\neg\phi$ is not in $\mathrm{SEN}_w(w')$; that $\phi \wedge \psi$ is in $\mathrm{SEN}_w(w')$ if and only if ϕ is in $\mathrm{SEN}_w(w')$ and ψ is in $\mathrm{SEN}_w(w')$; that $\phi \vee \psi$ is in $\mathrm{SEN}_w(w')$ if and only if ϕ is in $\mathrm{SEN}_w(w')$ or at least ψ is in $\mathrm{SEN}_w(w')$, etc. Hence $\mathrm{SEN}_w(w')$ mimics a world, although $\mathrm{SEN}_w(w')$ itself is not necessarily a world too; but of course it is possible for some world w'' that ϕ be in $\mathrm{SEN}_w(w')$ if and only if it is the case at w'' that ϕ , for any ϕ.

World-indexed sentences might be even more complex in form. Take, e.g., a sentence like $ww'w''\phi$. It says that according to w the world w' is such that it is the case that ϕ holds at w''; or: w reflects w' as saying that ϕ is the case at w''. Let the following condition hold:

(1.10) For each ϕ, $ww'w''\phi$ if and only if $ww''\phi$
(or, a bit more formal: $w(w'w'' \equiv w'')$ for each ϕ).

Then inside w the world w' adequately mirrors w''; note that w' does not mirror w'' as it 'really' is: rather the image or reflection of w' in w mirrors (in an adequate manner) the reflection of the image of w'' in w. And if for each ϕ it holds that $ww'w''\phi$ if and only if $ww'''\phi$ for some world w''', we might say that inside w the world w' mirrors w'' as w'''. The reflection relations just defined are different from the ones mentioned above; as opposed to 'outer' reflections, they express 'inner' reflections. In the former case we had: w reflects w'; in the latter one, the reflection relation comes relativized: in w, the world w' reflects w''. In the former case we had: w reflects w' as w''; in the latter one we get: in w, the world w' reflects w'' as w'''.

The question is, how very rich is a world w? As I shall show below, it is possible to provide a semantical frame that makes sense of a world's w being able to mirror other worlds as well as itself, and articulate inside itself a picture of all their relationships. But if that is the case, the other worlds appear redundant, with their ethereal counterparts in w as adequate substitutes. Whatever a world w' can say should be sayable within w. It is in this sense that the modal systems to be discussed below are local: they attempt to localize, internalize in any world w all the claims of all worlds.

However, since LK is modal, its semantics involves a new collection of worlds k, k', k'', etc. How are they related to our old worlds w, w', w''? As I shall try to show, under certain assumptions, our old worlds prove able to reflect what is going on at the new ones.

II.

Let Σ be a subset of SEN($\mathcal{L}$). I shall write $\Sigma \vdash_{LK} \phi$ for: ϕ is deducible in LK from Σ. A model for $\mathcal{L}$ is a structure $\mathfrak{C} = \langle K, R, F, \mho \rangle$, where K is a set of indices, R is a binary relation on K, F is a function from $W \times K$ to K, and $\mho$ is a function from SEN($\mathcal{L}$) $\times K$ to the set $\{1, 0\}$ of truth-values. In other words, $\mho$ assigns each sentence, relative to each world k in K, a truth-value. The definition of $\mho$ is the standard one, surely with a new case for sentences of the form $w\phi$, with w in W:

(D1) The definition of $\mho$:

(a) if ϕ is a sentence letter, then $\mho(\phi, k) = 1$ or $\mho(\phi, k) = 0$;
(b) if ϕ is $\neg\psi$, then $\mho(\phi, k) = 1$ iff $\mho(\psi, k) = 0$;
(c) if ϕ is $\psi \vee \chi$, then $\mho(\phi, k) = 1$ iff $\mho(\psi, k) = 1$ or $\mho(\chi, k) = 1$;
(d) if ϕ is $\psi \wedge \chi$, then $\mho(\phi, k) = 1$ iff $\mho(\psi, k) = \mho(\chi, k) = 1$;
(e) if ϕ is $\psi \rightarrow \chi$, then $\mho(\phi, k) = 1$ iff $\mho(\psi, k) = 0$ or $\mho(\chi, k)) = 1$;
(f) if ϕ is $\psi \equiv \chi$, then $\mho(\phi, k) = 1$ iff $\mho(\psi, k) = \mho(\chi, k))$;
(g) if ϕ is $\Box\psi$, then $\mho(\phi, k) = 1$ iff $\mho(\psi, k') = 1$ for all k' such that $R(k, k')$;
(h) if ϕ is $w\psi$, then $\mho(\phi, k) = 1$ iff $\mho(\psi, F(w, k)) = 1$.

I shall say that a sentence ϕ is true in a model $\mathfrak{C} = \langle K, R, F, \mho\rangle$, and write $\mathfrak{C} \models \phi$, iff $\mho(\phi, k) = 1$ for all k in K. A sentence ϕ is valid iff ϕ is true in all models, and I shall write $\models \phi$ in this case. Models for $\mathcal{L}$ differ from usual models in possible worlds semantics in that they contain the function F. The intuition is that the worlds we use in $\mathcal{L}$ mimic the indices in K: whenever at k it is true that ϕ is the case at w, then at the element k' of K corresponding by F to w (with respect to k) the sentence ϕ must be true. It follows that w is in k a mirror of $k' = F(w, k)$; k' is reflected in k as w. Now in general $F(w, k)$ varies with k. In different worlds k a world k' is not reflected as some fixed world w. Hence, in different k's the collection of sentences ϕ such that ϕ is the case at the world w do not necessarily coincide. Note also that in the usual parlour the indices in K are called "worlds"; but they are of course different from our old worlds w, w', w'', etc. Although one of my main aims in this paper is to closely connect "worlds" like w, w', w'', etc. and "worlds" like k, k', k'', etc., at the present stage of investigation I shall carefully distinguish them and call "WORLDS" the elements of K, while the members of W will still be appealed to as "worlds".

The first general result to be proved is this.

(2.1) $\vdash_{LK} \phi$ iff $\models_{LK} \phi$.

(Whenever there is no danger of confusion, I shall omit subscripts.) Sufficiency is the easiest to prove. It follows as a straight consequence of the following results:

(2.1a) If ϕ is a tautology, then $\models \phi$.

(2.1b) $\models \Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$.

(2.1c) If $\models \phi$, then $\models \Box\phi$.

(2.1d) $\models \neg w\phi \equiv w\neg\phi$

(2.1e) $\models w(\phi \wedge \psi) \equiv w\phi \wedge w\psi$

(2.1f) $\models w(\phi \vee \psi) \equiv w\phi \vee w\psi$

(2.1g) $\models w(\phi \rightarrow \psi) \equiv w\phi \rightarrow w\psi$

(2.1h) $\models w(\phi \equiv \psi) \equiv (w\phi \equiv w\psi)$

(2.1i) $\models w\phi \vee w\neg\phi$

(2.1j) $\models \neg(w\phi \wedge w\neg\phi)$

(2.1k) If $\models \phi$, then $\models w\phi$ for each ϕ.

For example, the proof of (2.1d) runs like this: let $\mathfrak{C}$ be a model and let k be a WORLD. Then: $\mho(\neg w\phi, k) = 1$ iff $\mho(w\phi, k) = 0$ iff $\mho(\phi, F(w, k)) = 0$ iff $\mho(\neg\phi, F(w, k)) = 1$ iff $\mho(w\neg\phi, k) = 1$. The proof of (2.1e) goes along the following lines: $\mho(w(\phi \wedge \psi), k) = 1$ iff $\mho(\phi \wedge \psi, F(w, k)) = 1$ iff $\mho(\phi, F(w, k)) = 1$ and $\mho(\psi, F(w, k)) = 1$, iff $\mho(w\phi, k) = 1$ and $\mho(w\psi, k) = 1$, iff $\mho(w\phi \wedge w\psi, k) = 1$. To prove (2.1k), suppose that $\mho(\phi, k) = 1$ for all k, but $\mho(w\phi, k') = 0$ for some k'. Now, $\mho(w\phi, k') = 0$ iff $\mho(\phi, F(w, k')) = 0$. But, since F is a function, $F(w, k')$ is a WORLD k'', of which we supposed that $\mho(\phi, k'') = 1$.

Necessity needs two lemmas:

(2.2) Each LK-consistent set Σ of sentences can be extended to a maximal LK-consistent set of sentences.

(2.3) (The local maximality lemma) Let Σ be a LK-maximal consistent set of sentences, and let $LOC_w(\Sigma)$ be the set of all sentences ϕ such that $w\phi$ is in Σ. Then $LOC_w(\Sigma)$ is a maximal LK-consistent set of sentences.

Proof. $LOC_w(\Sigma)$ is LK-consistent. For suppose it were inconsistent. Then for some ψ we would have both $\psi \in LOC_w(\Sigma)$ and $\neg\psi \in LOC_w(\Sigma)$. But, according to the definition of $LOC_w(\Sigma)$, we would have both $w\psi \in \Sigma$ and $w\neg\psi \in \Sigma$. By (1.3), we would also have $\neg w\psi \in \Sigma$, and this would contradict Σ's being LK-maximal consistent. Next, $LOC_w(\Sigma)$ is LK-maximal. Suppose it were not so. Then for some ψ both $\psi \in LOC_w(\Sigma)$ and $\neg\psi \in LOC_w(\Sigma)$ would not obtain. However, since (1.8) is a LK-theorem, $w\psi \vee w\neg\psi \in \Sigma$ and hence either $w\psi \in \Sigma$ or $w\neg\psi \in \Sigma$. But then we would have either $\psi \in LOC_w(\Sigma)$ or $\neg\psi \in LOC_w(\Sigma)$, which contradicts our assumption.

To prove the necessity part of (2.1), suppose that some ϕ is not LK-provable (i.e., $\vdash_{LK} \phi$ does not hold) and let us show that there is some model $\mathfrak{C}$ such that $\mathfrak{C} \models \phi$ does not hold. Now this is the case if for some WORLD k, $\mho(\phi, k) = 0$. So, let ϕ be not LK-provable. Then the set $\{\neg\phi\}$ is LK-consistent. Hence it can be extended to a LK-maximal consistent set Σ. We shall construct the model $\mathfrak{C}$. Let K be the set of all LK-maximal consistent sets. Obviously, $\Sigma \in K$. $R(\Sigma', \Sigma'')$ holds iff the set of all sentences ϕ such that $\Box\phi$ is in Σ' is included in Σ''. Further, let $\mathcal{F}(w, \Sigma') = LOC_w(\Sigma')$. By the above lemma, $LOC_w(\Sigma')$ is in K (intuitively, w is taken to be the reflection of $LOC_w(\Sigma')$ in Σ'). Finally, let $\mho(\phi, \Sigma') = 1$ iff $\phi \in \Sigma'$ whenever ϕ is a sentence letter. The proof consists in showing that for every sentence ϕ, $\phi \in \Sigma'$ iff $\mho(\phi, \Sigma') = 1$. The only difficult cases are for $\phi = \Box\psi$ and $\phi = w\psi$, for some w.

(i) ϕ has the form $w\psi$. Then:
$\mho(w\psi, \Sigma') = 1$ iff $\mho(\psi, \mathcal{F}(w, \Sigma')) = 1$
(1) iff $\psi \in \mathcal{F}(\psi, \Sigma')$
(2) iff $w\psi \in \Sigma'$.

First, (1) entails (2). Suppose that $w\psi$ does not belong to Σ'. According to the definition of $\mathcal{F}(w, \Sigma')$ and the local maximality lemma, ψ does not belong to $\mathcal{F}(w, \Sigma')$—contradiction. The converse follows by a simple application of the definition of $\mathcal{F}$.

(ii) ϕ has the form $\Box\psi$. Then:
$\mho(\Box\psi, \Sigma') = 1$ iff for each Σ'', if $R(\Sigma', \Sigma'')$, then $\mho(\psi, \Sigma'') = 1$
(1′) iff for each Σ'', if $R(\Sigma', \Sigma'')$, then $\psi \in \Sigma''$
(2′) iff $\Box\psi \in \Sigma'$.

The difficult step is to show that (1′) entails (2′). We shall show that if $\Box\psi$ is not in Σ', then there is some LK-maximal consistent set Σ'' such that $R(\Sigma', \Sigma'')$ and ψ is not in Σ''. The set $\Delta = \{\chi : \Box\chi \in \Sigma'\} \cup \{\neg\psi\}$ is LK-consistent. It can then be extended to a LK-consistent maximal set Δ'. One can easily see that for all χ such that $\Box\chi \in \Sigma'$, $\chi \in \Delta'$, and hence $R(\Sigma', \Delta')$. But, since Δ' contains $\neg\psi$ and is consistent, it is not the case that $\psi \in \Delta'$—contradiction. Now we can show that ϕ is not true in $\mathfrak{C}$. Indeed, since Σ is in K, and $\neg\phi \in \Sigma$, we have $\mho(\phi, \Sigma) = 0$.

A second example of completeness theorem concerns the system TRIV obtained by adding to LK all expressions of the form:

(2.4) $\phi \equiv w\phi$, for each world w

Let us observe that this makes LK collapse into K: a sentence ϕ is true at a WORLD k iff $w\phi$ is true at k. Now, a sentence is TRIV-valid iff it is true in every model $\mathfrak{C}$ such that the following holds:

(2.5) $\mathcal{F}(w, k) = k$ for every w and k

By (2.5), each world is the reflection in each WORLD of that very WORLD. The completeness theorem for TRIV says that a sentence ϕ is TRIV-deducible iff it is TRIV-valid. More rigorously, we have:

(2.6) $\vdash_{TRIV} \phi$ iff $\models_{TRIV} \phi$

Sufficiency is obvious: $\mho(w\phi \equiv \phi, k) = 1$ is an immediate consequence of $\mho(w\phi, k) = \mho(\phi, \mathcal{F}(w,k)) = \mho(\phi, k)$. Conversely, suppose that $\mathfrak{C}$ is a model of TRIV built as in the proof of the necessity part of (2.1). We show that for any world w and any TRIV-maximal consistent set of sentences Σ, $LOC_w(\Sigma) = \Sigma$. Or, to put it in other words, $\phi \in LOC_w(\Sigma)$ iff $w\phi \in \Sigma$. But by (2.4), $(w\phi \equiv \phi) \in \Sigma$ and thus $w\phi \in \Sigma$ iff $\phi \in \Sigma$, whence $\phi \in LOC_w(\Sigma)$ iff $\phi \in \Sigma$. Given the definition of $\mathcal{F}$, we get: $\mathcal{F}(w, \Sigma) = \Sigma$.

A third example concerns the system LS5R. This is defined by adding to LK all expressions of the form:

(2.7T) $\Box\phi \to \phi$

(2.7B) $\phi \to \Box\Diamond\phi$

(2.7S5) $\Box\phi \to \Box\Box\phi$

(2.7R) $w\phi \to \Box w\phi$

(where $\Diamond$ is $\neg\Box\neg$). The completeness theorem for LS5R states that

(2.8) $\vdash_{LS5R} \phi$ iff ϕ is true in all models such that
(i) R is total;
(ii) $\mathcal{F}$ is rigid: $\mathcal{F}(w,k) = \mathcal{F}(w,k')$ for every w, k, k'.

The rigidity condition states that in any WORLD each world reflects the very same WORLD; the way worlds reflect WORLDS is not context-sensitive.

To prove sufficiency, the only difficult part is with (2.7R). Suppose that in some $\mathfrak{C}$ there is a WORLD k such that $\mho(w\phi, k) = 1$ and $\mho(\Box w\phi, k) = 0$. But $\mho(w\phi, k) = \mho(\phi, \mathcal{F}(w,k)) = 1$. On the other hand, $\mho(\Box w\phi, k) = 0$ iff for some k' such that $R(k,k')$, $\mho(w\phi, k') = \mho(\phi, \mathcal{F}(w,k')) = \mho(\phi, \mathcal{F}(w,k)) = 0$ —contradiction. Necessity follows easily if we appeal to the substitution method from the theory of correspondence (van Benthem 1984). Again the focus is on (2.7R). Its translation is

(2.9.a) $(\forall P)(\forall w)(\forall k)(P(\mathcal{F}(w,k)) \to (\forall k')(R(k,k') \to P(\mathcal{F}(w,k'))))$

By substituting $\mathcal{F}(w,k) = *$ for $P(*)$, we get

$$(2.9.b) \quad (\forall w)(\forall k)(\mathcal{F}(w,k) = \mathcal{F}(w,k) \rightarrow (\forall k')(R(k,k') \rightarrow \mathcal{F}(w,k) = \mathcal{F}(w,k')))$$

and further

$$(2.9.c) \quad (\forall w)(\forall k)(\forall k')(R(k,k') \rightarrow \mathcal{F}(w,k) = \mathcal{F}(w,k'))$$

Since $R(k,k')$ holds for all k and k', (2.9.c) is equivalent to

(2.9.d) $(\forall w)(\forall k)(\forall k')(\mathcal{F}(w,k) = \mathcal{F}(w,k'))$, Q.E.D.

The following results can be proved in the same vein: at LK,

(2.10.1) $\Box\phi \rightarrow w\phi$ defines the condition that $R(k,\mathcal{F}(w,k))$ for all k and w;

(2.10.2) $w\Box\phi \rightarrow \phi$ defines the condition that $R(\mathcal{F}(w,k),k)$ for all k and w;

(2.10.3) $w\Box\phi \rightarrow \Box w\phi$ defines the condition that if $R(k,k')$, then $R(\mathcal{F}(w,k),k')$, for all k,k' and w;

(2.10.4) $\Box\phi \rightarrow w\Box\phi$ defines the condition that if $R(\mathcal{F}(w,k),k')$, then $R(k,k')$, for all k,k' and w.

(Hint: the appropriate substitutions in the translations of the sentences of $\mathcal{L}$ of $P(*)$ are: $R(k,*)$ in the case of (2.10.1); $\neg(k = *)$ in the case of (2.10.2); $\mathcal{F}(w,k) = *$ in the case of (2.10.3); and again $R(k,*)$ in the case of (2.10.4). Note that, since for all k and w the function $\mathcal{F}(w,k)$ is always defined, the condition $R(k,\mathcal{F}(w,k))$ entails that $(\exists k')R(k,k')$, and the condition $R(\mathcal{F}(w,k),k')$ entails that $(\exists k')R(k',k)$.)

III.

The aim of a local possible WORLDS semantics is not only to show that each world can say whatever some WORLD does, but also that each world can say whatever all WORLDS do. A world will then reflect inside itself anything that happens in a model $\mathfrak{C}$. This aim will be approached in two steps. In this section we shall define conditions to the effect that the reflection relation R is reduced to another relation the definition of which appeals to worlds. In the next one, we shall show how it is possible to design a copy of the whole model $\mathfrak{C}$ within each of the WORLDS in K. Hence, what we need is to create inside any WORLD $k \in K$ a full image of all the other WORLDS in K, to the effect that any claim of $\mathfrak{C}$ that a sentence ϕ is true in some WORLD k' is rendered inside k as a claim of k that ϕ holds at some world w. Suppose indeed that $\mho(\phi, k') = 1$, and suppose that this is rendered inside k as the claim that for some w the sentence $w\phi$ is true at k: $\mho(w\phi, k) = 1$. To put it in other words, each sentence ϕ is true at k' if and only if it is true at k that ϕ holds at w. But in this case the WORLD k provides a reflection of k' in it, and specifically it reflects the WORLD k' as the world w. As it looks from k, the world w is an exact copy of k'. If this happens, then we should agree that k fully reflects all the WORLDS in an 'adequate' manner. But then for defining what it is for a sentence ϕ to be true in a model $\mathfrak{C}$ it suffices to know what the WORLD k claims about it: if it claims that ϕ holds at any world w, then we may conclude that ϕ is true at any WORLD k in K.

If this aim were accomplished, our semantics would be 'local' in the sense that the elements of K (the WORLDS) simulate the whole of the model $\mathfrak{C}$. They turn out to be extremely rich entities: not only that they settle any issue about usual facts, but they also settle issues about WORLDS themselves. And we might try to retain just one WORLD k and get rid of the WORLDS k', k', k'', etc. while appealing only to their diaphanous counterparts in k—i.e., the worlds w, w', w'', etc.

To start with, consider the conditions under which a sentence $\Box\phi$ is true at some WORLD k: by D1g, $\mho(\Box\phi, k) = 1$ iff

$$(3.1) \qquad (\forall k')(R(k, k') \to \mho(\phi, k') = 1)$$

So, ϕ is necessary at k iff ϕ is true at all the WORLDS k' sueh that $R(k, k')$ holds. How could we render this condition inside k? The intuition is that k renders ϕ necessary iff at k according to all worlds w it is the case that ϕ. Hence, the claim that at k the sentence ϕ is necessary is rendered as:

$$(3.2) \qquad (\forall w)\mho(w\phi, k) = 1$$

A straight asymmetry appears to come into play with (3.2): for, while in expression (3.1) a sentence is regarded as necessary at k if true at all WORLDS that satisfy some clause, in (3.2) nothing comparable is involved. But this dissolves immediately, once we note that (3.2) also involves some restrictions. According to D1h, (3.2) is equivalent to:

$$(3.2') \qquad (\forall w)\mho(\phi, \mathcal{F}(w, k)) = 1$$

By (3.2′), ϕ is necessary at k iff it is true at all WORLDS k' which in k are reflected as some world. However, we have no guarantee that the set of all WORLDS k' such that there is a world w of which it holds that $\mathcal{F}(w,k) = k'$ is exactly K. That happens only if $\mathcal{F}$ maps $W \times K$ onto K. Hence (3.2′) requires that ϕ be true only at worlds $k' = \mathcal{F}(w,k)$.

The conditions expressed by (3.1) and (3.2) are equivalent if the following obtains:

$$(3.3) \qquad R(k,k') \leftrightarrow (\exists w)\mathcal{F}(w,k) = k'$$

Proof. First, (3.1) implies (3.2). Suppose (3.1) is true, but (3.2) is false. Then it is true that $(\exists w)\mho(\neg\phi, \mathcal{F}(w,k)) = 1)$. Let k'' be that WORLD such that $\mathcal{F}(w,k) = k''$. Then $\mho(\neg\phi, k'') = 1$. From this and (3.1) we get that $R(k,k'')$ is not the case. But, if (3.3) is true, then $(\exists w)\mathcal{F}(w,k) = k'' \rightarrow R(k,k'')$ and hence $R(k,k'')$ must be the case—contradiction. Conversely, (3.2) implies (3.1). Suppose that (3.2) is true, but (3.1) is false, i.e. $(\exists k')(R(k,k') \wedge \mho(\neg\phi,k') = 1)$. Since (3.3) is supposed to be true, $R(k,k') \rightarrow (\exists w)\mathcal{F}(w,k) = k'$. So $\mho(\neg\phi, \mathcal{F}(w,k)) = 1$, and hence $\mho(w\neg\phi, k) = 1$, and further $\mho(w\phi,k) = 0$, which contradicts the supposition that (3.2) is true.

Next, observe that to assert $(\exists w)\mathcal{F}(w,k) = k'$ is to assert that some binary relation holds between k and k'. Call this relation $\Re$. Intuitively, k and k' (in this order) are related by $\Re$ iff k' is reflected in k as some world. By (3.3), $\Re(k,k')$ obtains iff $R(k,k')$ does. It is also possible to prove that $\mho(\Box\phi,k) = 1$ iff

$$(3.4) \qquad (\forall k')(\Re(k,k') \rightarrow \mho(\phi,k') = 1)$$

Expression (3.4) is similar in form with (3.1), and thus the usual definition of $\mho$ for necessary sentences is preserved. (3.4) is equivalent with each of the following expressions:

$$(3.4.1) \qquad (\forall k')((\exists w)\mathcal{F}(w,k) = k' \rightarrow \mho(\phi,k') = 1)$$

$$(3.4.2) \qquad (\forall k')(\forall w)(\mathcal{F}(w,k) = k' \rightarrow \mho(\phi,k') = 1)$$

$$(3.4.3) \qquad (\forall w)(\forall k')(\mathcal{F}(w,k) = k' \rightarrow \mho(\phi,k') = 1)$$

$$(3.4.4) \qquad (\forall w)\neg((\exists k')(\mathcal{F}(w,k) = k') \wedge \neg\mho(\phi,\mathcal{F}(w,k)) = 1)$$

$$(3.4.5) \qquad (\forall w)\neg(\neg\mho(\phi,\mathcal{F}(w,k)) = 1)$$

$$(3.4.6) \qquad (\forall w)\mho(\phi,\mathcal{F}(w,k)) = 1$$

$$(3.4.7) \qquad (\forall w)\mho(w\phi,k) = 1$$

The move from (3.4.4) to (3.4.5) is based on the fact that $\mathcal{F}$ is a function, and hence that $(\exists k')(\mathcal{F}(w,k) = k')$ is always true. But we have already shown that $\mho(\Box\phi,k) = 1$ iff $(\forall w)\mho(w\phi,k) = 1$, and this ends the proof.

The problem is, what local modal logic is defined by the models in which (3.3) holds? A precise answer is given by:

(3.5) Let LT (local T) be the logic resulting by adding to LK:

(3.5.1) All sentences of the form $\Box\phi \rightarrow w\phi$, for all ϕ.

(3.5.2) The rule: if $\vdash_{LT} w\phi$ for each w, then $\vdash_{LT} \Box\phi$.

Then ϕ is LT-deducible iff ϕ is true in all models $\mathfrak{C}$ in which (3.3) holds, i.e. $\Re \leftrightarrow R$.

Condition (3.5.1) is a local counterpart of the T-principle (2.7T). The latter states that if a sentence ϕ is necessary, then ϕ is true (in the reference WORLD); (3.5.1) turns this to: if a sentence ϕ is necessary, then ϕ holds at every world w. Condition (3.5.2) is the local counterpart of the necessitation rule. According to it, if $w\phi$ is LT-provable for each world w, then $\Box\phi$ must be LT-provable. The intuition behind (3.5.1) and (3.5.2) is that worlds are similar to WORLDS, and by imposing the two conditions an attempt is made to render, with respect to worlds, standard conditions on WORLDS.

I shall sketch only the necessity part of the proof of (3.5). To do that, I shall use again the substitution method. According to (2.10.1) above, condition (3.5.1) defines

(3.5.1a) $(\forall w)(\forall k)R(k, F(w,k))$

By usual calculations, (3.5.1.a) is equivalent to

(3.5.1b) $(\forall w)(\forall k)(\forall k')(F(w,k) = k' \rightarrow R(k,k'))$

and further

(3.5.1c) $(\forall k)(\forall k')((\exists w)(F(w,k) = k') \rightarrow R(k,k'))$

i.e. $\Re \rightarrow R$. Expression (3.5.1c) renders half of (3.3). The other half comes with some transformations on the translation of (3.5.2). Indeed, we have:

(3.5.2a) $(\forall P)((\forall w)(\forall k)P(F(w,k)) \rightarrow (\forall k)(\forall k')(R(k,k') \rightarrow P(k')))$

Let us substitute $(\exists w')F(w',k) = *$ for $P(*)$:

(3.5.2b) $(\forall w)(\forall k)((\exists w')(F(w',k) = F(w,k))) \rightarrow$
$(\forall k)(\forall k')(R(k,k') \rightarrow (\exists w')(F(w',k) = k'))$

Since the antecedent of (3.5.2b) is always true, we get:

(3.5.2c) $(\forall k)(\forall k')(R(k,k') \to (\exists w')(\mathcal{F}(w',k) = k'))$,

and this states that $R \to \Re$. Hence, (3.5.1f) and (3.5.2c) together state that $\Re \leftrightarrow R$, Q.E.D.

Notice that, since $\mathcal{F}$ is a function, $\Re$ has the property that $(\forall k)(\exists k')\Re(k,k')$. This corresponds to

(3.6) $\vdash_{LT} \Box\phi \to \neg\Box\neg\phi$

which can be easily derived in LT by appeal to (3.5.1). Indeed, from $\Box\phi \to w\phi$ and $\Box\neg\phi \to w\neg\phi$ we get $(\Box\phi \wedge \Box\neg\phi) \to (w\phi \wedge w\neg\phi)$, i.e. $(\Box\phi \wedge \Box\neg\phi) \to (w\phi \wedge \neg w\phi)$, which renders $\neg(\Box\phi \wedge \Box\neg\phi)$; and this is in turn equivalent to (3.6).

IV.

Since LT is semantically defined by the condition $\Re \leftrightarrow R$, it follows that, given D1, the expression:

(4.1) $\mho(\Box\phi, k) = 1$ iff $\mho(\phi, k') = 1$ for all k' such that $\Re(k,k')$

holds. Now, this retains the standard way of defining the truth-value of necessary sentences at a world, except that $\Re$ replaces the more usual alternativeness relation R. Further, observe that the definition of $\Re$ is only in terms of $\mathcal{F}$, and does not depend upon the component R of the model. This suggests the possibility of letting a model $\mathfrak{C}$ be simply a structure $\langle K, \mathcal{F}, \mho\rangle$ and use (4.1) to modify D1g as:

(D1g$'$) If ϕ is $\Box\psi$, then $\mho(\phi, k) = 1$ iff $\mho(\phi, k')$ for all k' such that $\Re(k,k')$.

Keeping in mind expressions (3.1) and (3.2), this manoeuvre entails that $\mho(\Box\phi, k) = 1$ is definable as $(\forall w)\mho(w\phi, k) = 1$. Thus, at k a sentence ϕ is necessary iff for each w the sentence ϕ is the case at w.

Notice that if a modal logic contains LT, its semantics can be correspondingly simplified. Let us take as an example the local S5, for short LS5. Syntactically, it results by adding to LT all sentences of the forms:

(4.2.1) $\Box\phi \to \phi$

(4.2.2) $\Diamond\phi \to \Box\Diamond\phi$

Semantically, it is defined by the condition that $\Re$ be total:

(4.2.3) $(\forall k)(\forall k')\Re(k,k')$ or, equivalently: $(\forall k)(\forall k')(\exists w)\mathcal{F}(w,k)=k'$

The proof does not bring new elements as compared with that for standard S5.

Intuitively, (4.2.3) says that each WORLD is reflected in each WORLD as some world. And this is an essential part of the second step of the burden of a local possible WORLDS semantics, to show that each world can say within itself whatever all WORLDS do. For, indeed, by (4.2.3) worlds are able to create an inner image of some WORLD.

Further, we can easily check that at LS5 the following is always valid:[3]

(4.3.1) $w\Box\phi \rightarrow \Box w\phi$

(the reason is that $\mathcal{F}$ is always defined for all arguments w and k). Its converse, though, is not valid. At LS5 it defines a stronger condition on $\mathcal{F}$:

(4.3.2′) $(\forall w)(\forall k)(\exists k')\mathcal{F}(w,k')=k$

Indeed, starting with $\Box w\phi \rightarrow w\Box\phi$ we get its translation in predicate logic (remember that $\Re$ is total):

(4.3.2a) $(\forall P)(\forall w)((\forall k)P(\mathcal{F}(w,k)) \rightarrow (\forall k)P(k))$

and further, by substituting $(\exists k')\mathcal{F}(w,k')=*$ for P:

(4.3.2b) $(\forall w)((\forall k)(\exists k')\mathcal{F}(w,k')=\mathcal{F}(w,k) \rightarrow (\forall k)(\exists k')\mathcal{F}(w,k')=k)$

Since the antecedent of (4.3.2b) always obtains, elementary calculations prove that this expression entails (4.3.2′). Given the function $\mathcal{F}$, we may define a family of unary functions $\mathcal{F}_w : K \rightarrow K$ by letting $\mathcal{F}_w(k)=k'$ iff $\mathcal{F}(w,k)=k'$. For every w, (4.3.2′) comes to

(4.3.2c) $(\forall k)(\exists k')\mathcal{F}_w(k')=k$

and this states that $\mathcal{F}_w(K)=K$. (Exercise: What is the semantical condition for each world's being self-conscious?)

The implication is that whenever the converse of (4.3.1), and thus (4.3.2′), holds each world reflects each WORLD at some WORLD. Let

[3](4.3.1) is a far cry from the converse of Barcan's formula.

ϕ be some sentence such that it is not true in a model $\mathfrak{C}$. Then there is some WORLD k in K such that $\mho(\phi, k) = 0$. Now let k' be some WORLD in K. If the converse of (4.3.1) holds, then there is some world w such that at k' the WORLD k is reflected as world w. Hence at k' we have that $\neg\phi$ is the case at w, i.e. $\mho(\neg w\phi, k') = \mho(w\neg\phi, k') = 1$. The WORLD k' mimics the entire model $\mathfrak{C}$. All WORLDS are reconstructed in k' as worlds, and this operation is adequate in the sense that a sentence ϕ is true at some WORLD k iff at k' the world corresponding to k is such that ϕ is true at it. In this way, the claim that a sentence ϕ is true at some model $\mathfrak{C}$ can be established by concentrating exclusively on what is happening at any single WORLD k.

We could then get rid of all WORLDS, and reconstruct all they might say inside some k, as claims of worlds w, w', w'' , etc. All semantical claims would thus be rendered inside k. But this does not entail that k' reproduces within itself, not only each WORLD k in K, but also the structure of $\mathfrak{C}$. For it is possible that more worlds be such that at k' they (adequately) mimic the same WORLD k; and, conversely, if two WORLDS make true the same sentences, then nothing prevents that some world may be the reflection at k' of both. Hence it would be interesting to establish conditions that, given (any) WORLD k, then with respect to k WORLDS and worlds be one-to-one correlated.[4] However, this goes beyond the objectives of this paper.

References

van Benthem, J. 1984. Correspondence theory. In D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, vol. II*, pp. 167–247. D. Reidel, Dordrecht.

Plantinga, A. 1976. Actualism and possible worlds. *Theoria*, 42, pp. 139–160.

Faculty of Philosophy
Bucharest University, Romania

[4] I shall not explore in this paper the issue of extending the language $\mathcal{L}$ to allow quantifications over worlds. However, it is interesting to ask what logics different modal conditions define in this case. Suppose, indeed, that our language $\mathcal{L}$ is modified as follows: the world symbols from W are replaced with world variables, and quantifications over those variables are allowed. Then, e.g., $(\exists w)w\phi$ and $(\forall w)w\phi$ are sentences of $\mathcal{L}$. It is not difficult to see that the expressions of $\mathcal{L}$ can be translated into a modal language $\mathcal{L}'$ of the predicate calculus, with predicates of infinite arity. The question is, What constraints on the expressions of $\mathcal{L}'$ correspond to our modal conditions?