ON LOGICAL NECESSITY (*)

Philip P. HANSON

> *Some* would claim now a coherent understanding of modal contexts in the light of the extensive work on possible-worlds semantics .... There remains a nagging doubt ..., however, that the problem solved by the use of the possible worlds approach is *not quite* the original problem of modal logic, only an analogous one sharing certain formal similarities (Dana Scott, [27], p. 788).

In this paper I will consider an extension of S5 that I dub 'S'. The intuitions underlying S have received formal treatment at least as early as Carnap, and have resurfaced in the literature in different formal guises with surprising regularity.(1) Nevertheless, the present re-examination is justified on several grounds. First, S is here characterized semantically and syntactically making close use of ideas of Gentzen and Scott ([10], [27], [28], [29]). I think this mode of representation casts an interesting light on certain contentious issues in the 'metaphysics' of logical necessity. Secondly, some philosophical insight can be gained by properly locating S within the context of other recent work on necessity. And third, artificial intelligence researchers have recently become interested in so-called 'non-monotonic logics' for modelling inferential processes involving the rejection of previously held beliefs (cf. [20]). Non-monotonic logics are cha-

(*) I have benefitted from discussions on aspects of this paper with Raymond Bradley, David Copp, Raymond Jennings, Bernard Linsky and David Zimmerman. Hans Herzberger and Peter Schotch kindly commented on an early 1979 draft. Richard Routley brought the new literature on 'non-monotonic logics' to my attention.

(1) cf., e.g., [3], [4], [5], [19], [34]. Kit Fine mentioned in correspondence that David Lewis, citing Scott, suggested something like S several years ago in correspondence with him. An anonymous journal referee said that essentially the same semantical ideas form the basis of David Kaplan's S13 in his UCLA Ph. D. (1964) dissertation, *Foundations of Intensional Logic*.

racterizable as "... logics in which the introduction of new axioms can invalidate old theorems" ([20], p. 41). Such logics characteristically involve deliberately non-truth preserving rules. Interestingly enough, S also has a non-truth preserving rule, although the theoretical concerns leading to S are apparently quite distinct from those leading the artificial intelligence people to non-monotonic logics. In any case, the formal analogy is striking enough to merit attention.

I. The Problem

We will approach the several concerns of this paper by reviewing an issue in modal logic which has some claim to historical priority, if by 'history' we mean the history of modal since *Principia Mathematica* ([26]). C. I. Lewis, a father of modern modal logic, thought that 'A logically implies B' could be analysed as 'Necessarily if A then B;' in symbols: $\ulcorner \Box(A \supset B) \urcorner$.(2) By definitional abbreviation, $\ulcorner \Box(A \supset B) \urcorner$ became $\ulcorner (A \prec B) \urcorner$, which Lewis then read as 'A strictly implies B.'

Lewis' two desiderata for a 'logical calculus' were that it be both a "... canon and critique of deductive inference..." ([16], p. 247), that it be a system not only *of* implications but *about* implications, and took his own calculi of strict implication to be logical calculi in this sense. Modal logic as conceived by Lewis was thus not only *part* of logic, but also embodied an analysis or theory of logical implication. Its theorems were not only valid truths on a par with the law of the excluded middle, $\ulcorner (A \vee \neg A) \urcorner$, but at the same time expressed truths

(2) cf. [16]. 'A' and 'B' are metavariables ranging over declarative English sentences. They are therefore replacaeble by *names* of particular English sentences formed by putting the sentences in regular quotations. '$\supset$' is the dyadic sentential operator expressing the material conditional. '$\ulcorner (A \supset B) \urcorner$' is true just in case either A is false or B is true. '$\Box$' is the monadic sentential operator expressing necessity. The inward-facing corner quotes are Quine's 'quasi-quotations' (cf. [22], pp. 33-7). '$\ulcorner \Box(A \supset B) \urcorner$' refers to the result of putting in particular English sentences (*not* their names) in place of 'A' and 'B', appropriately replacing the sentential operators with their English translations, and then replacing the quasi-quotes with regular quotations marks. In effect, then '$\ulcorner \Box(A \supset B) \urcorner$' names the sentential context 'Necessarily, if ... then ---,' but by using a mixture of symbols from both the object language and the metalanguage. The use of quasi-quotes avoids use-mention confusion. Of course, Lewis did not himself have the paraphernalia of quasi-quotes; indeed, he apparently did not have clearly in mind the distinction of language levels signaled by their use.

about the binary relation, on sentences, of logical implication. It is because Lewis saw Russell's material conditional as capable of satisfying at best only the first desideratum for a logical calculus that he found it wanting.

But, as is well-known, Quine argued that nothing could coherently fulfill both these desiderata, and that Lewis' critique of Russell was conceived in several sins simultaneously ([22], pp. 14-33). At the level of syntax, Lewis had confused the distinct roles of binary sentential operators and two-place predicates. This syntactical confusion reflected a deeper semantical confusion about what could be properly expressed using such constructions. The relation of implication could only be expressed by a predicate, since sentential operators were syncategorematic and thus nonreferential. But the worst sin of all was confusing 'use' and 'mention.' For since implication was a relation between *sentences*, identifying an instance of this relation involved mentioning, i.e., *naming* the sentences so related, and this was properly done in the 'metalanguage' in which we talk about the 'object language' to which those sentences belonged, not in that object language itself. By contrast, a sentential operator functioned within the object language, forming more complex sentences, not out of the names of simpler sentences, but out of those sentences themselves. The latter were 'used,' not merely mentioned. Failure to observe this distinction led to dire paradox, as Tarski and others had shown.

But with the advent, in the fifties, of 'possible worlds semantics,' Quine's criticisms seemed to lose their sharp edge. His charge that modal logic was incoherent could now be answered in terms of its consistency and completeness under its formal model-theoretic interpretation. A strictly implied B just in case every possible world in which A was true was a world in which B held also; which was just another way of saying that $\ulcorner(A \supset B)\urcorner$ held in every world, i.e., was necessarily true. (3)

(3) See [12], [13]. My rendition of the possible worlds interpretation of strict implication, if the crudest, is also the most transparent. In certain modal systems an 'accessibility relation' on worlds is also invoked. In these systems A strictly implies B, *relative to a given world,* just in case every world *accessible from that given world* in which a is true is also one in which B is true. Different constraints on accessibility result in different systems. But whatever the merits of an accessibility relation for other analytical uses of 'strict implication,' there seems to be nothing in our intuitions about

The focus of philosophical defence and attack now shifted to the notion of possible worlds. How many were there and what were they like? Did accepting the model theory commit one to the existence of nonactual possible worlds in some 'fullblooded realist' sense, or, by a more 'moderate realism,' to their existence merely as, say, uninstantiated properties or as maximal consistent sets of propositions; or could 'possible worlds talk' be construed merely as talk about different conceivable states of affairs involving actually existing or imagined entities? (4) And by now modal logics were proliferating at an astounding, some said unseemly, rate. Which one embodied a conception of necessity appropriate for the analysis of logical implication as strict implication?

Some said 'none,' on such grounds as that if logical implication were the necessity of the material conditional, then A would logically imply B for *any A and any necessary* B. This violated intuitions that there must be a 'connection of meaning' between an *implicans* and its *implicandum*. (5) I will set such objections aside here, on the grounds

logical implication to which it corresponds. An accessible introduction to these complexities is [6].

(4) Naturally, all of these schools of thought and more are represented in current philosophical polemics on possible worlds semantics. Declared representative of 'full-blooded realism' is David K. Lewis (see [17], pp. 85-6). Representative of the 'imaginable counterfactual situation' interpretation is Saul Kripke (see [14], p. 267). Would-be 'moderate realists' include Robert Adams [1], and Robert Stalnaker [33].

(5) Lewis was well aware of this and the other so-called 'paradoxes of strict implication,' and of the fact that they paralleled the paradoxes of material implication which he had invoked against Russell's (confused) talk of the material conditional as expressing implication. But to the extent that Lewis does insist on implication involving a 'connection of meaning' between *implicans* and *implicandum* and on its analysis explicitly reflecting this, I think that Lewis' attempts to avoid the paradoxes of strict implication fail. 'Relevance theorists' make much of the latter paradoxes. Anderson's and Belnap's 'system of entailment' suffers from an attenuated form of the same problem, however: witness the *embedded* entailments in their "... strong and natural list of valid entailments ..." ([2], p. 26); e.g.,

$$(A \to B) \to (A \to (C \to C))), \text{ and}$$
$$(A \to A) \to B) \to B$$

Although the Anderson-Belnap system can no longer be faulted for lack of a semantics, it turns out that the semantics that it now has is a possible worlds semantics involving a complicated triadic counterpart to Kripke's diadic 'accessibility relation' on worlds (cf. n. 3 above, and also [23], [24], and [25]).

that we want a theory of *logical* implication to embody only the most general principles of implication, only those which hold independently of the meaining of sentences. C. I. Lewis, in fact, worked with a formal language whose 'atomic sentences' were completely unanlysed ([16], p. 122).

A more pressing reason for rejecting, say, Lewis' S5 as an account of logical implication in terms of strict implication, concerns our intuitions about possibility. If A is necessarily true just in case it is true in all possible worlds, A is possibly true just in case it is true in some possible world. If A is necessarily true, then our logic of necessity ought to say so by making $\ulcorner \Box A \urcorner$ a theorem. By the same token, if A is possibly true then $\ulcorner \neg\Box\neg A \urcorner$, or by definitional abbreviation, $\ulcorner \Diamond A \urcorner$, ought to be a theorem of our logic. Now it seems that for any declarative sentence not expressing a self contradiction we can imagine a possible world in which it holds. Therefore $\ulcorner \Diamond A \urcorner$ ought to be a theorem for *any* non-selfcontradictory A, and, for instance, this will include all *atomic sentences* of the formal language. But this does not hold for *any* atomic sentences in S5. So if our intuitions are correct, S5 is apparently, if anything, *not inclusive enough*. ([6]) This has obvious implications for the analysis of logical implication as strict implication. For all theorems of the form $\ulcorner \Diamond A \urcorner$, we will have $\ulcorner \Box(B \supset \Diamond A) \urcorner$, for any B, and therefore $\ulcorner B \mathrel{-\!\!3} \Diamond A \urcorner$. Given that we are setting aside considerations of meaning connection in our quest for the most general principles of implication (see above), B does, indeed, *logically* imply any such theorem. So S5 strict implication is apparently inadequate as a theory of logical implication. Is any modal system adequate?

(6) This problem does not arise in the same way in connection with S5's alternate semantics, involving a reflexive, symmetric, and transitive accessibility relation on worlds. Such a relation in effect exhaustively divides the worlds into mutually exclusive equivalence classes. So, A is possibly true *at a world w*, (say, the actual world), just in case it is true in some world accessible from w, *and therefore in the same equivalence class*. This can fail to be the case even though A is true in some world belonging to a different equivalence class. So our intuitions about non-self-contradictory sentences A being true in some world *can* be built into an S5 semantics. But then what notion of possibility does S5 capture if truth in a possible world is not enough to make $\ulcorner \Diamond A \urcorner$ a theorem? Surely not the notion of logical possibility. By the same token it is problematic as to whether S5 strict implication expresses logical implication.

II. A Formal Representation

Our language will consist *initially* of a denumberable set, $\mathfrak{S}$, of 'unanalysed' sentences, represented by 'propositional constants,' a_1, $a_2, \ldots$, closed under a relation of 'conditional assertion,' $\vdash$. Sentences are to be true or false. A valuation is a function, $\mathfrak{v}$, from sentences to truth values: $\mathfrak{v}: \mathfrak{S} \to \{T,F\}$. For each sentence $A \varepsilon\ \mathfrak{S}$, $\mathfrak{v}(A)$ is the truth value of A as assigned by the function $\mathfrak{v}$. $\vdash$ holds between *finite subsets* of $\mathfrak{S}$. In the metalanguage, 'A,' 'A_0,' 'A_1,'..., 'A_{n-1},' 'B,' 'B_0,' 'B_1'..., 'B_{m-1}' range over members of $\mathfrak{S}$. We let '$\mathfrak{A}$,' '$\mathfrak{B}$,' '$\mathfrak{C}$,'...'$\mathfrak{A}'$,' '$\mathfrak{B}'$,' '$\mathfrak{C}'$,' ... stand for finite subsets of $\mathfrak{S}$. Given some intended set of valuations, $\mathfrak{V}$, we further intend that

$\mathfrak{A} \vdash_{\mathfrak{V}} \mathfrak{B}$ if and only if (hereafter iff) for all $\mathfrak{v}\ \varepsilon\ \mathfrak{V}$, whenever $\mathfrak{v}(A_i) = T$ for all $A_i\ \varepsilon\ \mathfrak{A}$, $\mathfrak{v}(B_j) = T$ for some $B_j\ \varepsilon\ \mathfrak{B}$.

If we assume that the sets are such that $\mathfrak{A} = \{A_0, A_1, \ldots, A_{n-1}\}$ and $\mathfrak{B} = \{B_0, B_1, \ldots, B_{m-1}\}$, then we may write $\mathfrak{A} \vdash_{\mathfrak{V}} \mathfrak{B}$ as: $A_0, A_1, \ldots. A_{n-1} \vdash_{\mathfrak{V}} B_0, B_1 \ldots, B_{m-1}$ for short (i.e., minus the set brackets).

In a relationship $\mathfrak{A} \vdash_{\mathfrak{V}} \mathfrak{B}$ we call the set $\mathfrak{A}$ the *antecedent*, and the set $\mathfrak{B}$ the *consequent*. It is asserted that at least one of the members of the consequent is true, conditional on the truth of all the members of the antecedent. If $\mathfrak{A} \vdash_{\mathfrak{V}} \mathfrak{B}$ does not hold, we may write '$\mathfrak{A} \nvdash_{\mathfrak{V}} \mathfrak{B}$'. '$\vdash_{\mathfrak{V}} \mathfrak{B}$' will abbreviate '$\phi \vdash_{\mathfrak{V}} \mathfrak{B}$,' where ϕ is the null set; '$\vdash_{\mathfrak{V}} \mathfrak{B}$' asserts *un*conditionally that at least some member of $\mathfrak{B}$ (not necessarily the same for each valuation) is always true. If $\mathfrak{B}$ consists of a single sentence B_j, then B_j is unconditionally asserted.

What formal properties does such a binary relation, $\vdash$, on sets of sentences possess? It is not difficult to establish that the following rules obtain.

(R) $\mathfrak{A} \vdash \mathfrak{B}$ if $\mathfrak{A} \cap \mathfrak{B} \neq \phi$ ('reflexivity')

(M) $\dfrac{\mathfrak{A} \vdash \mathfrak{B}}{\mathfrak{A}, \mathfrak{A}' \vdash \mathfrak{B}, \mathfrak{B}'}$ ('monotonicity')

where the notation '$\mathfrak{A}, \mathfrak{A}'$' stands for '$\mathfrak{A} \cup \mathfrak{A}'$' (on either side of '$\vdash$')

(T) $\mathfrak{A} \vdash B, \mathfrak{C}$

$$\frac{\mathfrak{A}, \mathrm{B} \vdash \mathfrak{C}}{\mathfrak{A} \vdash \mathfrak{C}} \quad \text{('transitivity')}$$

where the notation 'B, $\mathfrak{C}$' stands for '$\{B\} \cup \mathfrak{C}$' (on either side of '$\vdash$')

Are there any other essential properties of such a relation? Suppose that we are given denumerable $\mathfrak{S}$, and a binary relation $\vdash$ defined on finite subsets of $\mathfrak{S}$ as satisfying the rules (R), (M), and (T). Is $\vdash$ then a relation of conditional assertion? How can we show this?

We can define $\mathfrak{V}_{\vdash}$ as the set of valuations 'consistent' with $\vdash$, i.e., as the set of all valuations $\mathfrak{v}$, such that whenever $\mathfrak{A} \vdash \mathfrak{B}$ and $\mathfrak{v}(A) = T$ for all $A \varepsilon \mathfrak{A}$, $\mathfrak{v}(B) = T$ for some $B \varepsilon \mathfrak{B}$. But given $\mathfrak{V}_{\vdash}$ we know that we can *define* a relation of conditional assertion on $\mathfrak{S}$ and relative to $\mathfrak{V}$ as follows.

$\mathfrak{A} \vdash_{\mathfrak{V}_{\vdash}} \mathfrak{B}$ iff whenever $\mathfrak{v} \varepsilon \mathfrak{V}_{\vdash}$ and $\mathfrak{v}(A) = T$ for all $A \varepsilon \mathfrak{A}$, then $\mathfrak{v}(B) = T$ for some $B \varepsilon \mathfrak{B}$.

It can be shown that, so long as $\vdash$ satisfies (R), (M), and (T), $\vdash = \vdash_{\mathfrak{V}_{\vdash}}$, and therefore that $\vdash$ is indeed a relation of conditional assertion. But since $\vdash$ was completely characterized by (R), (M), and (T), it follows that satisfaction of those three syntactical rules is necessary and sufficient for a binary relation on finite subsets of sentences to be a relation of conditional assertion in the sense of our semantic definition. (It is also not difficult to show that, provided $\mathfrak{V}$ is finite, $\mathfrak{V} = \mathfrak{V}_{\vdash_{\mathfrak{V}}}$.)

We can now *extend* our language by extending $\mathfrak{S}$ to a larger set $\mathfrak{S}_{PC}$, by adding closure under truth-functional negation, '$\neg$,' and the material conditional, '$\supset$.' We ignore other truth-functional connectives, as they are easily defined in terms of these. At the same time we extend each $\mathfrak{v} \varepsilon \mathfrak{V}$ to a $\mathfrak{v}_{PC}$ defined on $\mathfrak{S}_{PC}$ so as to be consistent with our intended interpretations of truth-functionally complex sentences. We intend for each $\mathfrak{v} \varepsilon \mathfrak{V}$, that

$$\mathfrak{v}_{PC}(\ulcorner \neg A \urcorner) = T \text{ iff } \mathfrak{v}_{PC}(A) = F\text{; and}$$
$$\mathfrak{v}_{PC}(\ulcorner A \supset B \urcorner) = T \text{ iff } \mathfrak{v}_{PC}(A) = F \text{ or } \mathfrak{v}_{PC}(B) = T.$$

We then have

$$\mathfrak{V}_{PC} = \{\mathfrak{v}_{PC} : \mathfrak{v} \varepsilon \mathfrak{V}\}$$

as the set of intended valuations for $\mathfrak{S}_{PC}$. This allows us to define $\vdash_{\mathfrak{V}_{PC}}$, which proves to be an extension of $\vdash_{\mathfrak{V}}$. It is then easy to show that our intended interpretations of '$\neg$' and '$\supset$' are completely and soudly captured syntactically by the following rules:

$$(\neg)\ \frac{\mathfrak{A} \vdash \mathrm{A},\ \mathfrak{B}}{\mathfrak{A},\ \ulcorner \neg \mathrm{A} \urcorner \vdash \mathfrak{B}}$$
$$(\supset)\ \frac{\mathfrak{A},\ \mathrm{A} \vdash \mathrm{B},\ \mathfrak{B}}{\mathfrak{A} \vdash \ulcorner \mathrm{A} \supset \mathrm{B} \urcorner,\ \mathfrak{B}}$$

where the double bar signifies that the rule holds in both directions. That is *given* that (R), (M), and (T) jointly capture, with respect to $\mathfrak{S}_{PC}$, our intended interpretation of $\vdash_{\mathfrak{V}_{PC}}$, we can show that $(\neg)$ and $(\supset)$ hold for $\mathfrak{S}_{PC}$ just in case $\ulcorner \neg \mathrm{A} \urcorner$ and $\ulcorner \mathrm{A} \supset \mathrm{B} \urcorner$ are interpreted as above.

We now extend our language yet further by extending $\mathfrak{S}_{PC}$ to $\mathfrak{S}_{\Box}$, by adding closure under the modal operators for necessity, '$\Box$,' possibility, '$\Diamond$,' and strict implication, '$\prec$.' '$\Box$' is introduced as a primitive, while the other two are defined by means of it as follows.

$$\ulcorner \Diamond \mathrm{A} \urcorner = \mathrm{df}\ \ulcorner \neg \Box \neg \mathrm{A} \urcorner$$
$$\ulcorner \mathrm{A} \prec \mathrm{B} \urcorner = \mathrm{df}\ \ulcorner \Box (\mathrm{A} \supset \mathrm{B}) \urcorner$$

At the same time we extend each $\mathfrak{v}_{PC} \varepsilon \mathfrak{V}_{PC}$ to a $\mathfrak{v}_{\Box}$ defined on $\mathfrak{S}_{\Box}$ so as to be consistent with our intended interpretation of modally complex sentences. We intend, for each $\infty_{PC} \varepsilon \mathfrak{V}_{PC}$, that

$\mathfrak{v}_{\Box}(\ulcorner \Box \mathrm{A} \urcorner) = \mathrm{T}$ iff $\mathfrak{v}_{\Box}(\mathrm{A}) = \mathrm{T}$ for all $\mathfrak{v}_{PC} \varepsilon \mathfrak{V}_{PC}$,
$\mathfrak{v}_{\Box}(\ulcorner \Diamond \mathrm{A} \urcorner) = \mathrm{T}$ iff $\mathfrak{v}_{\Box}(\mathrm{A}) = \mathrm{T}$ for some $\mathfrak{v}_{PC} \varepsilon \mathfrak{V}_{PC}$, and
$\mathfrak{v}_{\Box}/\ulcorner \mathrm{A} \prec \mathrm{B} \urcorner) = \mathrm{T}$ iff for all $\mathfrak{v}_{PC} \varepsilon \mathfrak{V}_{PC}$, if $\mathfrak{v}_{\Box}(\mathrm{A}) = \mathrm{T}$ then $\mathfrak{v}_{\Box}(\mathrm{B}) = \mathrm{T}$.

So we then have

$$\mathfrak{V}_{\Box} = \{\mathfrak{v}_{\Box} : \mathfrak{v}_{PC} \varepsilon \mathfrak{V}_{PC}\}$$

as the set of intended valuations for $\mathfrak{S}_{\Box}$. This again allows us to define $\vdash_{\mathfrak{V}_{\Box}}$, which again proves to be an extension of $\vdash_{\mathfrak{V}_{PC}}$ (and thus of $\vdash_{\mathfrak{V}}$).

In addition to the definitions for '$\Diamond$' and '$\prec$,' we need only the following two rules to completely characterize syntactically our intentions with respect to '$\Box$,' '$\Diamond$,' and '$\prec$.'

$$(\Box)\ \frac{\mathfrak{A} \vdash \mathrm{B}}{\ulcorner \Box \mathfrak{A} \urcorner \vdash \ulcorner \Box \mathrm{B} \urcorner}$$

where '$\ulcorner \Box \mathfrak{A} \urcorner$' abbreviates '$\{\ulcorner \Box \mathfrak{A}_i \urcorner : \mathrm{A}_i \varepsilon \mathfrak{A}\}$' and 'B' abbreviates '$\{\mathrm{B}\}$'.

$$(\neg\Box)\quad \frac{\nvdash A}{\vdash \ulcorner \neg\Box A \urcorner}$$

$(\neg\Box)$ says that if A is *not* unconditionally asserted, then $\ulcorner \neg\Box A \urcorner$ *is*.(7)

Of course $(\neg\Box)$ is not a rule if rules must take you from one conditional assertion to another. $(\neg\Box)$ cannot be used to establish conditional assertions on the basis of previously established conditional assertions, as is usual in natural deduction systems. Nor is $(\neg\Box)$ a 'truth preserving' rule as this is usually understood. Still, $(\neg\Box)$ is a perfectly coherent principle which, as we shall demonstrate, does help to capture, syntactically, our intended interpretation of '$\Box$'.(8)

Syntactically, our formal representation of logical implication, S, is constructed in $\mathfrak{S}_{\Box}$, and consists of the rules (R), (M), (T), $(\neg)$, $(\supset)$, $(\Box)$ and $(\neg\Box)$, together with the definitions for '$\Diamond$' and '$\prec$' (and definitions of other truth-functional operators, as required). To show that these rules capture our semantic intentions it remains simply to prove that $(\Box)$ and $(\neg\Box)$ hold for $\mathfrak{S}_{\Box}$ just in case: $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$ iff $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_{\Box}$, given that analogous results are previously obtainable for the non-modal fragment of S. It remains, therefore, to prove four conditionals.

1. To prove: If $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$ iff $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_{\Box}$, then $(\Box)$ holds.

(7) We could have preserved a certain aesthetic uniformity to our system by formulating $(\neg\Box)$, equivalently, as

$$\frac{\mathfrak{A} \nvdash B}{\ulcorner \neg\Box \mathfrak{A} \urcorner \vdash \ulcorner \neg\Box B \urcorner}\quad \text{, or indeed as}\quad \frac{\mathfrak{A} \nvdash B}{\mathfrak{C} \vdash \ulcorner \neg\Box B \urcorner}$$

(8) In a *limited* way $(\neg\Box)$ could, in fact, function deductively if applied to 'anti-theorems' *via* a 'rule' such as $\frac{A \vdash}{\nvdash A}$. But this would only get us to $\vdash \ulcorner \neg\Box A \urcorner$ when A is a contradiction, whereas $(\neg\Box)$ is also applicable to contingent propositions.

Clearly, in order that all of the intended theorems of the form $\ulcorner \neg\Box A \urcorner$ where A is contingent be *derivable* in S as it stands, we would have to add the relevant A as 'anti-theorems,' i.e. as '$\nvdash A$.' I have not bothered to do this because my concerns here are not with derivability but just with the notions of logically necessary truth (and logical implication) *per se*. For this purpose I need not prove semantic completeness and soundness in the usual sense. It will be sufficient to demonstrate that the rules for necessity obtain in S if and only if '$\Box$' is interpreted as we have done.

We assume that (1) $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$ iff $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$, and that (2) $\mathfrak{A} \vdash B$. We want to show that $\ulcorner \Box \mathfrak{A} \urcorner \vdash \ulcorner \Box B \urcorner$. Suppose that $\mathfrak{v}_{\Box_i}(B) = F$. Then, by (1), $\mathfrak{v}_{\Box_j}(B) = F$ for some $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$. But then, by (2) $\mathfrak{v}_{\Box_j}(A_i) = F$ for some $A_i \varepsilon \mathfrak{A}$, from which it follows, by (1), that $\mathfrak{v}_{\Box_i}(\ulcorner \Box A_i \urcorner) = F$ for some $A_i \varepsilon \mathfrak{A}$. Therefore, whenever $\mathfrak{v}_{\Box_i}(\ulcorner \Box A_i \urcorner) = T$ for all $A_i \varepsilon \mathfrak{A}$, $\mathfrak{V}_{\Box_i}(\ulcorner \Box B \urcorner) = T$; i.e. $\ulcorner \Box \mathfrak{A} \urcorner \vdash \ulcorner \Box B \urcorner$, by definition of $\vdash$. Q.E.D.

2. To prove: If $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$ iff $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$ then $(\neg\Box)$ holds.

We again assume that (1) $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$ iff $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$, and also that (2) $\nvdash A$. We want to show that $\vdash \ulcorner \neg\Box A \urcorner$. From (2) we know that $\mathfrak{v}_{\Box_k}(A) = F$ for some $\mathfrak{v}_{\Box_k} \varepsilon \mathfrak{V}_\Box$. But then from (1) it follows that $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = F$ *for any*, *i.e. all* $\mathfrak{v}_{\Box_i} \varepsilon \mathfrak{V}_\Box$. But then since $\mathfrak{S}_\Box$ is closed under '$\neg$' and $\mathfrak{v}_\Box$ is consistent with our intended interpretation of '$\neg$' we have that $\mathfrak{v}_{\Box_i}(\ulcorner \neg\Box A \urcorner) = T$ for all $\mathfrak{v}_{\Box_i} \varepsilon \mathfrak{V}_\Box$, which is just to say that $\vdash \ulcorner \neg\Box A \urcorner$. Q.E.D.

3. To prove: If $(\Box)$ and $(\neg\Box)$ hold, then if $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$, then $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$.

We assume $(\Box)$ and $(\neg\Box)$ and that (3) $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$. We must show that $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$. From (3) and our definition of '$\vdash$' it follows that $\vdash A$, and therefore, by $(\Box)$ that $\vdash \ulcorner \Box A \urcorner$. But then $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$. Q.E.D.

4. To prove: If $(\Box)$ and $(\neg\Box)$ hold, then if $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$ then $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$.

We assume $(\Box)$ and $(\neg\Box)$ and that (3) $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$. We must show that $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$. We proceed by *reductio ad absurdum*. Thus we suppose that (4) $\mathfrak{v}_{\Box_j}(A) = F$ for some $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$. From (4) and the definition of '$\vdash$' we would have that $\nvdash A$. From (3) and the definition of '$\neg$' we would have that $\mathfrak{v}_{\Box_i}(\ulcorner \neg\Box A \urcorner) = F$, and thus that $\nvdash \ulcorner \neg\Box A \urcorner$. But then using $(\neg\Box)$ we could infer not $\nvdash A$, i.e., $\vdash A$. So we would have both $\vdash A$ and $\nvdash A$. Therefore we must deny our *reductio* assumption. Therefore $\mathfrak{v}_{\Box_j}(A) = T$ for all $\mathfrak{v}_{\Box_j} \varepsilon \mathfrak{V}_\Box$. Q.E.D.

Now S contains S5. S5 is constructed in $\mathfrak{S}_\Box$, and consists of the principles (R), (M), (T), ($\supset$), ($\neg$), ($\Box$), the definitions of '$\Diamond$' and '$\dashv$,' and the following principles.

($\Box$ R) $\ulcorner \Box A \urcorner \vdash A$
($\Box$ T) $\ulcorner \Box A \urcorner \vdash \ulcorner \Box \Box A \urcorner$
($\Box$ S) $\ulcorner \Diamond \Box A \urcorner \vdash A$

To establish that S contains S5, it remains only to establish that ($\Box$ R), ($\Box$ T), and ($\Box$ S) are contained in S. Since we have shown that the syntax and semantics of S are matched, we may proceed semantically.

5. To prove: ($\Box$ R).

Assume, by *reductio*, that $\ulcorner \Box A \urcorner \nvdash A$. Then for some $A_i \,\varepsilon\, \mathfrak{S}_\Box$ and for some $\mathfrak{v}_{\Box_i} \varepsilon\, \mathfrak{V}_\Box$, $\mathfrak{v}_{\Box_i}(\ulcorner \Box A_i \urcorner) = T$ and $\mathfrak{v}_{\Box_i}(A_i) = F$. But $\mathfrak{v}_{\Box_i}(\ulcorner \Box A_i \urcorner) = T$ iff $\mathfrak{v}_{\Box_j}(A_i) = T$ for *all* $\mathfrak{v}_{\Box_j} \varepsilon\, \mathfrak{V}_\Box$. Thus $\mathfrak{v}_{\Box_i}(A_i) = T$, which conflicts with our previous result. Therefore we must reject our *reductio* assumption. Q.E.D.

6. To prove: ($\Box$ T).

Assume, by *reductio*, that $\ulcorner \Box A \urcorner \nvdash \ulcorner \Box \Box A \urcorner$. Then for some $A_i \,\varepsilon\, \mathfrak{S}_\Box$ and for some $\mathfrak{v}_{\Box_i} \varepsilon\, \mathfrak{V}_\Box$, $\mathfrak{v}_{\Box_i}(\ulcorner \Box A_i \urcorner) = T$ and $\mathfrak{v}_{\Box_i}(\ulcorner \Box \Box A_i \urcorner) = F$. If $\mathfrak{v}_{\Box_i}(\ulcorner \Box A_i \urcorner) = T$, then $\mathfrak{v}_{\Box_j}(A_i) = T$ for *all* $\mathfrak{v}_{\Box_j} \varepsilon\, \mathfrak{V}_\Box$. If $\mathfrak{v}_{\Box_i}(\ulcorner \Box \Box A_i \urcorner) = F$ then $\mathfrak{v}_{\Box_k}(\ulcorner \Box A_i \urcorner) = F$ for some $\mathfrak{v}_{\Box_k} \varepsilon\, \mathfrak{V}_\Box$. Take one such $\mathfrak{v}_{\Box_k}$. Then $\mathfrak{v}_{\Box_l}(A_i) = F$ for some $\mathfrak{v}_{\Box_l} \varepsilon\, \mathfrak{V}_\Box$, but this conflicts with our previous result. Therefore we must reject our *reductio* assumption. Q.E.D.

7. To prove ($\Box$ S).

Assume that $\mathfrak{v}_{\Box_i}(A) = F$. Then $\nvdash A$, by definition of '$\vdash$'. Therefore $\vdash \ulcorner \neg \Box A \urcorner$ by ($\neg \Box$). Therefore $\vdash \ulcorner \Box \neg \Box A \urcorner$ by ($\Box$), and by definition of '$\Diamond$' this transforms into $\vdash \ulcorner \neg \Diamond \Box A \urcorner$. By definition of '$\vdash$' it follows that $\mathfrak{v}_{\Box_i}(\ulcorner \neg \Diamond \Box A \urcorner) = T$ and by definition of '$\neg$' that $\mathfrak{v}_{\Box_i}(\ulcorner \Diamond \Box A \urcorner) = F$. So, whenever $\mathfrak{v}_{\Box_i}(A) = F$, $\mathfrak{v}_{\Box_i}(\ulcorner \Diamond \Box A \urcorner) = F$. Which is to say that whenever $\mathfrak{v}_{\Box_i}(\ulcorner \Diamond \Box A \urcorner) = T$ then $\mathfrak{v}_{\Box_i}(A) = T$, i.e., $\ulcorner \Diamond \Box A \urcorner \vdash A$. Q.E.D.

But though S contains S5, S5 does not contain S, for the rule ($\neg \Box$) fails in S5. For let $\mathfrak{v}_{\Box_i}(A) = F$, $\mathfrak{v}_{\Box_i}(\ulcorner \Box \Box A \urcorner) = F$, $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = F$, $\mathfrak{v}_{\Box_i}(\ulcorner \Diamond \Box A \urcorner) = F$, $\mathfrak{v}_{\Box_i}(\ulcorner \neg \Box A \urcorner) = T$; and, for some $\mathfrak{v}_{\Box_j} \neq \mathfrak{v}_{\Box_i}$, let

$\mathfrak{v}_{\Box_j}(\ulcorner \neg\Box A \urcorner) = F$. $\mathfrak{v}_{\Box_i}$ and $\mathfrak{v}_{\Box_j}$ are jointly consistent with all the principles of S5. Yet since $\mathfrak{v}_{\Box_i}(A) = F$ we have by defintion of '$\vdash$' that $\nvdash A$, and since $\mathfrak{v}_{\Box_j}(\ulcorner \neg\Box A \urcorner) = F$ we have by definition of '$\vdash$' that $\nvdash \ulcorner \neg\Box A \urcorner$. Q.E.D.

Just as there is an obvious relationship between our interpretations of '$\vdash$' and of'$\Box$', to wit:

$$\vdash A \text{ iff } \vdash \ulcorner \Box A \urcorner$$

so there is an obvious relationship between our interpretations of '$\vdash$' and of '$\mathbin{-\!\!\!3}$,' to wit:

$$A \vdash B \text{ iff } \vdash \ulcorner A \mathbin{-\!\!\!3} B \urcorner$$

We can think of '$\Box$' as a device for expressing truths, in the object language, about *logical truth*, that may also may also be expressed metalinguistically using '$\vdash$.' And we can think of '$\mathbin{-\!\!\!3}$' as a device for expressing truths in the object language about *logical implication*, that may also be expressed metalinguistically using '$\vdash$.' For we can take Gentzen's relation of conditional assertion as a formal representation, at the metalinguistic level, of the most general properties of implication. This ability of S, and indeed of S4 and S5, to 'confuse' the validity of a conditional assertion with the unconditional assertion of a strict implication can be represented syntactically by the rule:

$$\text{(C)} \quad \frac{A \vdash B}{\vdash \ulcorner A \mathbin{-\!\!\!3} B \urcorner}$$

which Scott dubs the 'rule of confusion.' (9) Apparently, then, what

(9) Cf. [27], [29]. Scott has a very nice result to the effect that for a certain class of modal systems containing the rule (C) and certain others, a 'vertical rule' of the form

$$\begin{array}{c} \mathfrak{A}_0 \vdash \mathfrak{B}_0 \\ \mathfrak{A}_1 \vdash \mathfrak{B}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathfrak{A}_{n-1} \vdash \mathfrak{B}_{n-1} \\ \hline \mathfrak{A}_n \vdash \mathfrak{B}_n \end{array}$$

is derivable within the system if and only if its 'horizontalization' also is: $\hat{\mathfrak{A}}_0 \mathbin{-\!\!\!3} \check{\mathfrak{B}}_0$, $\hat{\mathfrak{A}}_1 \mathbin{-\!\!\!3} \check{\mathfrak{B}}_1, \ldots, \hat{\mathfrak{A}}_{n-1} \mathbin{-\!\!\!3} \check{\mathfrak{B}}_{n-1} \vdash \hat{\mathfrak{A}}_n \mathbin{-\!\!\!3} \check{\mathfrak{B}}_n$, where '$\hat{\mathfrak{A}}_i$' signifies the result of conjoining all the members of A_i and '$\check{\mathfrak{B}}_j$' signifies the result of disjoining all the members of $\mathfrak{B}_j$ (with

Lewis wanted can be had. In Scott's pregnant phrase, "...metatheory can be (fragmentarily) self-applied..." ([27], p. 804).

Steven K. Thomason has kindly pointed out to me that S is closely related to a system he described in [34]. The relation is that Thomason's system is a *recursive axiomatization* of the valid *un*conditional assertions of S. Theorems of the form $\vdash \ulcorner \neg\Box A \urcorner$ for contingent A are, in effect, included in his axioms, which include all formulae of the form

$$\Diamond \& \{a_i{}^* \mid i = 1, \ldots, n\},$$

where $a_1, \ldots, a_n$ are *distinct propositional constants* and each $a_i{}^*$ is either a_i or $\neg a_i$. Thomason notes that such a system is not closed under substitution. His 'new representation of S5' is, in effect, that the theorems of S5 are precisely those formulae of his system such that all of their substitution instances are theorems of his system. ([10])

If S is not closed under substitution, *is it a logic*? That is, is it a logic in the sense that its theorems are truths of logic and its conditional assertions hold by virtue of logic? Surely one of our more persistent intuitions is that a logical truth is so merely by virtue of its *form,* where the notion of form includes the idea that each logical truth is a member of a whole class of logical truths *sharing the same form* and

association to the left). This result turns out to hold for S4 and S5, as well as our own S, provided it is remembered that S's characteristic rule ($\neg\Box$) does not qualify as A 'vertical rule' of the form circumscribed by Scott. Given a suitably defined 'accessibility relation' on valuations, S4 and S5 can also be provided with a 'worldless' semantics in the manner of S.

([10]) Since substitutivity fails in Thomason's system and in S, these systems are not 'normal extensions' of S5 in the sense of [21], p. 7. Therefore neither are these extensions 'quasi-normal' in the sense of Schiller Joe Scroggs in [30], p. 112, and thus are not covered by Scroggs' result: i.e., that all quasi-normal proper extensions of S5 have a finite characteristic matrix, where S5 is not a proper extension of itself, and where a finite characteristic matrix for a propositional calculus is a matrix of a finite number of elements (truth values) which satisfies those and only those formulae which are provable in it (see [7]). Scroggs says, "The class of quasi-normal extensions of S5 is a very broad class and actually includes all extensions which are likely to prove interesting." ([31], p. 12). I disagree. More recently, Krister Segerberg has defined 'modal logic' in such a way that if a set of formulae fails to be closed under substitution, it fails to be a modal logic ([31]). One can accept Segerberg's definition, in the context of his essay, without accepting its innuendo.

thus satisfying closure under some appropriately articulated rule of substitution.

Without having to legislate on this intuition, I think that, should it be sustained, its negative implications for the status of S as logic should not necessarily be viewed as problematic. Even if S were closed under substitution, that would not necessarily *be sufficient* for it to qualify as logic. For if S is, in Scott's phrase, "... metatheory self-applied..." (see above), then perhaps S is *not* logic *per se*, but rather a formalization of (part of) *the theory of logic*. The truths of the theory of logic need not be considered themselves to be truths of logic. To so consider them may well involve a kind of use mention confusion less forgiveable than Lewis'.

III. Discussion

A. *Worlds*

Notice that 'worlds' do not appear in the semantic for S. They do not appear, that is, unless 'worlds' are valuations. Of course if 'worlds' are valuations, then they are abstract entities of a familiar sort, which, if they exist at all, exist in all possible worlds including this one! So we were dealing with 'worlds' all along even when giving the standard semantics for the truth-functional propositional calculus.

But perhaps it is not that worlds *are* valuations, but that they may be *represented by* valuations in our semantics. We can, if we like, say that the valuation which maps all and only true sentences into the value, True, characterizes and *is realized by* the actual world. But do we need to suppose that the other valuations are realized by other, nonactual, worlds in order that those valuations make good sense of sentences, in e.g., the truth-functional propositional calculus? Surely their existence as unrealized functions is quite sufficient. If so, we do not need to suppose that valuations 'represent' nonactual worlds in any sense which implies that those valuations are realized. Valuations can represent possibilities without realizing possible worlds.

One can argue for this as follows. Intuitively, possible worlds *cannot* be *identified* with valuations *or any other abstract platonic entities*. The whole idea of possibilia was that they *need not be actual,* so to identity them with necessary and therefore actual existents would rob them of their distinguishing feature, and improperly conflate the notions of abstractness and possibility.

This suggests, in fact, that there may in general be something wrong with the standard set-theoretic representations of possible worlds and other unactualized possibilia if taken too literally. For consider a principle of set existence: that sets do not contain nonexistent members.([11]) Suppose we postulate possibilia – worlds, individuals, etc. – as the members of certain *sets*, as is standardly done in the semantics for modal logics. But since sets do not contain nonexistent members, therefore, by our principle, in postulating the sets, we postulate their members. But to postulate them is to postulate their *actual* existence. So if these sets existed, as per a literal construal of the standard semantics, so would the 'unactualized possibilia,' which seems incoherent.

David Lewis ([17], pp. 86-7) would take this somewhat controversial principle to beg questions about the interpretation of 'exist' and perhaps 'set.' It is one thing to exist, another to 'actually exist.' Perhaps it is one thing to be a set, another to be an 'actual set.' Lewis could then say that our principle holds only if given a narrow actualist interpretation: actual sets do not contain non-actually existing members. Our argument would not then go through, and set-theoretic representations of possibilia would be sustained at face value. These moves seem to me themselves questionbegging and controversial. But the issues, including intricacies of modal set theory, cannot be further explored here.

B. *Non-monotonic Logic*

Now S is 'monotonic': in particular, the introduction of new theorems *via* the rule ($\neg\Box$) does not invalidate any previous theorems. But the syntactical similarity between ($\neg\Box$) and an inference rule cited as a non-monotonic rule is striking:([12])

([11]) Cf. [11], p. 44. A recent and fairminded discussion of related set-theoretical principles is found in [9].

([12]) Cited in [20], p. 50. The authors there want to "capture the idea" of (N) (see main text), but claim that (N) as it stands is circular: " 'Derivable' means 'derivable

$$\text{(N)}\ \frac{\nvdash \neg A}{\vdash \Diamond A}$$

especially given that applying to it the definition of '$\Diamond$' we obtain

$$\frac{\nvdash \neg A}{\vdash \neg\Box\neg A}$$

which is contained in ($\neg\Box$). What then is non-monotonic about (N)? As far as I can tell, nothing about it *per se* is non-monotonic except perhaps that one is allowed to envisage it applying to 'theories' that include both 'logical' and 'nonlogical' axioms, and the *nonlogical* axioms might at one time include, e.g.:

(1) B
(2) $\ulcorner(B \,\&\, \neg\Box\neg A) \supset A\urcorner$
and (3) $\ulcorner C \supset \neg A\urcorner$ (cf. [20], p. 44)

Now if we also have at time that $\nvdash \neg A$, then by (N), together with (1) and (2), A will be derivable as a 'theorem.' But of course '$\neg\Box\neg A$' in this context has to mean something like 'A is consistent with the (other) assertions of the theory.' But if at a *later* time C gets added as a further nonlogical axiom, this can force the withdrawal of A as a theorem. So monotonicity fails, not because of the syntax of (N) *per se* but because of how it is being interpreted when applied to particular nonlogical theorems in this way.

In that case perhaps we may say that S has a syntax that lends itself to non-monotonic applications and interpretations. But the representation of logical necessity embodied in S is not non-monotonic *per se*.

C. *Iteration*

One objection to our approach that naturally arises is that while

from axioms by inference rules.' So we cannot define an inference rule in terms of derivability so casually" (p. 50). If this were a well-founded worry then most classical inference rules, e.g., *Modus Ponendo Ponens* would be objectionable on the same grounds.

It is probably worth emphasizing here that this notion of a 'non-monotonic logic' is unrelated to our so called rule of 'monotonicity', (M). I have carried over the latter terminology from Scott [27].

sentential operators iterate, predicates do not. So how can $\ulcorner \Box A \urcorner$ express in the object language the metalinguistic claim that A is logically valid? Does $\ulcorner \Box\Box A \urcorner$ then express that 'A is logically valid' is logically valid? What does this mean? M. H. Löb, in a ground-breaking paper, puts the problem succinctly, and suggests a solution.

> In a relation to a given formal system, S say, the notion of a world may be satisfactorily identified with that of a model. Then we are led to define $\Box\phi$ where ϕ is a P_0–formula, as 'ϕ is true in all models of P_0.' This approach is, however, not available when ϕ itself contains occurrences of the square. Consider, for instance, the formula $\Box\Box\phi$, where ϕ is a P_0–formula. For the inner formula, $\Box\phi$, as rendered by our interpretation, is not a P_0–formula, but belongs to some metalanguage, P_1 say. Therefore the notion 'model of P_0' is not applicable to it, thus preventing us from interpreting the outer square as we did the inner one. It is natural, however, to attempt to interpret the outer square analogously to the inner one by rendering it, for instance, as true in all of an appropriate class of models of P_1. Similarly consideration of formulae such as $\Box\Box\Box\phi$, etc., would lead us to repeat the previous observation and to talk about models of, P_2 say, and so on. We therefore wish to have at our disposal an infinite hierarchy of metalanguages $P_0, P_1, P_2, \ldots$ such that the notion of a model of P_i can be defined in P_{i+1} ([18], pp. 23-4).

Given such a hierarchy of metalanguages, we can make sense of *something like* iteration of the predicate '... is true in all models': namely that the predicate ''...' is true in all P_{i+1} models' is true of sentences of the form ''...' is true in all P_i models,' for $i \geq 0$. But does '$\Box$' express logical validity on this treatment, as opposed to some broader merely model-theoretic notion of validity-at-some-level-of-a-hierarchy-of-metalanguages? If A is a truth of logic, is 'A is valid in P_0' a truth of logic in P_1? It is unclear to me that we want $\ulcorner \Box\Box A \urcorner$ to mean that 'A is logically valid' is logically valid. But if not that, what?

We can avoid a hierarchy of metalanguages in the way that we have in fact done so. Following Scott, we have in our system S that for any sentence $A \,\varepsilon\, \mathfrak{S}_\Box$, modalized or not, $\mathfrak{v}_{\Box_j}(\ulcorner \Box A \urcorner) = T$ iff $\mathfrak{v}_{\Box_j}(A) = T$ in all $\mathfrak{v}_{\Box_j} \varepsilon\, \mathfrak{V}_\Box$. Of course it is true that the $\mathfrak{v}_{\Box_j} \varepsilon\, \mathfrak{V}_\Box$ are *extensions* of the valuations $\mathfrak{V}_{PC}$ for the propositional calculus, which in turn are extensions of the valuations $\mathfrak{V}$ of the 'calculus of unanalysed sentences' involving only '$\vdash$'. So we *could*, alternately, say that $\mathfrak{v}_{\Box_i}(\ulcorner \Box A \urcorner) = T$ iff $\mathfrak{v}_{PC_j}(A) = T$ for all $\mathfrak{v}_{PC_j} \varepsilon \mathfrak{V}_{PC}$ *as extended to* $\mathfrak{V}_{\Box}$ (or iff $\mathfrak{v}_j(A) = T$ for all $\mathfrak{v} \varepsilon \mathfrak{V}$ *as extended to* $\mathfrak{V}_{\Box}$). But such extensions do not have to involve a hierarchy of metalanguages. Indeed the $\mathfrak{V}_{PC}$ have been extended to $\mathfrak{V}_{\Box}$ precisely so as to make $\ulcorner \Box A \urcorner$ true in all $\mathfrak{V}_{\Box}$ if A is, for *all* $A \varepsilon \mathfrak{S}_{\Box}$. The deliberate effect of this is to make $\ulcorner \Box\Box A \urcorner$ and $\ulcorner \Box A \urcorner$ *express the same thing*: logical validity of A, where A is unmodalized. In Löb's construction they do not. This must not be confused with the fact that both $\ulcorner \Box A \urcorner \vdash \ulcorner \Box\Box A \urcorner$ and $\ulcorner \Box\Box A \urcorner \vdash \ulcorner \Box A \urcorner$ hold in S, S5, and S4. These principles hold in Löb's construction as well, but there $\ulcorner \Box A \urcorner$ and $\ulcorner \Box\Box A \urcorner$ make distinct claims. The point is that given Löb's semantics these principles are *non-trivial* and, if '$\Box$' is to express logical validity, *non-obvious*.

More recently, Brian Skyrms, citing Löb, has explored this sort of interpretation ([32]). Skyrms says, "You may, if you please, regard the modal operators as the object-linguistic shadows of metalinguistic predicates. The 'projecting down' of metalinguistic predicates to enrich an object language is, in fact, a process of general interest. Negation can be thought of as the projection down of falsity. In a slightly more complicated way, the existential quantifier can be thought of as the projection down of 'is satisfied.' From this standpoint, the problem of mixing modality and quantification is the problem of simultaneaously projecting down two metalinguistic categories" ([32], p. 387). I find this very suggestive and in the spirit of our own approach. But I would also then argue that the 'projecting down' of the notion of logical validity no more needs to be a projecting down from a hierarchy of metalanguages than the projecting down of falsity into the 'calculus of unanalysed propositions' (see above) needs to be a projecting down from a hierarchy of metalanguages. $\ulcorner \neg\neg\neg A \urcorner$ means the same as $\ulcorner \neg A \urcorner$, not "'A is false' is false' is false.'

Our interpretation of iterated modal operators in S receives further clarification and explanation when S is extended by quantification (see below).

D. *Adding Individual Quantifiers*

What happens whan we add to S the syntax and rules for the first order quantifiers and identity? Call such an extension $Q^{=}S$. An answer is given in a very interesting paper by Nino Cocchiarella ([5]). He establishes the following lemma with respect to sentences A, $B \,\varepsilon\, \mathfrak{S}_{Q^{=}_{\Box}}$, where $\mathfrak{S}_{Q^{=}_{\Box}}$ is $\mathfrak{S}_{\Box}$ extended in the usual way by n-adic predicates (including at least one non-monadic predicate; identity will do), and individual variables and quantifiers, where $\mathfrak{V}_{\Box}$ is extended to $\mathfrak{V}_{Q^{=}_{\Box}}$ *via* the standard truth conditions for sentences with individual quantifiers and identity. Each $\mathfrak{v}_{Q^{=}_{\Box_i}} \,\varepsilon\, \mathfrak{V}_{Q^{=}_{\Box}}$ remains a function from $\mathfrak{S}_{Q^{=}_{\Box}}$ into {T, F}, but the truth value a particular sentence takes in $\mathfrak{v}_{Q^{=}_{\Box_i}}$ is now dependent on values assigned by a 'satisfaction function' to its predicates and variables *in a domain of individuals.*

Lemma: If $A \,\varepsilon\, \mathfrak{S}_{Q^{=}_{\Box}}$ is satisfiable, but only in an infinite domain, and $B \,\varepsilon\, \mathfrak{S}_{Q^{=}_{\Box}}$ is a modal and identity free sentence, then $\ulcorner A \supset \neg\Box B \urcorner$ is *logically* true iff B is *not* logically true ([5], p. 17).

The proof makes use of the Lowenheim-Skolem theorem. The significance of this lemma is immediate. Since the $B_i \,\varepsilon\, \mathfrak{S}_{Q^{=}_{\Box}}$ such that B_i is modal and identity free and not logically true *are not recursively enumerable*, neither are the modal truths of the form $\ulcorner A \supset \neg\Box B_i \urcorner$, and therefore neither are the modal truths belonging to $Q^{=}S$, which include these. *$Q^{=}S$ is semantically incomplete.*

Why is $Q^{=}S$ incomplete while quantified S5, say, is complete under, e.g., Kripke's semantics? The lemma shows why $Q^{=}S$ is incomplete, but how does quantified S5 escape the lemma?

Cocchiarella's thought is that it does so because $\ulcorner \Box A \urcorner$, under Kripke's semantic's is evaluated with respect to 'possible worlds,' *where these may correspond to any arbitrary subset of the set* $\mathfrak{V}_{Q^{=}_{\Box}}$. $\ulcorner \Box A \urcorner$ is said to be 'universally valid' just in case 'true in all worlds,' but what is true in all 'worlds' need not be true in all $\mathfrak{V}_{Q^{=}_{\Box}}$. The suggestion then is that, insofar as quantified S5 is semantically complete, to that extent its necessity operator fails to express logical necessity, since, intuitively, logical necessity is a matter of truth in *all* valuations. Cocchiarella says:

> The significance of this lemma should not be seen as rendering suspect our primary semantics for logical necessity. Indeed,

> . . . . , what this lemma shows is that there is a complete concurrence between logical necessity as an internal condition of modal free propositions . . . and logical truth as a semantical condition of the modal free (first order) sentences expressing these propositions . . . And it is precisely this concurrence which must hold if the modal operator for logical necessity is to represent merely formal and no material content. Of course, this means that we can express in modal terms the non-logical truth of modal free sentences of standard predicate logic (with identity) . . . . ([5], p. 18).

It is revealing to compare these remarks with the following heuristic remarks of Kripke's, in which D is a non-empty domain of individuals, G is the valuation representing the actual world, and K is a set of such valuations, including G.

> In trying to construct a definition of universal logical validity, it seems plausible to assume not only that the universe of discourse may contain an arbitrary number of elements and that predicates may be assigned any given interpretation in the actual world, but also that any combination of possible worlds may be associated with the real world with respect to some group of predicates. In other words, it is plausible to assume that no further restrictions need be placed on D, G, and K, except the standard one that D be non-empty. This assumption leads directly to our definition of universal validity ([12], p. 3).

Cocchiarella wants the notion of logical necessity to be truly universal by requiring that $\ulcorner \Box A \urcorner$ is true just in case A is true *in all valuations*. Kripke wants the notion of logical necessity to be truly universal by requiring that *its principles hold for any arbitrary set of valuations*. Now which is the true measure of universally: the truth of $\ulcorner \Box A \urcorner$ being determined relative to the set of all valuations, or relative to any, and thus, in effect, all *subsets* of valuations (note: the set of all valuations is a subset of itself)?

I would say that both Cocchiarella and Kripke are concerned with universality: Cocchiarella with the all-embracing scope of the claim that a proposition is *logically* necessary, and Kripke with what properties hold of *any and all notions of 'necessity.'* What Cochiarella, and we, are calling 'logical necessity' is a limiting case of the latter. *But precisely because it is a limiting case, its special properties are of special interest.* And, as we have seen, $Q^{=}S$ is semantically incomplete. But, as Cocchiarella intimates, it surely should not have surprised us that the whole truth about what is *and is not* logically necessary, i.e., true in all $\mathfrak{B}_{Q^{=}_{\Box}}$, and expressible in $\mathfrak{S}_{Q^{=}_{\Box}}$ should defy recursive specification. Thus the fact that $Q^{=}S$ is incomplete tells in its favour if anything.

Cocchiarella enlarges the philosophical impact of his incompleteness result by showing that it is equivalent to the incompleteness of (a proper subsystem of) standard second order logic with respect to its primary or standard semantics. There is a transformation definable on sentences in $\mathfrak{S}_{Q^{=}_{\Box}}$ such that a sentence is logically true just in case its transformation is. The key clause of the definition is this:

> $t(\Box\,\phi) = \wedge\pi_0, \ldots, \wedge\pi_{n-1}\, t(\phi)$, where $\pi_0, \ldots, \pi_{n-1}$ are all the predicate constants, now reconstructed as predicate variables of the same addicty, occurring in ϕ ([5], p. 18).

Cocchiarella also is able to turn this relation to explanatory advantage in connection with the problem of iteration (see above).

> The modal operator for logical necessity is interpreted according to the above transformation as a universal quantifier binding all the predicates occuring within its scope. The null effect of iterated occurences of the operator is hereby explained by the null effect of iterated quantifiers binding variables already bound. ([13])

([13]) [5], p. 19. A similar explanation for the redundancy of iterated modal operators is suggested in [15], pp. 1-2, but in connection with a construal of the necessity operator as a first order universal quantifier, '(|x),' ranging over all *possible individuals*. 'Necessarily' is held to signal intensional relations among *particular predicates*. Thus '$\Box$(x) (Fx $\supset$ Gx)' goes into '(|x) (Fx $\supset$ Gx)' and can be true if the right intensional relation obtains between 'Fx' and 'Gx'. But on Cocchiarella's translation scheme the same sentence would go into '(F) (G) (x) (Fx $\supset$ Gx)' where 'F' and 'G' are now predicate *variables*. But this sentence is not a valid sentence of second order theory. Cocchiarella would say, I think, that this construal of necessity brings in the *descriptive*

The relation might also help us with the interpretation of 'quantification into modal contexts,' and with the question of modal logic's status as logic; but these possibilities would require careful exploration, the intricacies of which are beyond the scope of this paper. Of course, to the extent that atomic sentences of S receive a partial 'analysis' in $Q^{=}S$, $Q^{=}S$ already takes us beyond the scope of C.I. Lewis' conception of a canon and critique of *logical* implication (see above).

There remains a problem related to, but I think distinct from, the fact that $Q^{=}S$ is semantically incomplete. The truths of $Q^{=}S$ defied recursive specification, but this is a problem about whether all 'modal facts' are expressible in language, recursively or otherwise. Thus (bears, of course, on any identification of possible worlds with valuations, but it also) bears on the *representational suitability* of valuations, since of course valuations are functions from sentences into truth values. A pivotal point is that $Q^{=}S$ has a denumerably infinite set of *atomic* sentences. So if these are of cardinality $\aleph_0$, $\mathfrak{V}_{Q^{=}_{\square}}$ is at most of cardinality $2^{\aleph_0}$. But suposing there to be at least $2^{\aleph_0}$ non-modal facts, there would have to be on the order of $2^{2^{\aleph_0}}$ 'possible worlds' involving these. And *there are* at least $2^{\aleph_0}$ non-modal facts, if e.g., we admit the real numbers as existents, or all subsets of a denumerable domain of individuals, or all points in continuous space-time. This suggests that we ought to view 'possible worlds' not as valuations, but as *language independent abstract entities*.

If postulating full-blooded worlds in order to explain our talk of possibilities seems like explanatory over-kill, postulating language independent platonic structures may seem scarcely to connect in any explanatory way with our modal talk. But if we can think of a reasonable modal logic as a *regimentation and idealization* of our pre-theoretic notions of possibility, necessity, and implication, there are ways of construing these mathematical structures that would relate them to valuations. One way is to liberalize the notion of a language and envision 'superlogics' constructed in 'superlanguages' with uncountably many sentences.([14]) Each $\mathfrak{v} \in \mathfrak{V}_{Q^{=}_{\square}}$ would then de-

element of intensional relations among particular predicates and thus fails in this respect to express the purely structural notion of logical necessity.

([14]) Perhaps this could be done along the lines of Kit Fine's 'ideal languages,' in [8], termine a *set* of 'supervaluations,' each member of which would be consistent with 𝔟.

Insofar as these notions could be understood as abstractions from our more familiar notions of language, logic, and valuation, the resulting conception would still connect with, and help to explain, two rather common intuitions. The first is that the interpretation of sentences of a natural language like English is often relative to different conceivable states of affairs. The second is that what interpretative differences are due to attendent differences in conceived states of affairs, and indeed what differences in such states there may be, are only imperfectly and incompletely reflected *in the interpretive conventions of the language*. The transfinite abstractions we make in modal *theory* could then simply be seen as parallel to those made in other mathematical and scientific theorizing on behalf of abstract and concrete uncountables.

Simon Fraser University Philip P. HANSON

REFERENCES

[1] Adams, R. M., 1974, 'Theories of Actuality,' *Noûs* 8, pp. 211-31.
[2] Anderson, A. R., and Belnap, N. E., 1975, *Entailment, The Logic of Relevance and Necessity, Vol. I* (Princeton University Press, Princeton).
[3] Carnap, R., 1946, 'Modalities and Quantifications,' *Journal of Symbolic Logic*, 11 pp. 33-64.
[4] Cocchiarella, N.B., 1974, 'Logical Atomism and Modal Logic,' *Philosophia*, 4, pp. 41-66.
[5] Cocchiarella, N. B., 1975, 'On the Primary and Secondary Semantics of Logical Necessity,' *Journal of Philosophical Logic* 4, pp. 13-27.
[6] Cresswell, M. J., and Hughes, G. E., 1968, *An Introduction to Modal Logic*, (Methuen and Co. Ltd., London).
[7] Dugundji, J., 1940, 'Note on a Property of Matrices for Lewis and Langford Calculi of Propositions,' *Journal of Symbolic Logic* 5, pp. 150-1.
[8] Fine, K., 1977, 'Properties, Propositions, and Sets,' *Journal of Philosophical Logic* 6, pp. 135-91.
[9] Fine, K., 1978, 'Modal Set Theory,' (unpublished).

p. 137, David Lewis' adumbrations to the contrary notwithstanding (see [17], p. 90 n. *).

[10] Gentzen, G., 1934, 'Investigations in Logical Dedeuction,' in Szabo, M. E. (ed.), 1969, *The collected Papers of Gerhard Gentzen* (North-Holland Pub. Co., Amsterdam), pp. 68-131.

[11] Hambourger, H., 1977, 'A Difficulty with the Frege-Russell Definition of Number,' *Journal of Philosophy* 74, pp. 409-14.

[12] Kripke, S.A., 1959, 'A Completeness Theorem in Modal Logic,' *Journal of Symbolic Logic* 24, pp. 1-14.

[13] Kripke, S. A., 1963, 'Semantical Considerations in Modal Logic,' *Acta Philosophica Fennica* 16, pp. 83-94.

[14] Kripke, S. A., 1972, 'Naming and Necessity,' in Davidson, D., and Harman, G. (eds.), 1972, *Semantics for Natural Language* (D. Reidel Pub. Co., Dordrecht-Holland), pp. 253-355, 763-9.

[15] Lambert, K., and van Fraassen, B. C., 1970, 'Meaning Relations, Possible Objects, and Possible Worlds,' in Lambert, K. (ed.), 1970, *Philosophical Problems in Logic* (D. Reidel Pub. Co., Dordrecht-Holland), pp. 1-19.

[16] Langford, C. H., and Lewis, C. I., 1932, *Symbolic Logic* (The Century Co., New York).

[17] Lewis, D. K., 1973, *Counterfactuals* (Harvard University Press, Cambridge, Mass.).

[18] Löb, M. H., 1966, 'Extensional Interpretations of Modal Logics,' *Journal of Symbolic Logic* 31, pp. 23-45.

[19] Makinson, D., 1966, 'How Meaningful are Modal Operators?' *Australasian Journal of Philosophy* 44, pp. 331-337.

[20] McDermott, D., Doyle, J., 1980, 'Non-Monotonic Logic I,' *Artificial Intelligence* 13, pp. 41-72.

[21] McKinsey, J. C. C., and Tarski, A., 1948, 'Some Theorems about the Sentential Calculi of Lewis and Heyting,' *Journal of Symbolic Logic* 13, pp. 1-15.

[22] Quine, W. v. O., 1940, *Mathematical Logic* (Harvard University Press, Cambridge, Mass.; 2nd ed. 1957).

[23] Routley, R., and Meyer, R. K., 1972, 'The Semantics of Entailment I,' in Leblanc, H. (ed.), 1972, *Truth, Syntax, and Modality* (North-Holland, Pub. Co., Amsterdam), pp. 199-243.

[24] Routley, R., and Meyer, R. K., 1972, 'The Semantics of Entailment II,' *Journal of Philosophical Logic* 1, pp. 53-73.

[25] Routley, R., and Meyer, R. K., 1972, 'The Semantics of Entailment III,' *Journal of Philosophical Logic* 1, pp. 192-208.

[26] Russell, B., and Whitehead, A., 1910, *Principia Mathematica, Vol. I* (Cambridge University Press, Cambridge, England).

[27]) Scott, D., 1971, 'On Engendering an Illusion of Understanding,' *Journal of Philosophy* 63, pp. 787-807.

[28] Scott, D., 1972, 'Background to Formalization,' in Leblanc, H. (ed.), 1972, *Truth, Syntax, and Modality* (North-Holland pub. Co., Amsterdam), pp 244-73.

[29] Scott, D., 1974, 'Rules and Derived Rules,' in Stenlund, S. (ed.), 1974, *Logical Theory and Semantic Analysis* (D. Reidel Pub. Co., Dordrecht-Holland), pp. 147-61.

[30] Scroggs, S. J., 1951, 'Extensions of the Lewis System S5,' *Journal of Symbolic Logic* 16, pp. 112-20.
[31] Segerberg, K., 1971, *An Essay in Classical Modal Logic Vol. I-III, Filosofiska studier* 13 (Uppsala).
[32] Skyrms, B., 1978, 'An Immaculate Conception of Modality, or, How to Confuse Use and Mention,' *Journal of Philosophy* 75, pp. 368-87.
[33] Stalnaker, R., 1976, 'Possible Worlds,' *Noûs* 10, pp. 65-75.
[34] Thomason, S. K., 1973, 'A New Representation of S5,' *Notre Dame Journal of Symbolic Logic* 14, pp. 281-4.