Proceedings of the British Academy, **95**, 85–133

Iterated Attitudes

TIMOTHY WILLIAMSON

THE AIM OF THIS PAPER is not to advance a thesis but to explore a phenomenon. The phenomenon is an easily overlooked kind of intensionality, whose neglect causes philosophical problems. It can be put like this: sometimes the set of principles satisfied by one sentential operator differs from, and may even be inconsistent with, the set of principles satisfied by another sentential operator coextensive with the first. These notions will be formally defined. The phenomenon arises when an operator is iterated, in the loose sense of being applied to a sentence in which it itself occurs. Examples will be analysed in detail. Several involve operators by which propositional attitudes are attributed to subjects; one involves metaphysical modality. An application will be made to some arguments against mechanistic views of mind. The phenomenon also raises a number of technical questions in bimodal logic, some of which will be answered.

1. *Five examples*

Example 1. According to what is sometimes known as the KK principle, if one knows something, then one knows that one knows it (in the propositional sense of 'know'). The principle is open to several cogent objections (for one objection see Williamson 1992). Even so, qualified versions of it may hold when appropriate restrictions or idealizations are introduced. In such cases, the question arises: under what modes of presentation of oneself to oneself does the qualified KK principle hold? Let us use the phrase 'singular term' broadly, to cover definite descriptions as well as names, demonstratives and pronouns such as 'I', without prejudice to the question whether definite descriptions are to be analysed as quantifiers. For any singular term *t*, '*t* knows that . . .' is therefore a singulary sentence operator (interpret quotation marks as corner quotes where appropriate). Let us say, relative to a fixed context, that the term *t* fits the KK principle just in case this holds

> (KK_t) For all P, if *t* knows that P, then *t* knows that *t* knows that P.

Suppose (what is not the case) that in the context of its use by me, 'I' fits the KK principle. Suppose also that, although I am in fact the man with a hole in his pocket, I do not know that I am the man with a hole in his pocket. In this context, does the definite description 'the man with a hole in his pocket' fit the KK principle? Suppose that the man with a hole in his pocket knows that it is sunny. Does it follow that the man with a hole in his pocket knows that the man with a hole in his pocket knows that it is sunny? It certainly follows that I know that it is sunny ('. . . knows that P' is an extensional context for '. . .'). Since KK_t holds for *t* = 'I' in this context, it follows that I know that I know that it is sunny. Thus the man with a hole in his pocket knows that I know that it is sunny. But it does not follow that the man with a hole in his pocket knows that the man with a hole in his pocket knows that it is sunny, or that I know that the man with a hole in his pocket knows that it is sunny. For all I know, the man with a hole in his pocket is asleep and has no idea what the weather is like ('*t* knows that . . . knows that P' is an intensional context for '. . .'). Thus 'the man with a hole in his pocket' does not fit the KK principle in this context, even though in this context the definite description picks out the same individual as 'I' does, and 'I' fits the KK principle in this context.

Even if an individual satisfies the KK principle in the intuitively intended sense, it does not follow that KK_t holds whenever the term *t* picks out that individual. Nor does KK_t specify any property of the individual picked out by *t*. To do that, one would also need to specify how the knower is to be specified in the content of the knowledge, in a way that depends only on who the knower is. For example, one might stipulate that an individual *x* satisfies the KK principle just in case KK_t holds for *t* = 'I' in a context in which *x* is the speaker. This comes close to the intuitively intended sense of the principle. Sophisticated defenders of the KK principle entered just such qualifications, which its Cartesian spirit seems to demand (see Hintikka 1962, pp. 158–9, and Castañeda 1970). Failures of KK_t are not failures of the KK principle when t does not contribute what 'I' would have contributed to the proposition expressed. For present purposes what matters is that although the two sentence operators 'I know that . . .' and 'The man with a hole in his trousers knows that . . .' are coextensive in the present context, in the sense that one yields a truth when applied to a given sentence if and only if the other does, they do not satisfy the same principles: one iterates in the manner of KK_t but the other does not.

Example 2. Switch from the subject of the propositional attitude to the attitude itself. On a cartoon version of the Stoic idea of wisdom, x is wise if and only if x believes only what x knows, i.e.

(B/K_x) For all P, if x believes that P, x knows that P.

Given the simplifying assumption that knowledge entails belief, the converse of B/K_x holds automatically. Now suppose that Socrates is wise. To ensure that the case is genuinely different from Example 1, assume that Socrates knows perfectly well that he is Socrates. The operators 'Socrates knows that . . .' and 'Socrates believes that . . .' are coextensive, but they need not satisfy the same principles. For example, Socrates knows that if Socrates knows that there is a form of mud then there is a form of mud, simply because Socrates knows that knowledge is factive, and applies that knowledge to the present case. It does not follow that Socrates believes that if Socrates believes that there is a form of mud then there is a form of mud. For although Socrates is wise, he may not believe that he is wise. More generally, the correctness of the schema $O[Op \supset p]$ for the operator 'Socrates knows that' in place of O does not imply its correctness for 'Socrates believes that' in place of O, even granted the coextensiveness of the two operators.

Matters are slightly more complicated than has yet been indicated. For suppose that Socrates always knows by introspection whether he believes something; suppose also that he believes all the logical consequences of his beliefs. Then if he does not believe that P, he believes that Socrates does not believe that P. Since 'Socrates does not believe that P' entails 'If Socrates believes that P then P' (on a truth-functional reading of the conditional), Socrates believes that if Socrates believes that P then P. The same conclusion follows from the assumption that Socrates does believe that P, for P entails 'If Socrates believes that P then P'. Thus if the schema $O[Op \supset p]$ fails for 'Socrates believes that' in place of O, Socrates fails either to know whether he believes something or to believe the logical consequences of what he believes; both kinds of failure are consistent with wisdom as defined above. In particular, the technique of possible worlds semantics can be used to provide a model in which Socrates is wise, and knows that if he knows that P then P, but believes neither that P nor that if he believes that P then P.

For the proof of this, consider a possible worlds model based on a set of three worlds $\{a,b,c\}$. Interpret K ('Socrates knows that') and B ('Socrates believes that') like necessity operators whose accessibility relations are R_K and R_B respectively, where wR_Kx for all worlds w and x, and wR_Bx if and only if either $w = a$ or $x = c$. Since a has both R_K and R_B to all worlds, $K\alpha \equiv B\alpha$ is true at a for any formula α. However, if p is true at c but not at b, then $Bp \supset p$ is false at b, so $B(Bp \supset p)$ is false at a (as is $\sim Bp \supset B\sim Bp$). Since R_K is an equivalence relation, K satisfies all the principles of the modal system S5. Since R_B is serial, connected and transitive, B satisfies all the principles of KDG_14. See Hughes and Cresswell 1984 for the relevant background in modal logic.

Example 3 involves metamathematical provability operators, and their relevance to John Lucas's celebrated attempt to enlist Gödel's work to disprove mechanism. Put racily, Gödel's second incompleteness theorem says that Peano Arithmetic (PA)—or any recursively axiomatizable system extending it—can prove its own consistency only at the price of inconsistency. A little more formally: if it is provable in PA that it is not provable in PA that 0 = 1 then it is provable in PA that 0 = 1. But this formulation still has one occurrence of the informal sentence operator 'It is provable in PA that' inside the scope of another; it therefore treats the operator as coding something in the first-order language of PA, whose intended interpretation concerns numbers, not proofs. Gödel codes each formula α of L_{PA}, the language of PA, by a numeral $\ulcorner\alpha\urcorner$ of L_{PA}, and provability in PA by a formula Bew(x) of L_{PA} with just one free variable. Thus the informal sentence operator is formalized by Bew($\ulcorner...\urcorner$) in L_{PA}. The second incompleteness theorem says that $\vdash_{PA} \sim \text{Bew}(\ulcorner 0 = 1\urcorner)$ only if $\vdash_{PA} 0 = 1$.

Gödel could have used a different coding. What features of his coding justify the informal reading of Bew($\ulcorner...\urcorner$) as 'It is provable in PA that'? If L_{PA} had no intended interpretation, the informal reading would be hard to justify. But L_{PA} does have an intended interpretation, on which a sentence is true if and only if it is true in all standard models of PA. Since PA has standard models, and they are all isomorphic, exactly one of α and $\sim\alpha$ is true (given a classical metalogic). Since we have the notion of truth for L_{PA}, we can stipulate that a one-place sentential operator O in L_{PA} is a *provability operator for PA* just in case for each sentence α of L_{PA}, $O\alpha$ is true if and only if $\vdash_{PA} \alpha$. Bew($\ulcorner \ldots \urcorner$) is a provability operator for PA in this sense (if we count it an operator).

Any two provability operators O_1 and O_2 for PA are coextensive: $O_1\alpha \equiv O_2\alpha$ is true for each sentence α of L_{PA}. But it does not follow that the biconditionals are provable in PA. The unprovability in PA of ~Bew($\ulcorner 0 = 1 \urcorner$) does not imply the unprovability in PA of ~O(0 = 1) for every provability operator O for PA. For example, we can define an operator $O^\Diamond$ in L_{PA} thus:

$$O^\Diamond\alpha =_{df} \mathrm{Bew}(\ulcorner\alpha\urcorner) \;\&\; {\sim}\mathrm{Bew}(\ulcorner{\sim}\alpha\urcorner)$$

Since PA is in fact consistent, the second conjunct of $O^\Diamond\alpha$ is true whenever the first is, so $O^\Diamond\alpha$ is true if and only if Bew($\ulcorner\alpha\urcorner$) is true. Thus $O^\Diamond$ is a provability operator for PA. Yet $\vdash_{PA} \sim O^\Diamond(0 = 1)$, for it is easily checked that $\vdash_{PA}$ Bew($\ulcorner\sim 0 = 1\urcorner$). Thus, if $\perp$ is the negation of a theorem of PA, the result of replacing O . . . throughout the principle O~O$\perp$ by $O^\Diamond$ . . . is true, but the result of replacing it by Bew($\ulcorner \ldots \urcorner$) is false. Coextensive provability operators satisfy different principles in PA.

One might hope to exclude unwanted provability operators by analysing 'It is provable that' as the existential generalization of '*x* is a proof that'. Since the latter notion is effectively decidable, it can be subjected to tighter controls. More specifically, say that Pf[. . .](x) is a *proof operator for PA* just in case there is an effective procedure for assigning to each closed formula α of L_{PA} a formula Pf[α](x) of L_{PA} with just one free variable and coding each proof in an axiomatization of PA by a natural number such that for each natural number n, if n codes a proof of α then $\vdash_{PA}$ Pf[α](**n**) (where **n** is the L_{PA} numeral for n) and otherwise $\vdash_{PA}$ ~Pf[α](**n**). The operator of which Bew($\ulcorner \ldots \urcorner$) is the existential generalization is a proof operator in this sense. The definition ensures that the existential generalization of a proof operator is a provability operator. For let Pf[. . .](x) be a proof operator. If $(\exists x)$ Pf[α](x) is true then Pf[α](**n**) is true for some n, so not $\vdash_{PA}$ ~Pf[α](**n**) (since everything provable in PA is true), so n codes a proof of α (by definition of a proof operator), so $\vdash_{PA} \alpha$. Conversely, if $\vdash_{PA} \alpha$, then some number n codes a proof of α, so $\vdash_{PA}$ Pf[α](**n**) (by definition of a proof operator), so Pf[α](**n**) is true, so $(\exists x)$ Pf[α](x) is true.

It turns out that not even the existential generalization of a proof operator need satisfy the same principles as Bew($\ulcorner \ldots \urcorner$). Example: let

$\text{Bew}(\ulcorner\ldots\urcorner) = (\exists x)\ \text{B}(\ulcorner\ldots\urcorner,x)$. $\text{B}(\ulcorner\ldots\urcorner,x)$ is a proof operator. Define a new operator $\text{Pf}^+[\ldots](x)$ thus

$$\text{Pf}^+[\alpha]\ (x) =_{\text{df}} \text{B}(\ulcorner\alpha\urcorner,x)\ \&\alpha$$

$\text{Pf}^+[\ldots](x)$ is a proof operator, for if n codes a proof of α in PA then $\vdash_{\text{PA}} \alpha$ and $\vdash_{\text{PA}} \text{B}(\ulcorner\alpha\urcorner,\mathbf{n})$ (since $\text{B}(\ulcorner\ldots\urcorner,x)$ is a proof operator), so $\vdash_{\text{PA}} \text{B}(\ulcorner\alpha\urcorner,\mathbf{n})\ \&\ \alpha$; while if n does not code a proof of α in PA then $\vdash_{\text{PA}} {\sim}\text{B}(\ulcorner\alpha\urcorner,\mathbf{n})$, so $\vdash_{\text{PA}} {\sim}(\text{B}\ (\ulcorner\alpha\urcorner,\mathbf{n})\ \&\ \alpha)$. Now define a new provability operator O^+ as the existential generalization of $\text{Pf}^+[\ldots](x)$. Thus where x is not free in α

$$\text{O}^+\alpha =_{\text{df}} (\exists x)\ (\text{B}\ (\ulcorner\alpha\urcorner,x)\ \&\ \alpha)$$

Since x is not free in α, $\text{O}^+\alpha$ is provably equivalent in PA to $\text{Bew}(\ulcorner\alpha\urcorner)\ \&\ \alpha)$. Trivially, $\vdash_{\text{PA}} {\sim}(\text{Bew}(\ulcorner 0 = 1\urcorner)\ \&\ 0 = 1)$, i.e. $\vdash_{\text{PA}} {\sim}\text{O}^+(0 = 1)$. Thus the second incompleteness theorem does not apply to O^+, even though O^+ is the existential generalization of a proof operator. Like $\text{O}^\Diamond$, O^+ satisfies the principle $\text{O}{\sim}\text{O}\bot$, which is not satisfied by the coextensive operator $\text{Bew}(\ulcorner\ldots\urcorner)$.

The existence of deviant metamathematical codings was recognized early on. Kreisel (1953) showed that a problem of Henkin's could be resolved either way by alternative codings. Takeuti (1955) gave a Gentzen-style cut-free formulation of arithmetic in which the natural coding of the unprovability of $0 = 1$ is provable. Feferman (1960) discussed the problem of alternative presentations of formal systems in detail, and demonstrated the provability in PA of a deviant coding of consistency for the usual formulation of PA. Kreisel (1965, pp. 153–154) advocated the use of a notion of the *canonical presentation* of a formal system, a programme developed by Feferman (1982). If one can isolate the features of a coding relevant to the derivation of the second incompleteness theorem for PA, that analysis will carry with it a generalization of the theorem to other systems (the latter is the aspect of the problem with which Gödel was concerned in the footnote he added to the translation of Gödel 1932 in van Heijenoort 1967, discussed in Feferman 1990).

The usual contemporary response appeals to the Hilbert-Bernays-Löb derivability conditions, first isolated by Löb (1955; see e.g. Boolos 1993, p. 16). These are conditions on a *predicate* (as contrasted with an operator) F in the language L of a theory S, given a quotation functor $\ulcorner\ldots\urcorner$ in L from sentences to singular terms. They state that for all sentences α and β of L:

(1) If $\vdash_S \alpha$ then $\vdash_S F(\ulcorner\alpha\urcorner)$
(2) $\vdash_S F(\ulcorner\alpha \supset \beta\urcorner) \supset (F(\ulcorner\alpha\urcorner) \supset F(\ulcorner\beta\urcorner))$
(3) $\vdash_S (F(\ulcorner\alpha\urcorner) \supset F(\ulcorner F(\ulcorner\alpha\urcorner)\urcorner))$

If S satisfies (1)–(3) and a specifiable modicum of arithmetic can be embedded in S, then $\vdash_S \sim F(\ulcorner\bot\urcorner)$ only if $\vdash_S \bot$. This account lacks generality in at least one respect, because it does not make explicit the relevant features of the quotation functor $\ulcorner \ldots \urcorner$. What is really required is just that L should contain a function symbol # such that for all formulas α, β and γ of L:

(4) If the result of substituting $\ulcorner\beta\urcorner$ for all free variables in α is γ then $\vdash_S \ulcorner\alpha\urcorner \# \ulcorner\beta\urcorner = \ulcorner\gamma\urcorner$

No more arithmetic is needed in S. The required background logic is simply:

(5) If α is a truth-functional tautology then $\vdash_S \alpha$
(6) If $\vdash_S \alpha \supset \beta$ and $\vdash_S \alpha$ then $\vdash_S \beta$
(7) $\vdash_S s = t \supset (\alpha(s) \equiv \alpha(t))$

In (7), *s* and *t* are singular terms and the formula $\alpha(t)$ differs from $\alpha(s)$ only in having *t* somewhere where the latter has *s*. If S satisfies (1)–(7), then $\vdash_S \sim F(\ulcorner\bot\urcorner)$ only if $\vdash_S \bot$.

The crucial point is that (4)–(7) permit one to derive Gödel's diagonal lemma for S, that for every one-place predicate F(. . .) in L, there is a closed formula α of L such that $\vdash_S \alpha \equiv F(\ulcorner\alpha\urcorner)$. For the result of substituting the term $\ulcorner F(x \# x)\urcorner$ for the free variable in $F(x\#x)$ is $F(\ulcorner F(x \# x)\urcorner \# \ulcorner F(x \# x)\urcorner)$. Thus by (4)

$$\vdash_S \ulcorner F(x \# x)\urcorner \# \ulcorner F(x \# x)\urcorner = \ulcorner F(\ulcorner F(x \# x)\urcorner \# \ulcorner F(x \# x)\urcorner)\urcorner.$$

(7) then gives the required biconditional:

$$\vdash_S F(\ulcorner F(x \# x)\urcorner \# \ulcorner F(x \# x)\urcorner) \equiv F(\ulcorner F(\ulcorner F(x \# x)\urcorner \# \ulcorner F(x \# x)\urcorner)\urcorner).$$

From the diagonal lemma and (1)–(3) (and, of course, (5)–(6)) one can prove Löb's theorem, that $\vdash_S F(\ulcorner\alpha\urcorner) \supset \alpha$ only if $\vdash_S \alpha$ (for a clear presentation see Boolos 1993, pp. 56–57). The second incompleteness theorem is just the special case of Löb's theorem where $\alpha = \bot$.

What (1)–(7) do not provide are general sufficient conditions on an arbitrary sentential operator O for the derivability of the second incompleteness theorem. It would be naive simply to replace the operator $F(\ulcorner \ldots \urcorner)$ by O . . . in (1)–(3), giving:

(1′) If $\vdash_S \alpha$ then $\vdash_S O\alpha$
(2′) $\vdash_S O(\alpha \supset \beta) \supset (O\alpha \supset O\beta)$
(3′) $\vdash_S O\alpha \supset OO\alpha$

The claim is false that whenever S satisfies (1′)–(3′) and (4)–(7), $\vdash_S \sim O\bot$ only if $\vdash_S \bot$. For if O is a redundant operator, so that $O\alpha$ is provably equivalent to α, then (5)–(6) imply both (1′)–(3′) and the condition that $\vdash_S \sim O\bot$; thus the claim implies that S satisfies (4)–(7) only if S is inconsistent, to which PA is an obvious counterexample. The attempt to prove the claim breaks down at the application of the diagonal lemma. Although the lemma itself is still derivable, it does not provide the required formula α such that $\vdash_S \alpha \equiv \sim O\alpha$ (corresponding to the α such that $\vdash_S \alpha \equiv \sim F(\ulcorner\alpha\urcorner)$ in the proof of the incompleteness theorem), and when O is redundant such an α exists only if S is inconsistent. The redundant operator cannot be factorized into any predicate T and the quotation functor $\ulcorner\ldots\urcorner$ in PA, for otherwise $\alpha \equiv T(\ulcorner\alpha\urcorner)$ would be true for each formula α in L_{PA}, contrary to Tarski's theorem on the undefinability of truth in L_{PA}.

Even when O . . . is coextensive with $Bew(\ulcorner\ldots\urcorner)$, (1′)–(3′) are not sufficient for O . . . to satisfy the second incompleteness theorem when S = PA (which satisfies (4)–(7)). The operator O^+ defined above is a counterexample. That it satisfies (1′)–(3′) is an easy corollary of the fact that $Bew(\ulcorner\ldots\urcorner)$ does. Thus $O^+\alpha$ is not equivalent in PA to the application of a predicate to $\ulcorner\alpha\urcorner$. The source of the trouble is obviously the second conjunct of $O^+\alpha$, namely α. In contrast, $O^\Diamond\alpha$ is equivalent in PA to the application of a predicate to $\ulcorner\alpha\urcorner$, but $O^\Diamond$ does not satisfy (1′).

A systematic account of the principles satisfied by the various provability operators can be given within the framework of propositional modal logic (see Boolos 1993 for the relevant background and results on the modal logic of provability). Its language $L_\Box$ has infinitely many propositional variables $p_0, p_1, p_2, \ldots$, with material implication $\supset$, a falsity constant $\bot$ and the box $\Box$ as the only primitive operators; *p* and *q* will be used as metalinguistic variables ranging over distinct propositional variables of the object-language. For simplicity, assume that $\supset$ and $\bot$ are also in L_{PA}. Given a sentential operator O in L_{PA}, define an O-translation to be any mapping * of formulas of $L_\Box$ to closed formulas of L_{PA} such that for all formulas α and β of $L_\Box$, $*\bot = \bot$, $*(\alpha \supset \beta) = *\alpha \supset *\beta$ and $*\Box\alpha = O*\alpha$. Intuitively, * replaces $\Box$ by O and propositional variables by sentences of L_{PA} in a uniform way. For any formula α of $L_\Box$, say that O *satisfies* α just in case $*\alpha$ is true (in a standard model of PA) for each O-translation *. In effect, the propositional variables of $L_\Box$ are universally quantified over all closed formulas of L_{PA}.

It is convenient to begin by considering provable satisfaction, where O provably satisfies α just in case $*\alpha$ is provable in PA for each O-translation *. By (1)–(3) for Bew and Löb's theorem, Bew(⌜. . .⌝) provably satisfies all theorems of the system GL (= KW) of provability logic ('G' for Gödel, 'L' for Löb); likewise for any system satisfying (1)–(7) in place of PA. To axiomatize GL, take all truth-functional tautologies and all formulas of the forms $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$ and $\Box(\Box\alpha \supset \alpha) \supset \Box\alpha$ (Löb's axiom) as the axioms and *modus ponens* (MP) and necessitation (i.e. the rule that if α is a theorem then so is $\Box\alpha$) as the rules of inference. Solovay (1976) proved the converse for PA: if Bew(⌜. . .⌝) provably satisfies α, then $\vdash_{GL} \alpha$.

Now axiomatize an extension GLS of GL ('S' for Solovay) by taking all theorems of GL and all formulas of the form $\Box\alpha \supset \alpha$ as the axioms and MP as the only rule of inference (since GL is decidable, the axiomatization of GLS is recursive). Since everything provable in PA is true, Bew(⌜. . .⌝) satisfies every theorem of GLS. GLS is not closed under necessitation, for $\sim\Box\bot$ is a theorem but $\Box\sim\Box\bot$ is not (its negation is); although ~Bew(⌜⊥⌝) is true, it is not provable in PA. Again, Solovay (1976) proved the converse result: if Bew(⌜. . .⌝) satisfies α, then $\vdash_{GLS} \alpha$. Thus the theorems of GLS are exactly the principles that Bew(⌜. . .⌝) satisfies.

As for the deviant provability operator O^+, the principles that it satisfies constitute the modal logic Grz (= S4Grz = KGrz, named after Andrzej Grzegorczyk), which can be axiomatized by taking all truth-functional tautologies and all formulas of the forms $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$ and $\Box(\Box(\alpha \supset \Box\alpha) \supset \alpha) \supset \alpha$ (Grzegorczyk's axiom) as the axioms and MP and necessitation as the rules of inference (see Boolos 1993, pp. 155–164, for the relevant results about Grz). Every formula of the form $\Box\alpha \supset \alpha$ is derivable in Grz, and is subject to necessitation. In particular, $\Box(\Box\bot \supset \bot)$ is derivable. Thus $\Box\sim\Box\bot$ is a theorem of Grz but not of GLS, while $\sim\Box\sim\Box\bot$ is a theorem of GLS but not of Grz. O^+ . . . but not Bew(⌜. . .⌝) satisfies $\Box\sim\Box\bot$; Bew(⌜. . .⌝) but not O^+ . . . satisfies $\sim\Box\sim\Box\bot$.

The phenomenon is observable within GLS itself. Let $\blacksquare\alpha$ abbreviate $\Box\alpha$ & α. Then $\blacksquare\alpha \equiv \Box\alpha$ is a theorem of GLS for each formula α of $L_\Box$. Nevertheless, $\blacksquare\sim\blacksquare\bot$ but not $\Box\sim\Box\bot$ is a theorem of GLS, as is $\sim\Box\sim\Box\bot$ but not $\sim\blacksquare\sim\blacksquare\bot$. Unless α is a theorem of GLS (indeed, of

GL), $\Box(\blacksquare\alpha \equiv \Box\alpha)$ is not a theorem of GLS, so $\blacksquare$ and $\Box$ are not interchangeable within modal contexts in GLS.

Such possibilities cast doubt on attempts to use Gödel's incompleteness theorems to show that humans are not Turing machines (see Lucas 1961 and Penrose 1989, 1994). One argument runs something like this: if I am a Turing machine, then the set of mathematical and metamathematical sentences to which I eventually assent (given enough time and paper) is recursively enumerable, and therefore (it can be shown) recursively axiomatizable; since I assent to a modicum of logic and arithmetic, the second incompleteness theorem applies; but I assent to my own consistency, so it would follow that I am inconsistent, which I am not; therefore I am not a Turing machine. The argument is obviously very loose as it stands. The supposition that I am a Turing machine is quite unclear until my relevant states have been specified; the inference to the recursive enumerability of the set of sentences to which I eventually assent is also dubious until the notion of assent (is it revocable?) and the role of perceptual input have been clarified (a Turing machine that does not halt can write an infinite sequence of symbols that is not recursively enumerable on its tape, if such a sequence is provided for it to read). For the sake of argument, however, grant that these loose ends have been tidied up, and that a sense has been given to 'I am a Turing machine' from which it does follow that the set of sentences to which I eventually assent is recursively enumerable. Call the set 'my system'. It can be recursively axiomatized; note that there is no automatic guarantee that everything provable in my system is true. A standard objection is then that no good reason has been provided to believe that my system is consistent (see Putnam 1975, p. 366). This objection is unsatisfying, however, for the argument from the incompleteness theorem seems to show that I *must* be inconsistent if I eventually assent to my own consistency; yet it seems that even if I do so I *might* still be consistent.

A more significant fallacy in the argument is its neglect of the mode of presentation under which I assent to the consistency of my system. The Gödelian consistency sentence makes coded reference to my system by descriptive means: it codes a description of the axioms and inference rules of an axiomatized system that is in fact my own. The consistency sentence to which it seems that I can consistently assent makes reference to my system by indexical means, as my system. There is a corresponding difference between the descriptive provability operator 'A Turing machine with machine table T can prove that . . .', where

'T' abbreviates an appropriate description, and the indexical provability operator 'I can prove that . . .', in a context in which I am (i.e. realize) a Turing machine with machine table T (where 'can prove' abbreviates something like 'has a proof in its system'). Although the two operators are extensionally equivalent, they may satisfy different principles under embedding, if I cannot prove that I am a Turing machine with machine table T. As Paul Benacerraf pointed out long ago, I might be a Turing machine without knowing which Turing machine I am (Benacerraf 1967). In particular, if I can prove that I cannot prove that 0=1, then it follows that a Turing machine with machine table T can prove that I cannot prove that 0=1, but it does not follow that a Turing machine with machine table T can prove that a Turing machine with machine table T cannot prove that 0=1, just as $\mathrm{Bew}(\ulcorner \sim\mathrm{Bew}(\ulcorner \bot \urcorner)\urcorner)$ was seen above not to follow from $O{\sim}O\bot$ for an arbitrary provability operator O in PA. The present provability operators belong to a more expressive language than L_{PA}, but the logical point remains the same.

The objection applies even if the argument appeals to the first incompleteness theorem rather than the second. Unlike the latter, the former is not sensitive to differences between coextensive provability operators, because its statement—that in any consistent recursively axiomatizable extension of PA some sentence is neither provable nor disprovable—does not involve embedded occurrences of the provability operator. However, the anti-mechanistic argument claims that I can prove the supposedly undecidable sentence G for my system; but the purported proof requires the consistency [ω-consistency] of my system as a lemma, so the same objection applies.

I might even know that I am a Turing machine without knowing which Turing machine I am (see Reinhardt 1986). In effect, such a possibility is addressed by Penrose (1994, pp. 132–137). Penrose offers only plausibility arguments against it. He finds it unlikely that I could prove each sentence enumerated by a given algorithm without being able to prove the generalization that every sentence enumerated by the algorithm is true, but such claims of likelihood are unreliable in an area so full of surprises. The execution of the algorithm may involve me in a progressively more complex thought procedure for each sentence. There is no guarantee that I can discern the common pattern in these procedures; even if I am told what it is, there is no guarantee that I will be convinced that it always yields correct results, even though, in each particular case, running through the procedure convinces me that that particular result is correct.

The point essentially concerns provability operators, not provability predicates. If Bew_I is an indexical provability predicate, S is my system, and quotation marks are used in the ordinary way, then Bew_I would presumably satisfy the Hilbert-Bernays-Löb derivability conditions (1)–(3). My system presumably satisfies (5)–(7) and provides the elementary syntax needed for (4). However, when diagonalization is applied to formulas involving Bew_I in attempt to prove the second incompleteness theorem for Bew_I, it yields an intuitively paradoxical sentence whose intended reading is something like 'I cannot prove this sentence'; in contrast, Gödel's use of diagonalization yields a sentence whose intended reading makes an ordinary statement about numbers. Not surprisingly, reasoning with self-referential sentences involving Bew_I does lead to inconsistency (if consistency is claimed). The use of an indexical provability operator yields no such paradox, because it is not susceptible to diagonalization (these remarks address a worry developed in the appendix to Benacerraf 1967). For example, as noted above, the deviant provability operator O^+ cannot be factorized into a provability predicate and a quotation functor.

The deviant provability operators introduced by Kreisel, Takeuti and Feferman are defined from provability predicates, and so do not satisfy (1)–(3), otherwise the second incompleteness theorem would be derivable for them. They are therefore less threatening than O^+ to the anti-mechanistic argument, since they seem to differ so markedly from the indexical provability operator in the satisfaction of principles. Nevertheless, it would be interesting to know what deviant systems of modal logic are constituted by the principles that they do satisfy.

Example 3 combines elements of Examples 1 and 2. Like Example 1, it exploits the contrast between first-personal and descriptive self-presentation. Like Example 2, it concerns the principle expressed in $\text{L}_\Box$ as $\Box(\Box p \supset p)$. For present purposes, the principle at issue in Example 1, expressed in $\text{L}_\Box$ as $\Box p \supset \Box\Box p$, may be treated as unproblematic for both the descriptive and the indexical provability operators, although for somewhat different reasons in the two cases (it is also a theorem of both GLS and Grz). The analogy with Example 2 is strengthened if my system is defined as the set of sentences that express my mathematical and metamathematical *knowledge*, i.e. that are provable on the basis of my knowledge-yielding axiom schemata and inference rules. Since knowledge cannot be false, I can give a trivial consistency proof for my system, so defined; it will be necessary to use a knowledge operator rather than predicate to avoid Liar-like paradoxes. Some of the themes in this example are developed at more length in Shin and Williamson 1994 and Williamson 1998.

Example 4 involves metaphysical rather than epistemic or doxastic modalities. Nathan Salmon has argued that there are counterexamples concerning the possible constitution of artifacts to the analogue of the KK principle for metaphysical necessity, i.e. to the 4 axiom $\Box p \supset \Box\Box p$ (equivalently, $\Diamond\Diamond p \supset \Diamond p$) familiar from the modal system S4 (= KT4; $\Diamond$ is throughout a metalogical abbreviation for $\sim\Box\sim$). The soundness of his argument is not at issue here. (See his 1989 and, for some objections, Williamson 1990, pp. 126–143, to which Salmon 1993 is a reply. A reply to this must await another occasion.) What is relevant is one part of his response to those who find the modal system S5 (= KT5) compelling as a logic for metaphysical necessity. S5 adds to the system S4 the other half of what is required for necessity always to be a non-contingent feature, the 5 (= E) axiom $\sim\Box p \supset \Box\sim\Box p$ (equivalently, $\Diamond\Box p \supset \Box p$). S5 is the logic of necessity as truth in all possible worlds, on the assumption that it is not contingent what worlds are possible. Strictly, S5 also embodies the assumption that the actual world is possible: it contains the T axiom $\Box p \supset p$ (equivalently, $p \supset \Diamond p$). A logic stronger than S5 would result if and only if a specific finite upper bound to the number of possible worlds were imposed in addition to the T and 5 axioms (for the 'only if' direction, use Scroggs 1951; the 'if' direction is routine). Such a bound is surely incorrect when $\Box$ is read as metaphysical necessity: for each natural number n, it is metaphysically possible for there to be exactly n planets.

Salmon suggests that those who find S5 compelling as a logic for metaphysical necessity confuse necessity with actual necessity, i.e. with the property of being necessary in this, the actual world (1989, p. 30). For even if it is contingent whether it is necessary that P, it is non-contingent whether it is actually necessary that P, simply because it is non-contingent whether *anything* is actually the case. What is the case relative to a given world is not itself a world-relative matter. The logic of actual necessity is, uncontentiously, S5; necessity is easily confused with actual necessity, for the two are in some sense a priori equivalent (see Davies and Humberstone 1980 for some relevant background on 'actually').

To be more precise: First expand $L_\Box$ to a language $L_{\Box@}$ by adding a new primitive operator @, to be read as 'actually'. Axiomatize a preliminary bimodal system KT@ for necessity and actuality by taking all truth-functional tautologies and all formulas of the forms $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$, $\Box\alpha \supset \alpha$, $@(\alpha \supset \beta) \supset (@\alpha \supset @\beta)$, $@{\sim}a \equiv {\sim}@\alpha$, $@\alpha \supset @@\alpha$ and $@\alpha \supset \Box@\alpha$ as the axioms and MP, necessitation and actualization (i.e. the rule that if α is a theorem then so is @α) as the rules of inference. KT@ is a conservative extension of the familiar modal system KT (= T), in the sense that the theorems of KT@ in which @ does not occur are just the theorems of KT. To axiomatize KT, drop all the axiom schemata in which @ occurs and the rule of actualization from the axiomatization of KT@. KT and KT@ correspond to an interpretation of □ on which □α is true at a world *w* just in case α is true at all worlds accessible from *w*, where the only constraint on accessibility is that each world is to be accessible from itself. The theorems of KT are uncontentious when □ is read as metaphysical necessity. KT@ corresponds to an interpretation of @ on which @α is true at a world *w* just in case α is true at a fixed world *a*. However, its theorems are the formulas guaranteed to be true at *all* worlds; thus *a* is in no way privileged as the actual world in the semantics of KT@. It is privileged in the system KT@S, whose theorems are the formulas guaranteed to be true at *a* itself. KT@S is an extension of KT@ ('S' for Solovay, since the construction of KT@S from KT@ resembles that of GLS from GL). To axiomatize KT@S, take all theorems of KT@ and all formulas of the form $@\alpha \supset \alpha$ as the axioms and MP as the rule of inference (since KT@ is decidable, the axiomatization is recursive). All formulas of the form $\alpha \supset @\alpha$ are derivable in KT@S. KT@S is not closed under necessitation, for $@p \supset p$ is a theorem but $\Box(@p \supset p)$ is not. $@p \supset p$ is guaranteed to hold at the actual world, but not at all possible worlds.

A different way to define an operator that satisfies the S4 axiom out of one that does not is to let □*α be the infinite conjunction of α, □α, □□α, ... The accessibility relation for □* is the reflexive ancestral, which is automatically transitive, of the accessibility relation for □. This construction does not serve Salmon's purposes, however, for he also denies the logical truth of the B axiom $p \supset \Box\Diamond p$, which is derivable in S5 (it is not even entirely clear why Salmon should regard the T axiom as a logical truth; its obviousness alone does not qualify it as such). The B axiom is not guaranteed to hold for □*: it corresponds to the symmetry of the accessibility relation, and the reflexive ancestral of a non-symmetric relation may itself be non-symmetric.

What must be shown is that S5 is the logic of @□ in KT@S. More precisely, it must be shown that if α is any formula of $L_\Box$, and ^α is the result of replacing each occurrence of $\Box$ in α by $@\Box$, then α is a theorem of S5 if and only if $^\wedge\alpha$ is a theorem of KT@S. To axiomatize S5, add all formulas of the form $\sim\Box\alpha \supset \Box\sim\Box\alpha$ to the axiomatization of KT above. One can show by induction on the length of the proof of α that if α is a theorem of S5 then $^\wedge\Box\alpha$ is a theorem of KT@, and therefore of KT@S; but $^\wedge\Box\alpha \supset {}^\wedge\alpha$ is a theorem of KT@S, so $^\wedge\alpha$ is too. (One cannot argue directly that $^\wedge\alpha$ is a theorem of KT@S, because KT@S is not closed under necessitation and the induction step for the rule of necessitation in S5 would break down. And although KT@ is closed under necessitation, it does not follow that $^\wedge\alpha$ is a theorem of KT@, e.g. when $\alpha = \Box p \supset p$. But although $@\Box p \supset p$ is not a theorem of KT@, $@\Box(@\Box p \supset p)$ is; hence the indirect strategy used.) Conversely, if α is not a theorem of S5, then by a standard completeness theorem for S5 there is a possible worlds model in which α is false at a world w and the accessibility relation holds universally; this can easily be turned into a model of KT@S by the selection of w as the actual world; α is still false at w in the new model; but α and $^\wedge\alpha$ are true at the same worlds in this model (since the worlds accessible from a given world are those accessible from the actual world); thus $^\wedge\alpha$ is false at the actual world in the model of KT@S, so α is not a theorem of KT@S. (Some related and more general results are discussed in §5.) Although the operators $\Box$ and $@\Box$ are coextensive in the set of all models of KT@S, because $\Box\alpha \equiv @\Box\alpha$ is a theorem for every formula α, they satisfy different principles. For example, $@\Box p \supset @\Box@\Box p$ and $\sim@\Box p \supset @\Box\sim@\Box p$ are theorems but $\Box p \supset \Box\Box p$ and $\sim\Box p \supset \Box\sim\Box p$ are not. Correspondingly, $\Box(p \equiv @\Box p)$ is not a theorem of KT@S.

The relation between $\Box$ and $@\Box$ might be thought to be just a complex variation on the relation between the trivial identity operator and @, which would make the occurrence of $\Box$ in the example inessential. But in a language in which @ is the only primitive operator other than the usual truth-functional ones, with a semantics on which @ is coextensive with the identity operator, the two operators satisfy exactly the same principles. Another operator is required to bring out their latent difference.

Example 5. A similar phenomenon can arise even when the coextensive operators are in some sense synonymous. Consider a language L with atomic sentences q_0, q_1, q_2, . . ., the truth-functional operators $\supset$ and $\perp$, and two further operators O_1 and O_2. Let models be ordered pairs $\langle X,Y\rangle$, where X is a set of atomic sentences and Y a set of sentences of L. Define truth at a model ($\Vdash$) recursively, as follows:

$\langle X,Y\rangle \Vdash q$	just in case $q \in X$
$\langle X,Y\rangle \Vdash \bot$	in no case
$\langle X,Y\rangle \Vdash \alpha \supset \beta$	just in case either not $\langle X,Y\rangle \Vdash \alpha$ or $\langle X,Y\rangle \Vdash \beta$
$\langle X,Y\rangle \Vdash O_1\alpha$	just in case $\alpha \in Y$
$\langle X,Y\rangle \Vdash O_2\alpha$	just in case $\alpha \in Y$

Since the clauses for O_1 and O_2 have exactly the same right-hand side, one might regard O_1 and O_2 as synonymous, and conclude that they are interchangeable in all contexts *salva veritate*. Any formula of the form $O_1\alpha \equiv O_2\alpha$ is indeed true at any model, so O_1 and O_2 are coextensive. However, if $O_1q \in Y$ but $O_2q \notin Y$, then $\langle X,Y\rangle \Vdash O_1O_1q$ but not $\langle X,Y\rangle \Vdash O_2O_2q$. It does not immediately follow that O_1 and O_2 satisfy different principles with respect to the set of all models; indeed, if no further restrictions are imposed on the models, the symmetrical treatment of O_1 and O_2 will ensure that they satisfy the same principles. But if every model $\langle X,Y\rangle$ meets the condition that $O_1\alpha \in Y$ and $O_2\alpha \notin Y$, then O_1 will satisfy the principle OOp while O_2 satisfies the principle $\sim OOp$.

Of course, if the condition that $O_1\alpha \in Y$ and $O_2\alpha \notin Y$ followed from the semantics of O_1 and O_2, then they would not be synonymous. But the asymmetry could arise in other ways. For example, if the models are just contexts of utterance in the actual world, then the asymmetry might arise from mundane facts.

For any particular application of the formal semantics, some explanation would naturally be needed of the significance of the second element Y of a model. If the explanation permits a model $\langle X,Y\rangle$ for which $O_1q \in Y$ but $O_2q \notin Y$, so $\langle X,Y\rangle \Vdash O_1O_1q$ and $\langle X,Y\rangle \Vdash O_2O_1q$ but neither $\langle X,Y\rangle \Vdash O_1O_2q$ nor $\langle X,Y\rangle \Vdash O_2O_2q$, then it is natural to conclude that O_1 and O_2 have quotational features, since each yields two sentences of different truth-value when applied to the synonymous sentences O_1q and O_2q. For example, Y might contain all sentences uttered at a given time and place. Thus $O_1\alpha$ and $O_1\beta$ may differ in meaning when α and β have the same meaning; likewise for O_2. For this reason, one may be unwilling to class O_1 and O_2 as *propositional* operators; nevertheless, they remain well-defined operators on *sentences* (for a related discussion see Cresswell 1983). Note that Examples 1–4 do not depend on quotationality.

2. *A generalization*

It is time to give a precise general statement of the phenomenon exemplified in the previous section. Start with an interpreted language L, assumed to contain at least the truth-functional operators $\supset$ and $\perp$. Suppose that the sentences of L are evaluated as true or false at members of a set I, known as *indices*. They can be possible worlds, contexts of utterance, ordered pairs of worlds and contexts, or anything else; their role is simply to give the account generality. When the generality is not wanted, I can have just one member. In Examples 1–3 one can take the only index to be, respectively, a context of utterance in which I (the speaker) am the man with a hole in his pocket, a world in which Socrates is wise, and a context of utterance in which I am a Turing machine with machine table T. $\perp$ and $\supset$ will be assumed to behave according to their usual truth-tables at each index i: $\perp$ is not true at i and $\alpha \supset \beta$ is true at i if and only if either α is not true at i or β is true at i, for all sentences α and β of L.

If O_1 and O_2 are sentence operators of L, say that O_1 and O_2 are *coextensive* just in case, for every sentence α of L and every index i, $O_1\alpha$ is true at i if and only if $O_2\alpha$ is true at i (since the indices will sometimes be possible worlds, it would be just as appropriate to speak of cointensiveness). The indices are not required to include all the points at which the semantics of L evaluates its sentences; rather, they are a subset of such points in which we happen to be interested. For example, in Example 4 the semantics of the modal language L with the 'actually' operator may be two-dimensional, evaluating sentences at pairs of worlds $<w,x>$, of which w is the context of utterance and x the context of evaluation proper; thus $@\alpha$ is true at $<w,x>$ if and only if α is true at $<w,w>$. If we are interested in the evaluation of utterances at the context in which they are made, we can take the indices to be the pairs $<w,w>$ 'on the diagonal', with respect to which $\Box$ and $@\Box$ are coextensive. They are not coextensive if all pairs are taken as indices.

As in Example 3, a *principle* is a formula of $L_\Box$. If O is a sentence operator of L, an O-*translation* is a mapping * from formulas of $L_\Box$ to sentences of L such that for all formulas α and β of $L_\Box$, $*(\alpha \supset \beta) = *\alpha \supset *\beta$, $*\perp = \perp$ and $*\Box\alpha = O*\alpha$. O *satisfies* a principle α just in case $*\alpha$ is true at every index i for every O-translation *. The phenomenon of interest is just that not all coextensive operators satisfy the same principles.

The choice of $L_\Box$ to record principles is purely notational: it does not assume the operator O to have anything in common with a necessity operator beyond being a one-place sentential operator. The modal language has the merit of familiarity; many of its formulas and sets of formulas already have names.

Coextensiveness and satisfaction of the same principles are conditions at the same level of generality; they are defined by quantification over the same set of indices. If coextensiveness had been defined by quantification only over actual items of some kind, and satisfaction of the same principles by quantification over all logically possible items of that kind, then the failure of the former to imply the latter would not have been surprising; but there is no such disparity in the definitions above.

The principles over which coextensive operators can differ always involve some self-embedding of $\Box$. If a *first-degree* principle is a formula of $L_\Box$ in which no occurrence of $\Box$ lies within the scope of another, then coextensive operators O_1 and O_2 satisfy exactly the same first-degree principles. For let $*_1$ be any O_1-translation, and let $*_2$ be the O_2-translation that agrees with $*_1$ on propositional variables. If α does not contain $\Box$, then $*_1\alpha = *_2\alpha$, so $*_1\Box\alpha$ ($= O_1 *_1\alpha$) and $*_2\Box\alpha$ ($= O_2 *_2\alpha = O_2 *_1\alpha$) are true at the same indices by coextensiveness, so $*_1\alpha$ is true at the same indices as $*_2\alpha$ for any first-degree formula α. Thus if O_2 satisfies a first degree formula, then so too does O_1. The converse follows by a parallel argument. But even the simplest kind of self-embedding, as in $\Box\Box p$, permits coextensive operators to differ in satisfaction of principles.

One limit to the phenomenon of coextensiveness without satisfaction of the same principles is that it cannot involve truth-functional operators. If O is truth-functional, then any operator coextensive with O satisfies the same principles as O. In the present context, it is natural to stipulate that an operator O in L is *truth-functional* just in case for all sentences α and β of L and all indices i and j, if the truth-value of α at i is the truth-value of β at j then the truth-value of $O\alpha$ at i is the truth-value of $O\beta$ at j. However, the limit to the phenomenon is more general than that. Say that O is *extensional* just in case for all sentences α and β of L and all indices i, if the truth-value of α at i is the truth-value of β at i then the truth-value of $O\alpha$ at i is the truth-value of $O\beta$ at i, in other words, just in case O satisfies the principle $(p \equiv q) \supset (\Box p \equiv \Box q)$. All truth-functional operators are extensional, but not all extensional operators are truth-functional. For example, let O . . . be the material biconditional '. . . if and only if it is raining', and let I contain both an index i at which it is raining and an index j at which it is not raining; then, where $\top$ is $\bot \supset \bot$, $O\top$ is true at i and false at j, even though $\top$ has the same truth-value at i and j, so O is not truth-functional; nevertheless, O is obviously extensional. The example also shows that no principle characterizes truth-functionality in the way that $(p \equiv q) \supset (\Box p \equiv \Box q)$ characterizes extensionality: if there were such a principle, O would satisfy it too. (For a thorough discussion of a distinction closely related to that between truth-functionality and extensionality see Humberstone 1986. The present phenomenon is closely related to the fact, stated in his terminology, that if $O_1\alpha$ and $O_2\alpha$ always imply each other in a logic, it does not follow that each of O_1 and O_2 is a subconnective of the other. He points out that if $O_1\alpha$ always implies $O_2\alpha$, it does not follow that O_1 is a subconnective of O_2 (1986, p. 41). Note that his notions, unlike the present ones, are defined relative to a logic.) One limit to the phenomenon is this: if O is extensional, then any operator coextensive with O satisfies the same principles as O.

To establish the point, argue as follows. Let O be any extensional operator in L, and O′ any operator in L coextensive with O. Let * be any O-translation, and *′ the O′-translation such that $*p = *'p$ for each propositional variable p. It can be shown by induction on the complexity of $\alpha \in L_\Box$ that for any index i, the truth-value of $*\alpha$ at i is the truth-value of $*'\alpha$ at i. The only interesting case is the induction step for $\Box$. Suppose that the truth-value of $*\alpha$ at i is the truth-value of $*'\alpha$ at i (the induction hypothesis). Then the truth-value of $*\Box\alpha$ at i is the truth-value of $O*\alpha$ at i (since * is an O-translation), which is the truth-value of $O*'\alpha$ at i (from the induction hypothesis, since O is extensional), which is the truth-value of $O'*'\alpha$ at i (since O and O′ are coextensive), which is the truth-value of $*'\Box\alpha$ at i (since *′ is a O′-translation), as required. It follows that if O′ satisfies the principle α, then so too does O′. Conversely, by a parallel argument, if O satisfies α, then so too does O′. Thus O and O′ satisfy the same principles, QED.

Perhaps a more intuitive way of making the point is this. Evidently, any operator coextensive with an extensional operator is itself extensional. Now at any given index a (singulary) extensional operator behaves according to one of the four singulary truth-tables; which one is fixed by the truth-values of the results of applying it to $\top$ and $\bot$, which are the same from one of two coextensive operators to the other. Thus coextensive extensional operators behave according to the same truth-table at any given index. The principles that an extensional operator satisfies are fixed by the set of truth-tables according to which it behaves at at least one index; thus coextensive extensional operators satisfy the same principles. Since the set of singulary truth-tables has just 16 subsets, there are just 16 possibilities for the set of principles that an extensional operator satisfies (or 15 if the set of indices is required to be non-empty. Humberstone 1995 arrives at an answer greater than 16 to what may look like the same question by individuating operators according to consequence relations rather than in the present semantic manner.)

The phenomenon of coextensiveness without satisfaction of the same principles can be generalized to n-ary operators where $n > 1$, although a more complex language than $L_\Box$ would be needed to keep track of it. For example, one could seek coextensive conditionals that satisfy different principles. The preceding remarks show that the phenomenon arises only for non-truth-functional conditionals. Example 4 provides an obvious example, for strict implication ($\Box(p \supset q)$) and actualized strict implication ($@\Box(p \supset q)$) are coextensive in KT@S without satisfying the same principles. Since necessity is definable in terms of strict implication, this is just a notational variant of Example 4. Necessity can equally well be defined in terms of many other conditionals $\Rightarrow$ (e.g. the counterfactual conditional) by an equivalence between $\Box\alpha$ and $\sim\alpha \Rightarrow \alpha$, so similar examples can be manufactured for them too. In such cases, if $p \Rightarrow q$ can be interpreted as contingently true (as it can when $\Rightarrow$ is the counterfactual conditional), then the principle $(p \Rightarrow q) \supset (\sim(p \Rightarrow q) \Rightarrow (p \Rightarrow q))$ will fail, while the corresponding principle holds for the actualized variant of $\Rightarrow$, even if the logic of necessity is assumed to be S5. Similar remarks apply to the principle $(p \Rightarrow q) \Rightarrow (r \Rightarrow (p \Rightarrow q))$. It would be interesting to have examples in which the principles at stake have been subject to more controversy in discussion of conditionals (see Dorothy Edgington's contribution to this volume).

3. *Systems of principles*

To investigate the phenomenon of coextensiveness without satisfaction of the same principles more deeply, one must look at the systems of modal logic constituted by the principles that given operators satisfy. Some terminology will help. A *(mono)modal system* is a subset of the language (set of formulas) $L_\Box$ containing all truth-functional tautologies and closed under uniform substitution (US) and MP. A *normal* (mono)-modal system is a (mono)modal system containing the formula $\Box(p \supset q) \supset (\Box p \supset \Box q)$ and closed under necessitation. In effect, normality means that $\Box$ is closed under deductive consequence in the system. These definitions are extended to polymodal languages in the obvious way, e.g. if the language contains the primitive modal operators $\Box_1, \ldots \Box_n$ then normality requires the system to contain the formula $\Box_j(p \supset q) \supset (\Box_j p \supset \Box_j q)$ and to be closed under necessitation for $\Box_j$ for each j from 1 to n. If S is a modal system, write $\vdash_S \alpha$ for $\alpha \in$ S. A set of formulas in the language of S is *S-consistent* just in case it has no finite subset $\{\alpha_1, \ldots, \alpha_k\}$ such that $\vdash_S \sim(\alpha_1 \,\&\, \ldots \,\&\, \alpha_k)$.

The set of principles that an operator O satisfies is automatically closed under US. For given any substitution σ and O-translation *, the mapping of each formula α of $L_\Box$ to $*\sigma\alpha$ is also an O-translation; thus if O satisfies α, then $*\sigma\alpha$ is true at every index; thus every O-translation of $\sigma\alpha$ is true at every index, so O satisfies $\sigma\alpha$. Since the primitive truth-functional operators of $L_\Box$ ($\bot$ and $\supset$) behave in L according to their usual truth-tables at every index and commute with every O-translation, O satisfies all truth-functional tautologies in $L_\Box$, and if O satisfies $\alpha \supset \beta$ and α then it satisfies β. Thus the principles that an operator satisfies always constitute a modal system.

Conversely, any modal system S is the set of all principles satisfied by some operator O for some index set I. To establish this, let the base language L be $L_\Box$ itself, O be $\Box$, I be the set of all maximal S-consistent subsets of $L_\Box$, and α be true at i ($\in$ I) just in case $\alpha \in i$. Since S contains all truth-functional tautologies and is closed under MP, this gives the right truth-conditions for $\bot$ and $\alpha \supset \beta$. $\Box$-translations are simply uniform substitutions, so every $\Box$-translation of α is true at every index if and only if every substitution instance of α belongs to every maximal S-consistent set, i.e. if and only if every substitution instance of α is a theorem of S, i.e. if and only if α is a theorem of S (since S is closed under US). Thus the theorems of S are exactly the principles that $\Box$ satisfies.

In many cases, the principles that an operator satisfies do not constitute a normal modal system. Normal systems are especially tractable, however, and therefore make a good point of departure.

Some logical questions can now be raised. How widespread is the phenomenon of coextensiveness without satisfaction of the same principles? More specifically: given the principles that two operators O_1 and O_2 satisfy, when is it consistent to suppose that O_1 and O_2 are coextensive? The question can be taken in more than one way. One reading is this: given two monomodal systems S_1 and S_2, when is there a language L with a non-empty index set I and coextensive operators O_1 and O_2 such that O_1 satisfies all the principles in S_1 and O_2 satisfies all the principles in S_2? The index set is required to be non-empty in order not to make the answer trivially 'Always', since any operator vacuously satisfies any principle if the index set is empty. On this reading of the question, O_1 may also satisfy some principles not in S_1 and O_2 may also satisfy some principles not in S_2. It follows that this reading does not address the phenomenon at issue very well. For even if S_1 and S_2 are distinct systems, and O_1 satisfies every principle in S_1 and O_2 satisfies every principle in S_2, the two operators may satisfy exactly the same principles.

The extent of the problem can be illustrated as follows. Let S_1 and S_2 be any two normal systems. If $\Diamond\top$ ($\top$ a tautology) is a theorem of neither S_1 nor S_2, then every principle of both S_1 and S_2 is satisfied by any operator O such that, for each index *i* and sentence α of L, $O\alpha$ is true at *i*. For example, $O\alpha$ could be $\alpha \supset \alpha$. Proof: Let Ver be the smallest normal system of which $\Box\bot$ is a theorem. If $\Diamond\top$ is not a theorem of a normal system S, then Ver is an extension of S. For let $S^+ = \{\alpha \in L_\Box : \vdash_S \Box\bot \supset \alpha\}$. S^+ is easily shown to be a consistent normal extension of Ver, and therefore to be Ver (which is Post-complete). If $O\alpha$ is true at every index for every $\alpha \in L_\Box$, then O satisfies every principle in Ver, and therefore every principle in S, of which S^+ is an extension.

Similarly, if both S_1 and S_2 have consistent normal extensions of which $\Diamond\top$ is a theorem, then every every principle of both S_1 and S_2 is satisfied by any operator O such that, for each index *i* and sentence α of L, $O\alpha$ is true at *i* if and only if α is true at *i*. For example, $O\alpha$ could be α itself. Proof: Let Triv be the smallest normal system of which $\Box p \equiv p$ is a theorem. If a system S has a consistent normal extension S′ of which $\Diamond\top$ is a theorem, then Triv is an extension of S. For let σ be the substitution such that $\sigma p = \top$ for each propositional variable *p*, and put $S^+ = \{\alpha \in L_\Box : \vdash_S \sigma\tau\alpha$ for every substitution $\tau\}$. S^+ is easily shown to be a consistent normal extension of Triv, and therefore to be Triv (which is Post-complete). If $O\alpha$ is true at an index if and only if α is, for every $\alpha \in L_\Box$, then O satisfies every principle in Triv, and therefore every principle in S′, of which S^+ is an extension, and therefore every principle in S.

Each result covers a wide class of cases, in all of which one can choose O_1 and O_2 to be identical, so that they automatically satisfy the same principles. Thus, on this reading, the question does not elicit coextensiveness without satisfaction of the same principles.

There are, however, some pairs of coextensive operators O_1 and O_2 such that O_1 satisfies every principle in a normal system S_1, O_2 satisfies every principle in a normal system S_2, and no operator could consistently satisfy every principle in both S_1 and S_2. For example, let S_1 be the smallest normal system of which $\Box\Box\bot$ is a theorem, and S_2 be KD (=D), the smallest normal system of which $\Diamond\top$ is a theorem. Since $\sim\Box\Box\bot$ is a theorem of KD, no operator could satisfy every principle in both S_1 and S_2 with respect to a non-empty set of indices. Nevertheless, define two operators as follows. There is only one index, so this parameter can be suppressed. Truth and an auxiliary notion of pseudo-truth are defined by simultaneous recursion. Take the truth or falsity of atomic sentences as given. An atomic sentence is pseudo-true just in case it is true. Both truth and pseudo-truth behave in the usual way with respect to $\bot$ and $\supset$. $O_1\alpha$ is true just in case α is pseudo-true; $O_2\alpha$ is true just in case α is pseudo-true; $O_1\alpha$ is pseudo-true in any case; $O_2\alpha$ is pseudo-true just in case α is pseudo-true. The semantics works just like a possible worlds semantics for O_1 and O_2 as necessity-like operators whose accessibility relations are $\{<i,j>\}$ and $\{<i,j>,<j,j>\}$ respectively on a set of worlds $\{i,j\}$, where truth is truth at the actual world i and pseudo-truth is truth at j. Validity is taken as truth at the actual world for any truth-assignment to atomic formulas at worlds. In this semantics it is stipulated that only the actual world is an index. By induction on the length of proofs in the natural axiomatizations of S_1 and S_2 with MP and necessitation as the only rules of inference, it is easy to check that every O_1-translation of every theorem of S_1 and every O_2-translation of every theorem of S_2 is both true and pseudo-true, and therefore that O_1 and O_2 satisfy every principle of S_1 and S_2 respectively. In fact, O_1 and O_2 satisfy more principles than those of S_1 and S_2 respectively; for example, O_1 satisfies $\Diamond\top$, which is not a principle of S_1 (but O_1 does not satisfy $\Box\Diamond\top$; the principles that it satisfies do not constitute a normal system). Note that truth is all that matters for satisfaction of principles, but pseudo-truth must also be established for the inductive argument to go through. In particular, O_1 satisfies $\Box\Box\bot$ and O_2 does not. Thus O_1 and O_2 satisfy different principles. Indeed, they satisfy inconsistent principles, for O_2 satisfies $\sim\Box\Box\bot$. Nevertheless, O_1 and O_2 are coextensive, for each of $O_1\alpha$ and $O_2\alpha$ is true just in case α is pseudo-true.

A different question needs to be asked. Given two monomodal systems S_1 and S_2, when is there a language L with an index set I and coextensive operators O_1 and O_2 such that O_1 satisfies all and only the principles in S_1 and O_2 satisfies all and only the principles in S_2? Note that if S_1 and S_2 are distinct any case in which such operators exist is a case of coextensiveness without satisfaction of the same principles.

The question can be reformulated in terms of bimodal logic. Let $L_{\Box}$ be the result of extending $L_{\Box\blacksquare}$ by a second modal operator $\blacksquare$. Given an interpreted language L with at least the operators $\bot$, $\supset$, O_1 and O_2, define an O_1,O_2-translation to be a mapping * from formulas of $L_{\Box\blacksquare}$ to sentences of L such that for all formulas α and β of $L_{\Box\blacksquare}$, $*(\alpha \supset \beta) = *\alpha \supset *\beta$, $*\bot = \bot$, $*\Box\alpha = O_1*\alpha$ and $*\blacksquare\alpha = O_2*\alpha$. The ordered pair $<O_1,O_2>$ satisfies a principle α ($\in L_{\Box\blacksquare}$) just in case $*\alpha$ is true at every index in the relevant set for every O_1,O_2-translation *. If S_1 and S_2 are monomodal systems in $L_{\Box}$, let $S_1 \circledcirc S_2$ be the smallest bimodal system in $L_{\Box\blacksquare}$ containing every theorem of S_1 and the result of substituting $\blacksquare$ for $\Box$ throughout every theorem of S_2. Let $S_1 \dagger S_2$ be the smallest bimodal system containing $S_1 \circledcirc S_2$ and the formula $\Box p \equiv \blacksquare p$ (by US, $S_1 \dagger S_2$ contains every formula of the form $\Box\alpha \equiv \blacksquare\alpha$). A bimodal system S in $L_{\Box\blacksquare}$ *conservatively* $\Box$*-extends* a monomodal system S′ in $L_{\Box}$ just in case the $\blacksquare$-free theorems of S are exactly the theorems of S′; S *conservatively* $\blacksquare$*-extends* S′ just in case the results of substituting $\Box$ for $\blacksquare$ throughout the $\Box$-free theorems of S are exactly the theorems of S′. Then there is a language L with coextensive operators O_1 and O_2 and an index set I with respect to which O_1 satisfies all and only the principles in S_1 and O_2 satisfies all and only the principles in S_2 if and only if $S_1 \dagger S_2$ conservatively $\Box$-extends S_1 and conservatively $\blacksquare$-extends S_2.

The equivalence can be established as follows. Suppose that there is a language L with coextensive operators O_1 and O_2 and an index set I with respect to which O_1 satisfies all and only the principles in S_1 and O_2 satisfies all and only the principles in S_2. Let S be the set of all principles in $L_{\Box\blacksquare}$ that $<O_1,O_2>$ satisfies. S is a bimodal system. S extends S_1, for if α is a theorem of S_1 then O_1 satisfies α, so every O_1-translation of α is true at every index; but the restriction of any O_1,O_2-translation to $L_{\Box}$ is an O_1-translation, so every O_1,O_2-translation of α is true at every index, so α is a theorem of S. Similarly, S contains the result of substituting $\blacksquare$ for $\Box$ throughout every theorem of S_2. Thus S extends $S_1 \circledcirc S_2$. Since O_1 and O_2 are coextensive, S contains $\Box p \equiv \blacksquare p$. Thus S extends $S_1 \dagger S_2$. Now S conservatively $\Box$-extends S_1. For let α be a $\blacksquare$-free theorem of S; thus every O_1,O_2-translation of α is true at every index; but every O_1-translation has a (unique) extension to an O_1,O_2-translation, so every O_1-translation of α is true at every index, so α is a theorem of S_1. Similarly, S conservatively $\blacksquare$-extends S_2. Since S extends $S_1 \dagger S_2$, the latter conservatively $\Box$-extends S_1 and conservatively $\blacksquare$-extends S_2 too. This establishes one half of the equivalence. Conversely, suppose that $S_1 \dagger S_2$ conservatively $\Box$-extends S_1 and conservatively $\blacksquare$-extends S_2. Let the base language L be $L_{\Box\blacksquare}$ itself, O_1 and O_2 be $\Box$ and $\blacksquare$ respectively, I be the set of all maximal $S_1 \dagger S_2$-consistent subsets of $L_{\Box\blacksquare}$, and α be true at i ($\in$ I) just in case $\alpha \in i$. By an earlier argument, $\Box$ satisfies a principle α of $L_{\Box}$ if and only if α is a theorem of $S_1 \dagger S_2$; since the latter conservatively $\Box$-extends S_1, $\Box$ satisfies α if and only if α is a theorem of S_1. Similarly, $\blacksquare$ satisfies α if and only if α is a theorem of S_2. Since every formula of the form $\Box\alpha \equiv \blacksquare\alpha$ is a theorem of $S_1 \dagger S_2$, $\Box$ and $\blacksquare$ are coextensive. Thus there is a language L with coextensive operators O_1 and O_2 and an index set I with respect to which O_1 satisfies all and only the principles in S_1 and O_2 satisfies all and only the principles in S_2, as required.

Our question is therefore this: when does $S_1 \dagger S_2$ conservatively $\Box$-extend a monomodal system S_1 and conservatively $\blacksquare$-extend a monomodal system S_2? In brief, when is $S_1 \dagger S_2$ a conservative extension? As noted in §2, coextensive operators satisfy the same first-degree principles. Thus a necessary condition for $S_1 \dagger S_2$ to be a conservative extension is that for every first-degree formula α of $L_{\Box}$, α is a theorem of S_1 if and only if α is a theorem of S_2. However, the condition is not sufficient. For let S_1 be the smallest normal modal system containing $\Box\Box\bot$, and let S_2 be the smallest normal modal system containing $({\sim}p \mathbin{\&} \Box p) \supset \Box\Box\bot$. S_1 and S_2 can be shown to have exactly the same first-degree theorems (Appendix, Proposition 1). Nevertheless, $S_1 \dagger S_2$ does not conservatively $\blacksquare$-extend S_2, for it has the following sequence of theorems:

1	$({\sim}p \mathbin{\&} \blacksquare p) \supset \blacksquare\blacksquare\bot$	S_2
2	$({\sim}\bot \mathbin{\&} \blacksquare\bot) \supset \blacksquare\blacksquare\bot$	1 US
3	$({\sim}\Box\bot \mathbin{\&} \blacksquare\Box\bot) \supset \blacksquare\blacksquare\bot$	1 US
4	$(\blacksquare\bot \vee ({\sim}\Box\bot \mathbin{\&} \blacksquare\Box\bot)) \supset \blacksquare\blacksquare\bot$	2, 3 PC
5	$\Box p \equiv \blacksquare p$	$S_1 \dagger S_2$
6	$\Box\bot \equiv \blacksquare\bot$	5 US
7	$\blacksquare\Box\bot \supset \blacksquare\blacksquare\bot$	4, 6 PC
8	$\Box\Box\bot \equiv \blacksquare\Box\bot$	5 US
9	$\Box\Box\bot$	S_1
10	$\blacksquare\blacksquare\bot$	7, 8, 9 PC

Informally, the derivation shows that no operator coextensive with one that satisfies $\Box\Box\bot$ is truth-entailing, which is enough to discharge the antecedent in the S_2 axiom. Thus if $S_1\dagger S_2$ conservatively $\blacksquare$-extends S_2, then $\Box\Box\bot$ is a theorem of S_2. But $\Box\Box\bot$ is not a theorem of S_2, for, since $({\sim}p \ \& \ \Box p) \supset \Box\Box\bot$ is a tautological consequence of $\Box p \supset p$, all theorems of S_2 are theorems of KT (=T), which $\Box\Box\bot$ certainly is not.

In spite of such examples, when $S_1\dagger S_2$ is a conservative extension S_1 and S_2 often agree on little but their first-degree theorems, as earlier examples and later results show. It would be interesting to know whether agreement on first-degree theorems between S_1 and S_2 is sufficient for $S_1\dagger S_2$ to be a conservative extension when S_1 and S_2 are normal extensions of KT, or more generally of KD.

Similar remarks apply to a slightly stronger way of combining monomodal systems. If S_1 and S_2 are systems in $L_\Box$, let $S_1\oplus_2$ be the smallest bimodal system in $L_{\Box\blacksquare}$ containing every theorem of S_1 and the result of substituting $\blacksquare$ for $\Box$ throughout every theorem of S_2, and closed under the rules that if $\alpha \equiv \beta$ is a theorem then so are $\blacksquare\alpha \equiv \blacksquare\beta$ and $\Box\alpha \equiv \Box\beta$. In particular, if S_1 and S_2 are normal systems, then so is $S_1\oplus S_2$, because it is closed under necessitation for both $\Box$ and $\blacksquare$ (for if α is a theorem of $S_1\oplus S_2$, then so is $\alpha \equiv \top$, and therefore $\Box\alpha \equiv \Box\top$ and $\blacksquare\alpha \equiv \blacksquare\top$; but since S_1 and S_2 are normal, $\Box\top$ is a theorem of each, so both $\Box\alpha$ and $\blacksquare\alpha$ are theorems of $S_1\oplus S_2$). In fact $S_1\oplus S_2$ is the smallest normal bimodal system to contain every theorem of S_1 and the result of replacing $\Box$ by $\blacksquare$ throughout every theorem of S_2. Very roughly indeed, the difference between $\oplus$ and $\circledcirc$ is that in $S_1\oplus S_2$, but not in $S_1\circledcirc S_2$, each of $\Box$ and $\blacksquare$ knows what the logic of the other is. Let $S_1\ddagger S_2$ be the smallest bimodal system containing $S_1\oplus S_2$ and the formula $\Box p \equiv \blacksquare p$ (by US, $S_1\oplus S_2$ contains every formula of the form $\Box\alpha \equiv \blacksquare\alpha$).

$S_1\ddagger S_2$ will not in general be a normal system, even if S_1 and S_2 are normal, for it will not in general contain $\Box(\Box p \equiv \blacksquare p)$. Obviously, $S_1\ddagger S_2$ extends $S_1\dagger S_2$. The converse often fails, even when S_1 and S_2 are normal. For example, $\Box\blacksquare\Box\top$ is a theorem of K‡K (indeed of K⊕K), but not of K†K (Appendix, Proposition 2). Another example of the difference between † and ‡ can be gleaned from Example 4. On a view such as Salmon's, the pair of operators $<\Box,@\Box>$ satisfies all the principles of KT†S5, but not all the principles of KT‡S5. For $\Box\Box(\blacksquare p \supset p)$ is a theorem of KT⊕S5 (indeed, of K⊕KT) and therefore of KT‡S5, but it is not satisfied by the pair $<\Box,@\Box>$. If it were, $\Box\Box(@\Box p \supset p)$ would be a theorem of KT@S, in which case $\Box\Box@\Box p \supset \Box\Box p$ would also be a theorem of KT@S; but since $\Box p \supset \Box\Box@\Box p$ is a theorem of KT@S, so too would be $\Box p \supset \Box\Box p$, which it is not (for an easy semantic argument shows KT@S to be a conservative extension of KT). Thus $\Box\Box(\blacksquare p \supset p)$ is not a theorem of KT†S5.

From a semantical point of view, $S_1 \oplus S_2$ is usually a nicer system than $S_1 \circledcirc S_2$, for it can be hard to give $\Box$ and $\blacksquare$ nice semantics when they do not act congruentially. Correspondingly, one might prefer to work with S_1‡S_2 rather than with S_1†S_2. However, the latter system is the relevant one in some applications, as has just been noted in respect of Example 4. Again, if the operators O_1 and O_2 satisfy just the principles of the monomodal systems S_1 and S_2 respectively, and S_1 is not closed under the rule that if $\alpha \equiv \beta$ is a theorem then so is $\Box\alpha \equiv \Box\beta$, it follows that S_1‡S_2 does not conservatively $\Box$-extend S_1, so the only interesting question will be whether the pair of operators $<O_1,O_2>$ satisfies all the principles of S_1†S_2. Such a situation arose in Example 3, for $\Box$ is not congruential in the system GLS, which has $(\Box p \supset p) \equiv \top$ but not $\Box(\Box p \supset p) \equiv \Box\top$ as a theorem. GLS†Grz conservatively $\Box$-extends GLS and conservatively $\blacksquare$-extends Grz; GLS‡Grz is inconsistent, for the closure of GLS under necessitation is inconsistent. More generally, an earlier argument showed that S_1†S_2 is a conservative extension if and only if there are coextensive operators O_1 and O_2 such that O_1 satisfies all and only the principles in S_1 and O_2 satisfies all and only the principles in S_2; no parallel result has been proved for S_1‡S_2. However, in the cases for which it is proved below that S_1†S_2 conservatively $\Box$-extends a normal system S_1 and conservatively $\blacksquare$-extends a normal system S_2, it is also proved that S_1‡S_2 conservatively $\Box$-extends S_1 and conservatively $\blacksquare$-extends S_2. It would be interesting to know whether there are normal systems S_1 and S_2 such that S_1†S_2 but not S_1‡S_2 conservatively $\Box$-extends S_1 and conservatively $\blacksquare$-extends S_2.

It must be checked whether, if S_1†S_2 or S_1‡S_2 is not a conservative extension, the reason is not simply that $S_1 \circledcirc S_2$ or $S_1 \oplus S_2$ is not a conservative extension; if that were the reason, non-conservativeness would show nothing specific about coextensiveness. Call a modal system S in $L_\Box$ *congruential* just in case $\Box\alpha \equiv \Box\beta$ is a theorem of S whenever $\alpha \equiv \beta$ is. Thomason (1980) proved that if S_1 and S_2 are consistent congruential systems, then $S_1 \oplus S_2$ (and a fortiori $S_1 \circledcirc S_2$) conservatively $\Box$-extends S_1 and conservatively $\blacksquare$-extends S_2. (Kracht and Wolter 1991 give a different proof for the case in which S_1 and S_2 are normal. See the latter paper and Fine and Schurz 1996 for many results transferring properties of normal S_1 and S_2 to $S_1 \oplus S_2$.) The result is trivial if S_1 and S_2 are both inconsistent. Since an operator satisfies an inconsistent set of principles if and only if the relevant index set is empty, if O_1 and O_2 are operators in L, and S_1 and S_2 are the systems of the principles satisfied by O_1 and O_2 respectively, then S_1 is consistent if and only if S_2 is. Thus if the congruential systems S_1 and S_2 are the systems of the principles satisfied by the operators O_1 and O_2 respectively in L, then $S_1 \oplus S_2$ (and a fortiori $S_1 \circledcirc S_2$) conservatively $\square$-extends S_1 and conservatively $\blacksquare$-extends S_2. If S_1 or S_2 is non-congruential, then $S_1 \oplus S_2$ is not a conservative extension. But one can still prove that if S_1 and S_2 are consistent modal systems, then $S_1 \circledcirc S_2$ conservatively $\square$-extends S_1 and conservatively $\blacksquare$-extends S_2 (Appendix, Proposition 4). Thus if S_1 and S_2 are the systems of the principles satisfied by the operators O_1 and O_2 respectively in L, then $S_1 \circledcirc S_2$ conservatively $\square$-extends S_1 and conservatively $\blacksquare$-extends S_2.

A sufficient condition for $S_1 \ddagger S_2$ (and therefore $S_1 \dagger S_2$) to be a conservative extension can be developed in semantic terms, as follows. A *frame* is an ordered pair $<W,R>$, where W is a set (of 'worlds') and R is a binary relation (of 'accessibility') on W. $<W,R>$ is said to be reflexive just in case R is reflexive; likewise for other frame conditions. A *model based on* $<W,R>$ is a triple $<W,R,V>$, where V is a mapping from formulas of $L_\square$ and members of W to $\{0,1\}$ (0 for falsity, 1 for truth) such that for all $w \in W$ and $\alpha, \beta \in L_\square$, $V(\bot, w) = 0$, $V(\alpha \supset \beta, w) = 1$ if and only if $V(\alpha, w) \leq V(\beta, w)$, and $V(\square\alpha, w) = 1$ if and only if $V(\alpha, x) = 1$ for every $x \in W$ such that wRx. A formula α is *valid* on a frame $<W,R>$ just in case for every model $<W,R,V>$ based on $<W,R>$ and $w \in W$, $V(\alpha, w) = 1$. A system S is valid on $<W,R>$ just in case every theorem of S is; if S is valid on $<W,R>$, say that $<W,R>$ is a frame *for* S. The *logic* of a class of frames Γ is the set of all formulas valid on all members of Γ. The logic of a class of frames is automatically a normal system, but not every normal system is the logic of a class of frames (the exceptions are the incomplete systems). Now let Γ_1 and Γ_2 be classes of frames; let S_1 and S_2 be the logics of Γ_1 and Γ_2 respectively. Suppose that for each frame $<W_1,R_1> \in \Gamma_1$ and $w \in W_1$ there is a frame $<W_2,R_2> \in \Gamma_2$ such that $w \in W_2$ and $\{x \in W_1: wR_1x\} = \{x \in W_2: wR_2x\}$, and that for each frame $<W_2,R_2> \in \Gamma_2$ and $w \in W_2$ there is a frame $<W_1,R_1> \in \Gamma_1$ such that $w \in W_1$ and $\{x \in W_1: wR_1x\} = \{x \in W_2: wR_2x\}$. Then $S_1 \ddagger S_2$ (and therefore $S_1 \dagger S_2$) conservatively $\square$-extends S_1 and conservatively $\blacksquare$-extends S_2 (Appendix, Proposition 6).

Informally, the condition says that the two classes of frames are locally alike, in the sense that what can be reached in one step of accessibility from a given point is the same. This condition may be compared with the syntactic necessary condition discussed above for $S_1 \dagger S_2$ (and therefore $S_1 \ddagger S_2$) to be a conservative extension: that S_1 and S_2 should have the same first-degree theorems. The two conditions are closely related, for whether a first-degree formula is true at a world *w* depends only on worlds at most one step of accessibility from *w*. The two conditions are not equivalent, for the constraints that a system imposes on the one-step environments of worlds are not exhausted by its first-degree theorems. For example, if $\Box\Box\bot$ is a theorem of S, then S is valid only on irreflexive frames, even though S may have the same first-degree theorems as a system that is valid on some (or even all) reflexive frames. Whereas a frame is reflexive if and only if $\Box p \supset p$ is valid on it, there is no formula $\alpha \in L_\Box$ (let alone a first-degree one) such that a frame is irreflexive if and only if α is valid on it, for the smallest normal system K is complete both for the class of all irreflexive frames and for the class of all frames. Irreflexivity is however characterized in a weaker sense by the rule that if $(\sim p \;\&\; \Box p) \supset \alpha$ is a theorem then so is α, where the propositional variable *p* does not occur in α (Gabbay 1981). An example of just this kind was used above to show that for S_1 and S_2 to have the same first-degree theorems is not sufficient for $S_1 \ddagger S_2$ or even $S_1 \dagger S_2$ to be a conservative extension. The semantic sufficient condition above will be applied in §5.

Within the framework just constructed, the rest of the paper addresses some more specific logical issues raised by the examples in §1.

4. *Solovay extensions*

Examples 2 and 3 (as regards the Solovay logic GLS) can be generalized. Consider an operator O in a base language L with respect to an index set I. Suppose that for each index *i*, $O\alpha$ is true at *i* only if α is true at *i* (e.g. O is 'Socrates believes that . . .', if Socrates is in fact a wise man at all indices) . . . Define a new operator O^+ in L by the stipulation that $O^+\alpha = O\alpha \;\&\; \alpha$. Thus O and O^+ are coextensive. The principles that O^+ satisfies are related in a simple way to the principles that O satisfies. Inductively define a mapping τ from $L_\Box$ to $L_\Box$ thus:

$$\tau p = p$$
$$\tau\bot = \bot$$
$$\tau(\alpha \supset \beta) = \tau\alpha \supset \tau\beta$$
$$\tau\Box\alpha = \Box\tau\alpha \;\&\; \tau\alpha$$

For each O^+-translation $*$, let $*^-$ be the unique O-translation such that $*^-p = *p$ for each propositional variable p; for each O-translation $*$, let $*^+$ be the unique O^+-translation such that $*^+p = *p$ for each p. It is easy to prove by induction on the construction of a formula α in $L_\Box$ that for each O^+-translation $*$, $*\alpha = *^-\tau\alpha$, and that for each O-translation $*$, $*^+\alpha = *\tau\alpha$. Now if O^+ does not satisfy α, then $*\alpha$ is false at some index i for some O^+-translation $*$, so $*^-\tau\alpha$ is false at i, so O does not satisfy $\tau\alpha$. Conversely, if O does not satisfy $\tau\alpha$, then $*\tau\alpha$ is false at some index i for some O-translation $*$, so $*^+\alpha$ is false at i, so O^+ does not satisfy α. Thus O^+ satisfies the principle α if and only if O satisfies the principle $\tau\alpha$. Let S be the system of principles that O satisfies. Then a necessary and sufficient condition for O and O^+ to satisfy the same principles is that for each formula α in $L_\Box$, $\tau\alpha$ is a theorem of S if and only if α is a theorem of S.

By hypothesis, O satisfies the principle $\Box p \supset p$, which is therefore a theorem of S. Using this fact, one can show by induction on the construction of α that if S is congruential then $\tau\alpha \equiv \alpha$ is a theorem of S. So O and O^+ satisfy the same principles, if S is congruential. For present purposes, the interesting systems are therefore non-congruential ones. In some of the examples in §1, O and O^+ satisfied different principles because O^+ satisfied $\Box(\Box p \supset p)$ while O did not. Correspondingly, the system of principles satisfied by O was non-congruential because $(\Box p \supset p) \equiv \top$ but not $\Box(\Box p \supset p) \equiv \Box\top$ was a theorem.

Non-congruential systems are often semantically recalcitrant. Sometimes, however, the non-congruential system S′ with the theorem $\Box p \supset p$ has a congruential subsystem S without the theorem $\Box p \supset p$ such that S′ can be axiomatized with all theorems of S and all substitution instances of $\Box p \supset p$ as the axioms and MP as the only rule of inference. The non-congruential Solovay logic GLS stands in exactly this relation to the provability logic GL, which is normal, not just congruential. GL has a nice semantics: it is sound and complete for the class of all possible worlds models based on finite transitive irreflexive trees. In contrast, there is no frame <W,R> and world $w \in W$ such that all theorems of GLS are true at w in all models based on the frame <W,R>. Proof: The theorem $\Box(\Box p \supset p) \supset \Box p$ of GLS is true at $w \in W$ in all models based on <W,R> only if not wRw (consider an assignment on which p is true everywhere except at w); $\Box p \supset p$ is true at w in all models based on <W,R> only if wRw.

More generally, let the *Solovay extension* of a system S be the smallest system containing $\Box p \supset p$ and every theorem of S; since

Solovay extensions are systems, they are automatically closed under MP and contain every substitution instance of $\Box p \supset p$. One can reverse the direction of investigation and examine the Solovay extensions of given congruential systems.

Let S be a normal system, and S^+ the Solovay extension of S. Suppose that S^+ contains all and only the principles satisfied by an operator O. Since S^+ contains the principle $\Box p \equiv (\Box p \,\&\, p)$, O and O^+ are coextensive. What conditions on S ensure that O and O^+ satisfy different principles? Obviously, if S already contains $\Box p \supset p$, then S^+ is S itself (and conversely); in that case, S^+ is normal, and therefore congruential, so by remarks above O and O^+ satisfy the same principles. More interestingly, suppose that S admits the rule of disjunction, in the sense that whenever $\Box\alpha_1 \vee \ldots \vee \Box\alpha_n$ is a theorem of S, so too is α_m for some m $(1 \leq m \leq n)$. It can then be shown that any formula of the form $\Box\alpha$ is a theorem of S^+ only if α is already a theorem of S (Appendix, Proposition 10). In particular, if $\Box(\Box p \supset p)$ is a theorem of S^+, then $\Box p \supset p$ is a theorem of S, and S^+ is just S itself. Thus if S^+ is a proper extension of S, then $\Box(\Box p \supset p)$ is not a theorem of S^+, so O does not satisfy the principle $\Box(\Box p \supset p)$. On the other hand, $\tau\Box(\Box p \supset p) = \Box((\Box p \,\&\, p) \supset p) \,\&\, ((\Box p \,\&\, p) \supset p)$, which is a theorem of every normal system, hence of S, hence of S^+; thus O certainly satisfies the principle $\tau\Box(\Box p \supset p)$, so O^+ satisfies the principle $\Box(\Box p \supset p)$. What all this shows is that if S is a normal system that admits the rule of disjunction and lacks the theorem $\Box p \supset p$, and an operator O satisfies all and only the principles of the Solovay extension S^+, then O and O^+ are coextensive operators that satisfy different principles. Many normal systems do admit the rule of disjunction and lack the theorem $\Box p \supset p$, e.g. K, KD, K4, KD4 and GL itself. Thus Solovay extensions provide numerous examples of the phenomenon with which this paper is concerned. Examination of the proof shows that the remarks in the text can be generalized from the case in which S is normal to the case in which it is a congruential system with the theorem $\Box\top$ (thus S admits the rule of necessitation, but $\Box(p \supset q) \supset (\Box p \supset \Box q)$ need not be a theorem). For more on the rule of disjunction see Hughes and Cresswell 1984.

In the case of GL and GLS, it is not surprising that $\Box\alpha$ should be a theorem of GLS only if α (and therefore $\Box\alpha$) is already a theorem of GL. For GLS states what is true about provability in PA ($\Box$), while GL states what is provable in PA about provability in PA. If the result of replacing the propositional variables in $\Box\alpha$ by sentences of L_{PA} is always true, then the result of so replacing them in α is always provable. Subtler issues arise over the interpretation of Proposition 10 for GL and GLS when $n>1$. If $\Box\alpha \vee \Box\beta$ is always true, it is less trivial that either α is always provable in PA or β is.

An example of a system that does not admit the rule of disjunction and in which things go very differently is K5, the smallest normal system containing $\sim\Box p \supset \Box\sim\Box p$. Since $\Box(\Box p \supset p)$ is a theorem of K5, the Solovay extension of K5 is S5, whose normality means that any corresponding operators O and O^+ satisfy the same principles. In contrast, if the system S admits the rule of disjunction, then any normal subsystem of the Solovay extension S^+ of S is a subsystem of S, for if α is a theorem of a normal subsystem of S^+, then $\Box\alpha$ is too, so α is a theorem of S (Appendix, Proposition 10 for $n=1$).

5. *Introspection and 'actually' operators*

The following are sometimes described as principles of introspection:

4 $\Box p \supset \Box\Box p$
5 $\sim\Box p \supset \Box\sim\Box p$

Such a description is given when $\Box$ is interpreted as a knowledge or belief operator. On such an interpretation, 4 and 5 claim that if one knows something, one knows that one knows it, and that if one does not know it, one knows that one does not know it, or that if one believes something, one believes that one believes it, and that if one fails to believe something, one believes that one fails to believe it. Really, the term 'introspection' is more naturally associated with the bimodal claims that if one believes something, one knows that one believes it, and that if one fails to believe it, one knows that one fails to believe it, for introspection is supposed to be a faculty of inner perception by which one comes to know one's internal mental states. The present concern, however, is not with the nature of introspection, nor with the sorts of idealization that would be needed to make 4 or 5 plausible for knowledge or belief (considered as an empirical generalization about ordinary subjects, each of the preceding epistemic and doxastic claims is false). A simple creature that has knowledge and beliefs but lacks the concepts of knowledge and belief would constitute a counter-example to both 4 and 5 on both the epistemic and doxastic interpretations. The argument of Davidson 1984, p. 170, to the effect that no creature can have beliefs without having the concept of belief seems to exploit an equivocation between *de re* and *de dicto* readings of 'understands the possibility of being mistaken'. If one believes that P, one must understand what it would be for it not to be the case that P, which is the condition for one's belief to be false; but to assume that one must grasp it *as* the condition for one's belief to be false is to beg the question.

The question is rather: when is a given operator coextensive with one that satisfies 4 or 5? Such an operator need not be epistemic or doxastic: consider Salmon's treatment of necessity and actual necessity, discussed in Example 4. Answers to that question may help to explain the deceptive appeal of 4 and 5 in some cases, but the aim of this section is mainly to clarify the formal position.

Not every operator is coextensive with an operator that satisfies 4 and 5. For example, if negation were coextensive with an operator O that satisfied 4, then, at any index, since $\sim\perp$ is true, $O\perp$ would be true (by coextensiveness), so $OO\perp$ would be true (by 4), so $\sim O\perp$ would be true (by coextensiveness), so $O\perp$ would be false, so $\sim\perp$ would be false (by coextensiveness yet again), which is impossible. Similarly, if negation were coextensive with an operator O that satisfied 5, then, at any index, $O\top$ would be false (by coextensiveness), so $O\sim O\top$ would be true (by 5), so $\sim\sim O\top$ would be true (by coextensiveness), so $O\top$ would be true, which is a contradiction. Thus negation is not coextensive with any operator O that satisfies either 4 or 5.

Nevertheless, the treatment of the relation between necessity and actual necessity in Example 4 suggested that under fairly general conditions a given operator is coextensive with an operator that satisfies both 4 and 5. Such conditions will now be identified. An obvious strategy is to introduce something like an 'actually' operator while assuming as little about the background as possible.

Given an operator O in a language L, L itself will often contain no operator coextensive with O that satisfies 4 and 5. The obvious question is whether L can be extended to a language that does contain an operator coextensive with O that satisfies 4 and 5. But it may be unclear whether O in the extended language retains the meaning that it had in the original language, for it applies to sentences of the extended language that are not in the original language. In what follows, a sentence of the original language turns out to be true at an index in the extended language if and only if it is true at that index in the original language; moreover, O turns out to satisfy exactly the same principles with respect to the extended language as it did with respect to the original language. It is left to the reader to judge whether O retains its original meaning.

For simplicity, consider a language L_O made up of atomic sentences, the truth-functional operators $\bot$ and $\supset$ and the operator O. The arguments below can be generalized to languages with a more extensive vocabulary. Extend L_O to a language $L_{O@}$ by adding the singulary operator @. For each index *i*, define a mapping ϕ_i from $L_{O@}$ to L_O inductively as follows:

$$\begin{aligned}
\phi_i q &= q \quad (q \text{ atomic}) \\
\phi_i \bot &= \bot \\
\phi_i(\alpha \supset \beta) &= \phi_i\alpha \supset \phi_i\beta \\
\phi_i O\alpha &= O\phi_i\alpha \\
\phi_i @\alpha &= \top \quad \text{if } \phi_i\alpha \text{ is true at } i \text{ in } L_O \\
\phi_i @\alpha &= \bot \quad \text{otherwise}
\end{aligned}$$

For any $\alpha \in L_{O@}$ and index *i*, let α be true at *i* in $L_{O@}$ just in case $\phi_i\alpha$ is true at *i* in L_O. Since $\bot$ and $\supset$ have their usual truth-tables in L_O and ϕ_i commutes with them for each *i*, they have their usual truth-tables in $L_{O@}$. If $\alpha \in L_O$, then $\phi_i\alpha = \alpha$, so α is true at *i* in $L_{O@}$ if and only if α is true at *i* in L_O. Thus the new clauses for truth do not remove any counter-examples to principles that O did not satisfy with respect to the original language. It is not hard to show that for any principle $\alpha \in L_\Box$, O satisfies α with respect to $L_{O@}$ if and only if it satisfies it with respect to L_O. For let * be a O-translation into L_O; thus * is also a O-translation into $L_{O@}$. Moreover, if $\alpha \in L_\Box$ then, for any index *i*, $\phi_i{*\alpha} = {*\alpha}$, so $*\alpha$ is true at *i* in $L_{O@}$ if and only if $*\alpha$ is true at *i* in L_O. Thus if O satisfies α with respect to $L_{O@}$, it satisfies it with respect to L_O. Conversely, let * be a O-translation into $L_{O@}$, and *i* an index. Let $*_i$ be the O-translation into L_O such that $*_i p = \phi_i {*p}$ for each propositional variable *p*. By a routine induction on the complexity of $\alpha \in L_\Box$, $*_i\alpha = \phi_i{*\alpha}$. Hence $*\alpha$ is true at *i* in $L_{O@}$ if and only if $*_i\alpha$ is true at *i* in L_O. Thus if O satisfies α with respect to L_O, it satisfies it with respect to $L_{O@}$.

Thus the expansion of the language does not introduce any counter-examples to principles that O did satisfy with respect to the original language. Henceforth, truth at an index and satisfaction of a principle will be in and with respect to $L_{O@}$ unless otherwise specified.

Given this account of @, it should be stressed that in most cases the analogy between @ and 'actually' is no more than an analogy. If the operators 'It is known that' and 'It is believed that' do not satisfy 4 or 5, no more do the operators 'It is actually known that' and 'It is actually believed that'. There is no obvious general way of rendering @ in English. Nevertheless, truth-conditions are as well-defined for sentences of $L_{O@}$ as for sentences of L_O. @ can be interpreted as mapping all true propositions to one tautologous proposition and all false propositions to one self-contradictory proposition. It might be objected that the proposition expressed by α will then not in general be a constituent of the proposition expressed by $@\alpha$. But that is an inessential feature of the construction. $\phi_i@\alpha$ could just as well have been defined as the tautology $\phi_i\alpha \supset \phi_i\alpha$ if $\phi_i\alpha$ is true at i in L_O and as the contradiction $\phi_i\alpha$ & $\sim\phi_i\alpha$ otherwise, in which case the proposition expressed by α would be a constituent of the proposition expressed by $@\alpha$.

The construction has some unusual features. For concreteness, write O . . . as 'John knows that' and $\top$ as 'If it is raining then it is raining'; for simplicity, suppress the index *i*. Suppose that there is in fact intelligent extraterrestrial life, but that John is wholly ignorant as to whether there is intelligent extraterrestrial life, although he does know the tautology that if it is raining then it is raining. Thus ϕ maps the sentence '@ there is intelligent extraterrestrial life' of $L_{O@}$ to the sentence 'If it is raining then it is raining' of L_O. Consequently, ϕ maps the sentence 'John knows that @ there is intelligent extraterrestrial life' of $L_{O@}$ to the sentence 'John knows that if it is raining then it is raining' of L_O; since the latter sentence is true, so too is the former. Nevertheless, if one asks John the question 'Is it the case that @ there is intelligent extraterrestrial life?' he will have to shrug his shoulders, even though he knows the semantic rules of the language (e.g. the definition of ϕ) and desires to be helpful. Since he does in fact know that @ there is intelligent extraterrestrial life (i.e. he knows the truth expressed by the sentence), it would be inaccurate of him to answer 'I don't know'. His problem is presumably that he fails to know which proposition is expressed by the sentence '@ there is intelligent extraterrestrial life'. That ignorance is not due to ignorance of its linguistic meaning, for he knows the relevant semantic rules; what he does not know is which proposition those rules assign to the sentence, for that depends on non-semantic facts of which he is ignorant. One might compare him to someone who knows the linguistic meaning of 'That spider is poisonous' but cannot identify the spider which the speaker is pointing at, except that John's predicament is likely to be the rule, not the exception, for speakers of $L_{O@}$. The language has grave practical disadvantages. Nevertheless, one can put these doubts to one side and use $L_{O@}$ to explore the phenomenon of coextensive operators that differ in satisfaction of the principles 4 and 5.

For any $\alpha \in L_{O@}$ and index *i*, $@\alpha$ is true at *i* if and only if α is true at *i*. In particular, $@O\alpha$ and $O\alpha$ are true at the same indices, so $@O$ and O are coextensive operators. For $@\alpha$ is true at *i* if and only if $\phi_i @\alpha$ is true at *i* in L_O, which is so if and only if $\phi_i @\alpha$ is $\top$, which is so if and only if $\phi_i \alpha$ is true at *i* in L_O, which is so if and only if α is true at *i*. It follows immediately that for all $\alpha \in L_{O@}$, the sentence $@\alpha \equiv \alpha$ is true at all indices, so for all $\alpha, \beta \in L_{O@}$ each of $@(\alpha \supset \beta) \supset (@\alpha \supset @\beta)$, $@{\sim}\alpha \equiv {\sim}@\alpha$ and $@\alpha \supset @@\alpha$ is true at all indices. It also follows that if α is true at all indices so too is $@\alpha$ (compare the axioms and rules of the systems KT@ and KT@S in Example 4 that do not involve $\Box$). If @ is read as 'actually' within the framework of possible worlds semantics, the fact that $@\alpha$ and α are true at the same indices means that indices are functioning more like contexts of utterance than like contexts of evaluation. However, the construction of $L_{O@}$ out of L_O, with the semantics for @ in terms of the mappings ϕ_i, does not rest on any special assumptions about the operator O.

It must now be checked that the composite operator @O satisfies the principles 4 and 5. More generally, @O will be shown to satisfy any monomodal principle of the form $\alpha \supset \Box\alpha$, where α is *fully modalized*, in the sense that propositional variables occur in α only inside the scope of $\Box$. 4 and 5 are themselves of this form, and all principles of this form are theorems of K45, the smallest normal system containing 4 and 5. Since not every operator is coextensive with an operator that satisfies 4 and 5, some assumption will be needed about the operator O. It is enough to assume that whenever $\alpha \in L_O$ is a truth-functional tautology, $O\alpha$ is true at every index. This is a considerable weakening of the rule of necessitation ($\Box\top$ must be a theorem but $\Box\Box\top$ need not be). Now let $\alpha \in L_\Box$ be fully modalized, and * an @O-translation into $L_{O@}$. By definition of ϕ_i and the fact that α is fully modalized, $\phi_i{*}\alpha$ is either a tautology or a contradiction. Thus if $*\alpha$ is true at an index *i*, then $\phi_i{*}\alpha$ is true at *i* in L_O, so $\phi_i{*}\alpha$ is a tautology, so $O\phi_i{*}\alpha = \phi_i O{*}\alpha$ is true at *i* in L_O, so $\phi_i{*}\Box\alpha = \phi_i @O{*}\alpha = \top$, so $*\Box\alpha$ is true at *i*. Thus $*(\alpha \supset \Box\alpha)$ is true at every index for every @O-translation *, so @O satisfies the principle $\alpha \supset \Box\alpha$, for any fully modalized α. In particular, @O satisfies 4 and 5. By similar reasoning, the pair of operators <O,@> satisfies the bimodal principle $\blacksquare p \supset \Box\blacksquare p$ (compare the axiom $@p \supset \Box @p$ of the system KT@ in Example 4).

If α is fully modalized, then so too is the formula $\alpha \supset \Box\alpha$. By the reasoning above, @O also satisfies the principle $(\alpha \supset \Box\alpha) \supset \Box(\alpha \supset \Box\alpha)$, and therefore the principle $\Box(\alpha \supset \Box\alpha)$. By iteration of the argument, @O satisfies $\Box^n(\alpha \supset \Box\alpha)$, the n-fold necessitation of $\alpha \supset \Box\alpha$, for each n. In particular @O satisfies all necessitations of 4 and 5. Similarly, if α is a truth-functional tautology, @O satisfies $\Box^n\alpha$ for each n. If O satisfies the principle $\Box(p \supset q) \supset (\Box p \supset \Box q)$, the K axiom, so too does @O, because it is coextensive with O and the principle is first-degree; and since the K axiom is fully modalized, @O also satisfies $\Box^n(\Box(p \supset q) \supset (\Box p \supset \Box q))$ for each n. It follows that in this case @O satisfies every principle in the system K45, which can be axiomatized by those necessitations, with MP as the only primitive rule of inference (necessitation becomes a derived rule).

One corollary of what has just been shown is that if S is any normal modal system, then S†K45 conservatively □-extends S. For, as noted in §3, there is an operator O that satisfies all and only the principles in S. By the preceding construction, the pair <O,@O> satisfies all principles in S†K45 (since S is normal), and the only principles in $L_\Box$ it satisfies are in S, so S†K45 conservatively □-extends S. S†K45 conservatively ■-extends K45 only if all the first-degree theorems of S are also theorems of K45.

It can be shown that for any normal modal system S, there is exactly one extension S′ of K45 such that S†S′ conservatively □-extends S and conservatively ■-extends S′. There is at least one, for if S is the system of principles satisfied by an operator O and S′ by @O, the system satisfied by the pair <O,@O> both extends S†S′ and conservatively □-extends S and conservatively ■-extends S′, which extends K45. S′ is unique, for any system with the required properties is the smallest extension of K45 containing all first-degree theorems of S (by Proposition 9 of the Appendix; as a corollary, if all the first-degree theorems of S are also theorems of K45, then S†K45 conservatively ■-extends K45). Thus the system of principles satisfied by @O is the smallest extension of K45 containing all first-degree principles satisfied by O, provided that the system of principles satisfied by O is normal.

It follows that if S_1 and S_2 are normal systems with the same first-degree theorems, and S′ is the smallest extension of K45 containing those theorems, and $S_1 \subseteq S'$ and $S_2 \subseteq S'$, then $S_1 \dagger S_2$ conservatively □-extends S_1 and conservatively ■-extends S_2. For by what has just been noted, $S_1 \dagger S'$ conservatively □-extends S_1; since $S_2 \subseteq S'$, $S_1 \dagger S_2$ conservatively □-extends S_1. By parallel reasoning, $S_1 \dagger S_2$ conservatively ■-extends S_2 (compare Proposition 8 of the Appendix). This result means that over a wide range of cases, for S_1 and S_2 to have the same first-degree theorems is necessary and sufficient for $S_1 \dagger S_2$ to be a conservative extension.

Some remarks on the operator @ may be useful. Without any assumptions about the operator O, one can show that @ is eliminable from $L_{O@}$, in the sense that for any $\alpha \in L_{O@}$ there is some $\beta \in L_O$ such that $\alpha \equiv \beta$ is true at all indices. Moreover, @ can be eliminated in a uniform way: for any $\alpha \in L_{\Box\blacksquare}$ there is some $\beta \in L_\Box$ such that the pair <O,@> satisfies the principle $\alpha \equiv \beta$. To prove this, it is useful to define the result of substituting γ for $\blacksquare\delta$ throughout $\alpha \in L_{\Box\blacksquare}$, $[\gamma{:}\delta]\alpha$, as follows (a rigorous definition is needed because one substitution may create new occurrences of $\blacksquare\delta$):

$$[\gamma{:}\delta]p = p \quad (p \text{ a propositional variable})$$
$$[\gamma{:}\delta]\bot = \bot$$
$$[\gamma{:}\delta](\alpha \supset \beta) = [\gamma{:}\delta]\alpha \supset [\gamma{:}\delta]\beta$$
$$[\gamma{:}\delta]\Box\alpha = \Box[\gamma{:}\delta]\alpha$$
$$[\gamma{:}\delta]\blacksquare\alpha = \gamma \quad \text{if } \alpha = \delta$$
$$[\gamma{:}\delta]\blacksquare\alpha = \blacksquare[\gamma{:}\delta]\alpha \quad \text{otherwise}$$

By induction on the complexity of α, for any O,@-translation *, if $*\delta$ is true at an index i then $\phi_i{*}[\top{:}\delta]\alpha = \phi_i{*}\alpha$, so $*[\top{:}\delta]\alpha$ is true at i iff $*\alpha$ is true at i. Similarly, if $*\delta$ is false at i, then $*[\bot{:}\delta]\alpha$ is true at i iff $*\alpha$ is true at i. Thus the pair <O,@> satisfies the principle

$$\alpha \equiv ((\delta \supset [\top{:}\delta]\alpha) \;\&\; ({\sim}\delta \supset [\bot{:}\delta]\alpha))$$

If $\blacksquare\delta$ occurs in α, then each of δ, $[\top{:}\delta]\alpha$ and $[\bot{:}\delta]\alpha$ has fewer occurrences of $\blacksquare$ than α. By iteration of the argument, <O,@> satisfies a principle $\alpha \equiv \beta$ such that $\blacksquare$ does not occur in β.

This result means that the bimodal system $S_{O@}$ of principles in $L_{\Box\blacksquare}$ satisfied by <O,@> maximally conservatively $\Box$-extends the monomodal system S_O of principles in $L_\Box$ satisfied by O, in the sense that no proper extension of $S_{O@}$ in $L_{\Box\blacksquare}$ conservatively $\Box$-extends S_O. For let the extension S of $S_{O@}$ in $L_{\Box\blacksquare}$ conservatively $\Box$-extend S_O, and α be a theorem of S. For some $\beta \in L_\Box$, $\alpha \equiv \beta$ is a theorem of $S_{O@}$, and therefore of S (which extends $S_{O@}$), so β is a theorem of S, and therefore of S_O (which is conservatively $\Box$-extended by S), and therefore of $S_{O@}$, so α is a theorem of $S_{O@}$. Thus S is $S_{O@}$. Indeed, similar reasoning shows that $S_{O@}$ is simply the bimodal extension of S_O by the biconditionals $\alpha \equiv ((\delta \supset [\top{:}\delta]\alpha) \;\&\; ({\sim}\delta \supset [\top{:}\delta]\alpha))$ used in the proof above. Since those biconditionals should be satisfied by <O,@> if @ is to stand to O as 'actually' stands to 'necessarily', $S_{O@}$ is the only system that both conservatively $\Box$-extends S_O and contains the intuitively desirable biconditionals. These remarks might be regarded as an informal argument for the soundness and completeness of $S_{O@}$ relative to its intended interpretation in some contexts.

As Hazen (1978) has pointed out in a similar connection, the eliminability of @ is peculiar to propositional logic (he establishes eliminability for one such system). The very point of introducing an 'actually' operator is that it is ineliminable in the context of first-order quantificational logic (consider 'It could have happened that everyone who actually voted for the motion voted against it'). The definition of $\phi_i@\alpha$ does not generalize in a suitable way to the case in which α has free variables. Of course, quantified modal logic can be used to code the quantified principles that operators satisfy, in a generalization of the approach of §2, but such a generalization lies outside the scope of this paper. Moreover, the case of provability logic shows that the system of quantified principles that an operator satisfies may be unmanageably complex, even though the system of propositional principles that it satisfies is attractively simple (Boolos 1993, pp. 219–255).

Suppose now that the operator O satisfies the principles of the normal system S. One disadvantage of the 'actually' construction is that, although the pair <O,@O> satisfies the principles of S†K45, it need not satisfy the principles of the stronger composite S‡K45. In particular, suppose that O satisfies the T principle $\Box p \supset p$ but does not already satisfy the 4 principle $\Box p \supset \Box\Box p$. For some O-translation * into L_O and index i, $O{*}p$ is true and $OO{*}p$ false at i. Extend * to a O,@O-translation $*'$. But $\blacksquare p \supset p$ is a theorem of S‡K45 (indeed, of S†K), so $\Box\Box(\blacksquare p \supset p)$ is also a theorem of S‡K45 (although not of S†K45). Now $*'\Box\Box(\blacksquare p \supset p) = OO(@O{*}p \supset {*}p)$; since ${*}p \in L_O$, $\phi_i{*}p = {*}p$, so $\phi_i O{*}p = O{*}p$, which is true at i, so $\phi_i @O{*}p = \top$, so $\phi_i OO(@O{*}p \supset {*}p) = OO(\top \supset {*}p)$, which is false at i, for it is true at the same indices as $OO{*}p$ (since S is normal); hence $OO(@O{*}p \supset {*}p)$ is false at i. Thus the pair <O,@O> does not satisfy the principle $\Box\Box(\blacksquare p \supset p)$. But this does not imply that S‡K45 does not conservatively $\Box$-extend S, i.e. that no operator O satisfying exactly the principles of S can be coextensive with an operator O′ such that the pair <O,O′> satisfies the principles of S‡K45. In many such cases, S‡K45 does conservatively $\Box$-extend S: for example, when S is KT. It is just that, to establish such results, O′ must be chosen to be something other than @O.

A semantic method is now convenient. It is somewhat less general, however, than the syntactic method used to introduce @. The relevant proof in the Appendix applies only to complete (and therefore normal) modal systems, i.e. to those consisting of all formulas valid in all models based on a given class of frames. The normal systems that arise in practice are complete, but some non-normal systems are of interest in connection with epistemic or doxastic applications (ordinary subjects do not know or believe everything that follows from what they know or believe). At any rate, let S be a complete system. Then it can be shown that S‡K45 conservatively □-extends S. Moreover, if S′ is the smallest extension of K45 by all first-degree theorems of S, then S‡S′ conservatively □-extends S and conservatively ■-extends S′ (Appendix, Proposition 7).

To illustrate the scope of these results, let S_1 and S_2 be any complete systems such that $KT \subseteq S_1 \subseteq S5$ and $KT \subseteq S_2 \subseteq S5$. Now the smallest extension of K45 by all first-degree theorems of S_1 includes S5, for the T axiom is a first-degree theorem of S_1, since $KT \subseteq S_1$. Conversely, S5 includes the smallest extension of K45 by all first-degree theorems of S_1, all of which are theorems of S5, since $S_1 \subseteq S5$. Thus the smallest extension of K45 by all first-degree theorems of S_1 is exactly S5. Thus S_1‡S5 conservatively □-extends S_1. Since $S_2 \subseteq S5$, S_1‡S_2 conservatively □-extends S_1. Similarly, S_1‡S_2 conservatively ■-extends S_2. Thus S_1‡S_2 is a conservative extension. By similar arguments, one can show that if S_1 and S_2 are complete and either $KD \subseteq S_1 \subseteq KD45$ and $KD \subseteq S_2 \subseteq KD45$ or $K \subseteq S_1 \subseteq K45$ and $K \subseteq S_2 \subseteq K45$ then S_1‡S_2 is a conservative extension. Many other results can be proved along these lines. They cover a wide range of the normal systems of most interest.

Many operators that satisfy neither 4 nor 5 are coextensive with operators that satisfy both 4 and 5. By the same token, many operators that satisfy both 4 and 5 are coextensive with operators that satisfy neither 4 nor 5. In such cases, it is impossible to decide whether an operator satisfies one of those principles on the basis of arguments that do not discriminate between coextensive operators. For instance, some kinds of functional consideration about propositional attitudes are insensitive to differences between coextensive operators. Take as an example the principle ‘When one desires that P and one believes that if one does A then P, one usually does A’; if it holds at all, it holds whenever operators coextensive with ‘one desires that’ and ‘one believes that’ respectively are substituted for them. It is therefore impossible to determine on the basis of such considerations whether, if one believes something, one also believes that one believes it. An adequate account of the operators '*x* believes that' and '*x* desires that' must be based on considerations fine-grained enough to discriminate between them and operators merely coextensive with them.

Note: Earlier versions of this material were presented at Oriel College Oxford, the University of Canterbury (Christchurch), Monash University and the University of Edinburgh. I thank the audiences there and at the British Academy Symposium, and in particular Dorothy Edgington, for their constructive comments. At an earlier stage, George Boolos, Max Cresswell, Kit Fine and Lloyd Humberstone asked fruitful technical questions.

Appendix

This appendix collects the longer proofs needed for the paper.

Notation: L_{PC} is the language of the non-modal propositional calculus. As usual, a *substitution* from L to L′ is a mapping from formulas to formulas of L′ that commutes with all operators in L (thus all operators in L must also be in L′). The smallest modal system is K.

PROPOSITION 1. The smallest normal modal system containing $\Box\Box\bot$ has the same first-degree theorems as K; so has the smallest normal modal system containing $(p \mathbin{\&} \Box{\sim}p) \supset \Box\Box\bot$.

PROOF. Let S be the smallest normal modal system containing $\Box\Box\bot$, and S′ the smallest normal modal system containing $(p \mathbin{\&} \Box{\sim}p)\ \varepsilon\ \Box\Box\bot$. Define two functions ϕ and ψ from $L_\Box$ to $L_\Box$ thus:

$$\phi p = \psi p = p$$
$$\phi\bot = \psi\bot = \bot$$
$$\phi(\alpha \supset \beta) = \phi\alpha \supset \phi\beta \qquad \psi(\alpha \supset \beta) = \psi\alpha \supset \psi\beta$$
$$\phi\Box\alpha = \top \qquad \psi\Box\alpha = \Box\phi\alpha$$

Axiomatize S with all truth-functional tautologies and all formulas of the form $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$ and $\Box\Box\bot$ as the axioms and MP and necessitation as the rules of inference. By induction on the length of proofs, if $\vdash_S \alpha$ then $\vdash_K \phi\alpha$ and $\vdash_K \psi\alpha$. Now if α is first-degree, $\psi\alpha = \alpha$, so any first-degree theorem of S is a theorem of K; since S extends K, S and K have the same first-degree theorems. Since $K \subseteq S' \subseteq S$, S and S′ have the same first-degree theorems.

An alternative proof of the same result would use Proposition 6 and the fact that K is the logic of the class of irreflexive frames and S the logic of the class of frames <W,R> satisfying the condition that if wRx then not xRy to show that K‡S conservatively □-extends K and conservatively ■-extends S.

PROPOSITION 2. $\Box\blacksquare\Box\top$ is not a theorem of K†K.
PROOF. Define three functions ϕ, ψ and χ from $L_{\Box\blacksquare}$ to $L_{\Box}$ thus:

$$\phi p = \psi p = \chi p = p$$
$$\phi\bot = \psi\bot = \chi\bot = \bot$$
$$\phi(\alpha \supset \beta) = \phi\alpha \supset \phi\beta \quad \psi(\alpha \supset \beta) = \psi\alpha \supset \psi\beta \quad \chi(\alpha \supset \beta) = \chi\alpha \supset \chi\beta$$
$$\phi\Box\alpha = \bot \quad \psi\Box\alpha = \Box\psi\alpha \quad \chi\Box\alpha = \Box\psi\alpha$$
$$\phi\blacksquare\alpha = \Box\phi\alpha \quad \psi\blacksquare\alpha = \Box\phi\alpha \quad \chi\blacksquare\alpha = \Box\psi\alpha$$

$K \subseteq L_{\Box}$, but axiomatize $K\Box \subseteq L_{\Box\blacksquare}$ with all truth-functional tautologies and all formulas of the form $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box b)$ as the axioms and MP and necessitation for □ as the rules of inference. By induction on the length of proofs, if $\vdash_{K\Box} \alpha$ then $\vdash_K \psi\alpha$ and $\vdash_K \chi\alpha$. Similarly, axiomatize K■ in $L_{\Box\blacksquare}$ with all truth-functional tautologies and all formulas of the form $\blacksquare(\alpha \supset \beta) \supset (\blacksquare\alpha \supset \blacksquare\beta)$ as the axioms and MP and necessitation for ■ as the rules of inference. By induction on the length of proofs, if $\vdash_{K\blacksquare} \alpha$ then $\vdash_K \phi\alpha$, $\vdash_K \psi\alpha$ and $\vdash_K \chi\alpha$. Now axiomatize K†K with all theorems of K□ and K■ and formulas of the form $\Box\alpha \equiv \blacksquare\alpha$ as axioms and MP as the only rule of inference. By the previous results and induction on the length of proofs in K†K, if $\vdash_{K\dagger K} \alpha$ then $\vdash_K \chi\alpha$. But $\chi\Box\blacksquare\Box\top = \Box\psi\blacksquare\Box\top = \Box\Box\phi\Box\top = \Box\Box\bot$, and not $\vdash_K \Box\Box\bot$.

If X is a formula or set of formulas, let |X be the result of interchanging □ and ■ throughout X.

PROPOSITION 3. Let $S_1,S_2 \subseteq L_{\Box}$ be modal systems, $w_1 \subseteq L_{\Box}$ a maximal S_1-consistent set and $w_2 \subseteq L_{\Box}$ a maximal S_2-consistent set such that $w_1 \cap L_{PC} = w_2 \cap L_{PC}$. Then $w_1 \cup |w_2$ is an $S_1 \circledcirc S_2$-consistent set.
PROOF. Define mappings ϕ_1 and ϕ_2 from $L_{\Box\blacksquare}$ to $L_{\Box}$ thus:

$$\phi_1 p = \phi_2 p = p$$
$$\phi_1\bot = \phi_2\bot = \bot$$

$\phi_1(\alpha \supset \beta) = \phi_1\alpha \supset \phi_1\beta$	$\phi_2(\alpha \supset \beta) = \phi_2\alpha \supset \phi_2\beta$
$\phi_1\Box\alpha = \Box\phi_1\alpha$	$\phi_2\Box\alpha = \top$ if $\Box\phi_1\alpha \in w_1$
	$\phi_2\Box\alpha = \bot$ otherwise
$\phi_1\blacksquare\alpha = \top$ if $\Box\phi_2\alpha \in w_2$	$\phi_2\blacksquare\alpha = \Box\phi_2\alpha$
$\phi_1\blacksquare\alpha = \bot$ otherwise	

Now define a mapping ψ from $L_{\square\blacksquare}$ to L_{PC} thus:

$$\psi p = p$$
$$\psi\bot = \bot$$
$$\psi(\alpha \supset \beta) = \psi\alpha \supset \psi\beta$$
$$\psi\square\alpha = \phi_2\square\alpha$$
$$\psi\blacksquare\alpha = \phi_1\blacksquare\alpha$$

Set $U = \{\alpha\in L_{\square\blacksquare}: \psi\alpha\in w_1\}$. The plan is to show that U is an $S_1 \circledcirc S_2$-consistent set such that $w_1 \cup |w_2 \subseteq U$.

(i) $U = \{\alpha\in L_{\square\blacksquare}: \psi\alpha\in w_2\}$. Proof: $\psi\alpha\in L_{PC}$ and $w_1\cap L_{PC} = w_2\cap L_{PC}$.

(ii) For $\alpha\in L_{\square\blacksquare}$, $\alpha\in U$ if and only if $\phi_1\alpha\in w_1$. Proof: By induction on the complexity of α.

(iii) For $\alpha\in L_{\square\blacksquare}$, $\alpha\in U$ if and only if $\phi_2\alpha\in w_2$. Proof: By induction on the complexity of α and (i).

(iv) $w_1\subseteq U$. Proof: For $\alpha\in w_1\subseteq L_{\square}$, $\phi_1\alpha = \alpha$; use (ii).

(v) $|w_2\subseteq U$. Proof: For $\alpha\in |w_2\subseteq L_{\blacksquare}$, $\phi_2\alpha = |\alpha$; use (iii) and the fact that $\alpha\in |w_2$ if and only if $|\alpha\in w_2$.

(vi) If $\alpha\in L_{\square}$ and σ is a substitution from $L_{\square}$ to $L_{\square\blacksquare}$, then $\phi_1\sigma\alpha$ is a substitution instance of α. Proof: Let σ' be the substitution such that $\sigma' p = \phi_1\sigma p$ for every propositional variable p. By induction on the complexity of $\alpha\in L_{\square}$, $\sigma'\alpha = \phi_1\sigma\alpha$.

(vii) If $\alpha\in L_{\square}$ and σ is a substitution from $L_{\blacksquare}$ to $L_{\square\blacksquare}$, then $\phi_2\sigma|\alpha$ is a substitution instance of α. Proof: Like that of (vi).

(viii) If $\alpha\in S_1$ and σ is a substitution from $L_{\square}$ to $L_{\square\blacksquare}$, then $\sigma\alpha\in U$. Proof: Since $S_1\subseteq L_{\square}$, $\phi_1\sigma\alpha$ is a substitution instance of α by (vi). Since $\alpha\in S_1$ and S_1 is closed under US, $\phi_1\sigma\alpha\in S_1$. Since w_1 is maximal S_1-consistent, $\phi_1\sigma\alpha\in w_1$. By (ii), $\sigma\alpha\in U$.

(ix) If $\alpha\in S_2$ and σ is a substitution from $L_{\blacksquare}$ to $L_{\square\blacksquare}$, then $\sigma|\alpha\in U$. Proof: Like that of (viii), using (iii) and (vii).

(x) U is $S_1\circledcirc S_2$-consistent. Proof: Since w_1 is PC-consistent, U is closed under MP and does not contain $\bot$; then use (viii) and (ix).

(xi) $w_1\cup |w_2$ is $S_1\circledcirc S_2$-consistent. Proof: By (iv), (v) and (x).

PROPOSITION 4. If $S_1,S_2\subseteq L_{\square}$ are consistent modal systems, then $S_1\circledcirc S_2$ conservatively $\square$-extends S_1 and conservatively $\blacksquare$-extends S_2.

PROOF. Suppose that $\alpha\in L_{\square}$ but $\alpha\notin S_1$. Let $w_1\subseteq L_{\square}$ be a maximal S_1-consistent set such that $\sim\alpha\in w_1$. Since S_2 is a consistent modal system, $w_1\cap L_{PC}$ is S_2-consistent. Let $w_2\subseteq L_{\square}$ be a maximal S_2-consistent extension of $w_1\cap L_{PC}$. By Proposition 3, $w_1\cup |w_2$ is $S_1\circledcirc S_2$-consistent. Since

$\sim\alpha \in w_1$, $\alpha \notin S_1 © S_2$. Thus $S_1 © S_2$ conservatively □-extends S_1. The second half is similar.

PROPOSITION 5. If S_1 is the logic of a class of frames Γ, and for each $<W_1,R_1> \in \Gamma$ and $w \in W_1$ there is a frame $<W_2,R_2>$ for the system S_2 such that $w \in W_2$ and $\{x \in W_1: wR_1x\} = \{x \in W_2: wR_2x\}$, then $S_1 ‡ S_2$ conservatively □-extends S_1.

PROOF. Suppose that $\alpha \in L_\Box$ and not $\vdash_{S1} \alpha$. We show that not $\vdash_{S1‡S2} \alpha$. Since S_1 is the logic of Γ, there are $<W_1,R_1> \in \Gamma$, $w \in W_1$ and a model $<W_1,R_1,V>$ based on $<W_1,R_1>$ such that $V(\alpha,w) = 0$. By hypothesis, there is a frame $<W_2,R_2>$ for S_2 such that $w \in W_2$ and $\{x \in W_1: wR_1x\} = \{x \in W_2: wR_2x\}$. Extend R_1 and R_2 to relations R_1^+ and R_2^+ on $W_1 \cup W_2$ thus:

xR_1^+y just in case either		(a) xR_1y
	or	(b) $x \notin W_1$ and $x = y$ and R_1 is serial on W_1.
xR_2^+y just in case either		(a) xR_2y
	or	(b) $x \notin W_2$ and $x=y$ and R_2 is serial on W_2.

Define a bimodal model $<W_1 \cup W_2, R_1^+, R_2^+, V^+>$ based on the frame $<W_1 \cup W_2, R_1^+, R_2^+>$ by putting $V^+(p,x)=V(p,x)$ for $x \in W_1$ and $V^+(p,x)=0$ for $x \notin W_1$, for each propositional variable p. Now note the following facts:

(i) $<W_1 \cup W_2, R_1^+>$ is a frame for S_1. Proof: If $x \in W_1$, then the subframe of $<W_1 \cup W_2, R_1^+>$ generated by x is the subframe of $<W_1,R_1>$ generated by x; since $<W_1,R_1>$ is a frame for S_1, so is the latter. If $x \notin W_1$ and R_1 is serial, then the subframe of $<W_1 \cup W_2, R_1^+>$ generated by x is $<\{x\},\{<x,x>\}>$; but since R_1 is serial, this is a p-morphic image of any generated subframe of $<W_1,R_1>$, and so is again a frame for S_1. If $x \notin W_1$ and R_1 is not serial, then the subframe of $<W_1 \cup W_2, R_1^+>$ generated by x is $<\{x\},\{\}>$; but since R_1 is not serial, $<W_1,R_1>$ has a generated subframe of the form $<\{y\},\{\}>$, so $<\{x\},\{\}>$ is a frame for S_1. Thus every generated subframe of $<W_1 \cup W_2, R_1^+>$ is a frame for S_1, so $<W_1 W_2, R_1^+>$ itself is a frame for S_1.

(ii) $<W_1 \cup W_2, R_2^+>$ is a frame for S_2. Proof: As for (i).

(iii) $<W_1 \cup W_2, R_1^+, R_2^+>$ is a frame for $S_1 \oplus S_2$. Proof: Routine from (i) and (ii). The attempt to dispense with the completeness assumption by arguing in terms of models rather than frames would break down because there would be no guarantee that substitution instances in $L_{\Box\blacksquare}$ of theorems of S_1 or S_2 would be true in the new model.

(iv) If $<W_1 \cup W_2, R_1^+, R_2^+, V'>$ is a model based on the frame $<W_1 \cup W_2, R_1^+, R_2^+>$ and $\beta \in L_{\Box\blacksquare}$, then $V'(\Box\beta \equiv \blacksquare\beta, w) = 1$. Proof: $\{x \in W_1 \cup W_2: wR_1^+x\} = \{x \in W_1: wR_1x\} = \{x \in W_2: wR_2x\} = \{x \in W_1 \cup W_2: wR_2^+x\}$.

(v) If $<W_1 \cup W_2, R_1^+, R_2^+, V'>$ is a model based on the frame $<W_1 \cup W_2, R_1^+, R_2^+>$ and $\vdash_{S1\ddagger S2} \beta$, then $V'(\beta, w)=1$. Proof: Routine from (iii) and (iv).

(vi) $V^+(\alpha, w)=0$. Proof: The subframe of $<W_1 \cup W_2, R_1^+>$ generated by w is the subframe of $<W_1, R_1>$ generated by w and $V^+(p,x)=V(p,x)$ for $x \in W_1$. Thus, since $\alpha \in L_\Box$, $V^+(\alpha, w)=V(\alpha, w)=0$.

(vii) Not $\vdash_{S1\ddagger S2} \alpha$. Proof: From (v) and (vi).

PROPOSITION 6. Suppose that S_1 is the logic of a class of frames Γ_1, S_2 is the logic of a class of frames Γ_2, for each $<W_1, R_1> \in \Gamma_1$ and $w \in W_1$ there is a $<W_2, R_2> \in \Gamma_2$ such that $w \in W_2$ and $\{x \in W_1: wR_1x\} = \{x \in W_2: wR_2x\}$, and for each $<W_2, R_2> \in \Gamma_2$ and $w \in W_2$ there is a $<W_1, R_1> \in \Gamma_1$ such that $w \in W_1$ and $\{x \in W_1: wR_1x\} = \{x \in W_2: wR_2x\}$. Then $S_1 \ddagger S_2$ conservatively $\Box$-extends S_1 and conservatively $\blacksquare$-extends S_2.
PROOF. That $S_1 \ddagger S_2$ conservatively $\Box$-extends S_1 follows from Proposition 5. That it conservatively $\blacksquare$-extends S_2 is shown by an argument parallel to the proof of Proposition 5.

PROPOSITION 7. Let S be a complete system, and S' the smallest extension of K45 containing all first-degree theorems of S. Then $S \ddagger S'$ conservatively $\Box$-extends S and conservatively $\blacksquare$-extends S' (as does $S \dagger S'$).
PROOF. Let S be the logic of a class of frames Γ, $<W, R> \in \Gamma$ and $w \in W$. Set $W_w = \{w\} \cup \{x: wRx\}$ and define R_w on W_w by:

$$xR_wy \text{ just in case } wRy.$$

Let S' be the smallest extension of K45 containing all first-degree theorems of S.

(i) K45 is valid on $<W_w, R_w>$. Proof: R_w is obviously transitive and euclidean.

(ii) If $\alpha \in L_\Box$ is first-degree and $\vdash_S \alpha$, then α is valid on $<W_w, R_w>$. Proof: Let $<W_w, R_w, V_w>$ be a model based on $<W_w, R_w>$, and fix $x \in W_w$. Define a mapping π from W_w to W_w thus:

$$\pi w = x$$
$$\pi x = w \text{ if } wRw$$
$$\pi x = x \text{ if not } wRw$$
$$\pi y = y \text{ if } y \neq w \text{ and } y \neq x$$

Let $\langle W,R,V\rangle$ be a model based on $\langle W,R\rangle$ such that $V(p,y) = V_w(p,\pi y)$ for each $y \in W_w$ and propositional variable p. Thus if β is $\Box$-free, then $V(\beta,y) = V_w(\beta,\pi y)$. Since the restriction of π to $\{y\colon wRy\}$ $(=\{y\colon \pi wR_wy\})$ is a permutation, $V(\Box\beta,w) = V_w(\Box\beta,\pi w) = V_w(\Box\beta,x)$ for $\Box$-free β. Thus if β is first-degree, then $V(\beta,w) = V_w(\beta,x)$. By hypothesis, α is first-degree, $\vdash_S \alpha$ and $\langle W,R\rangle$ is a frame for S, so $V(\alpha,w) = 1$. Thus $V_w(\alpha,x) = 1$. Since x and V_w were arbitrary, α is valid on $\langle W_w,R_w\rangle$.

(iii) $\langle W_w,R_w\rangle$ is a frame for S′. Proof: From (i) and (ii).

(iv) S‡S′ conservatively $\Box$-extends S. Proof: By (iii), for each $\langle W,R\rangle \in \Gamma$ and $w \in W$ there is a frame $\langle W_w,R_w\rangle$ for S′ such that $w \in W_w$ and $\{x \in W\colon wRx\} = \{x \in W_w\colon wR_wx\}$. By Proposition 5, S‡S′ conservatively $\Box$-extends S.

(v) For each $\alpha \in L_\Box$ there is a first-degree formula 1α such that $\vdash_{K45} \alpha \equiv 1\alpha$. Proof: It is not hard to show that $\vdash_{K45} \Box(\Box\alpha_1 \vee \ldots \Box\alpha_m \vee \Diamond\beta \vee \gamma) \equiv (\Box\alpha_1 \vee \ldots \Box\alpha_m \vee \Diamond\beta \vee \Box\gamma)$ $(m \geq 0)$. By standard manipulations, one can then show that any formula of degree >1 is equivalent in K45 to a formula of lower degree.

(vi) S‡S′ conservatively ■-extends S′. Proof: Suppose that $\alpha \in L_\Box$ and $\vdash_{S\ddagger S'} |\alpha$. What must be shown is that $\vdash_{S'} \alpha$. By (v), $\vdash_{K45} \alpha \equiv 1\alpha$, so $\vdash_{S'} \alpha \equiv 1\alpha$, so $\vdash_{S\ddagger S'} |\alpha \equiv |1\alpha$, so $\vdash_{S\ddagger S'} |1\alpha$. Since $|1\alpha$ is first-degree in ■, $\vdash_{S\ddagger S'} 1\alpha \equiv |1\alpha$, so $\vdash_{S\ddagger S'} 1\alpha$. Since $1\alpha \in L_\Box$, $\vdash_S 1\alpha$ by (iv). Since 1α is first-degree, $\vdash_{S'} 1\alpha$. Since $\vdash_{S'} \alpha \equiv 1\alpha$, $\vdash_{S'} \alpha$.

PROPOSITION 8. Let S_1 and S_2 be complete systems with the same first-degree theorems, and S′ the smallest extension of K45 containing those theorems. If $S_1 \subseteq S'$ and $S_2 \subseteq S'$ then $S_1\ddagger S_2$ conservatively $\Box$-extends S_1 and conservatively ■-extends S_2 (as does $S_1\dagger S_2$).

PROOF. By Proposition 7, $S_1\ddagger S'$ conservatively $\Box$-extends S_1. Since $S_2 \subseteq S'$, $S_1\ddagger S_2$ conservatively $\Box$-extends S_1. By parallel reasoning, $S_1\ddagger S_2$ conservatively ■-extends S_2.

PROPOSITION 9. Let S and S′ be systems such that S′ extends K45 and S†S′ (or S‡S′) conservatively $\Box$-extends S and conservatively ■-extends S′. Then S′ is the smallest extension of K45 containing all first-degree theorems of S.

PROOF. Suppose that $\vdash_S \alpha$ and α is first-degree. Then $\vdash_{S\dagger S'} \alpha$ and $\vdash_{S\dagger S'} \alpha \equiv \mathsf{I}\alpha$, so $\vdash_{S\dagger S'} \mathsf{I}\alpha$. Since S†S′ conservatively ■-extends S′, $\vdash_{S'} \alpha$. Thus S′ is an extension of K45 containing all first-degree theorems of S. Let S″ be any extension of K45 containing all first-degree theorems of S. What must be shown is that $S' \subseteq S''$. Suppose that $\vdash_{S'} \alpha$. By reasoning as for (vi) in the proof of Proposition 7, $\vdash_{S\dagger S'} 1\alpha$. Since $1\alpha \in L_\Box$ and S†S′ conservatively □-extends S, $\vdash_S 1\alpha$. Since 1α is first-degree, $\vdash_{S''} 1\alpha$. Since S″ extends K45, $\vdash_{S''} \alpha \equiv 1\alpha$, so $\vdash_{S''} \alpha$, as required. Thus S′ is the smallest extension of K45 containing all first-degree theorems of S.

PROPOSITION 10. If the system S admits the rule of disjunction, and $\Box\alpha_1 \vee \ldots \vee \Box\alpha_n$ is a theorem of the Solovay extension of S, then $\vdash_S \alpha_m$ for some m $(1 \leq m \leq n)$.

PROOF. Grant the hypothesis of the proposition. Then for some $\beta_1, \ldots, \beta_k$:

$$\vdash_S [(\Box\beta_1 \supset \beta_1) \ \& \ldots \& \ (\Box\beta_k \supset \beta_k)] \supset [\Box\alpha_1 \vee \ldots \vee \Box\alpha_n]$$

Assume without loss of generality that k is the least number for which S has such a theorem. By propositional logic:

$$\vdash_S \Box\beta_1 \vee \ldots \vee \Box\beta_k \vee \Box\alpha_1 \vee \ldots \vee \Box\alpha_n$$

Since S admits the rule of disjunction, either $\vdash_S \beta_m$ or $\vdash_S \alpha_m$ for some m. If the former, then $\vdash_S \Box\beta_m \supset \beta_m$, so that conjunct was redundant, so k was not minimal, contrary to hypothesis.

REFERENCES

Benacerraf, Paul 1967: 'God, the devil and Gödel'. *Monist* 51, pp. 9–32.

Boolos, George 1993: *The Logic of Provability*. Cambridge: Cambridge University Press.

Castañeda, Hector-Neri 1970: 'On knowing (or believing) that one knows (or believes)'. *Synthese* 21, pp. 187–203.

Cresswell, Max 1983: 'Bigelow's semantic nominalism'. *Australasian Journal of Philosophy* 61, pp. 78–80.

Davidson, Donald 1984: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Davies, Martin and Humberstone, Lloyd 1980: 'Two notions of necessity'. *Philosophical Studies* 38, pp. 1–30.

Feferman, Solomon 1960: 'Arithmetization of metamathematics in a general setting'. *Fundamenta Mathematicae* 49, pp. 35– 92.
—— 1982: 'Inductively presented systems and the formalization of metamathematics', in van Dalen, Lascar and Smiley 1982, pp. 95–128.
—— 1990: 'Remark 1', in Gödel 1990, pp. 282–287.
Fine, Kit and Schurz, Gerhard 1996: 'Transfer theorems for multimodal logics' in Copeland, J., ed. *Logic and Reality: Essays on the Legacy of Arthur Prior.* Oxford: Oxford University Press.
Gabbay, Dov 1981: 'An irreflexivity lemma with applications to axiomatizations of conditions on tense frames', in Mönnich 1981.
Gödel, Kurt 1932: 'Über Vollständigkeit und Widerspruchsfreiheit', *Ergebnisse eines Mathematischen Kolloquiums* 3, pp. 12–13. Translated as 'On completeness and consistency' in van Heijenoort 1967, pp. 616–617.
—— 1990: *Collected Works*, vol. 2, ed. Feferman, S. Oxford: Oxford University Press.
Hazen, Allen 1978: 'The eliminability of the actuality operator in propositional modal logic'. *Notre Dame Journal of Formal Logic* 19, pp. 617–622.
Heijenoort, J. van, ed. 1967: *From Frege to Gödel.* Cambridge, MA: Harvard University Press.
Hintikka, Jaakko 1962: *Knowledge and Belief.* Ithaca, NY: Cornell University Press.
Hughes, George and Cresswell, Max 1984: *A Companion to Modal Logic*. London: Methuen.
Humberstone, Lloyd 1986: 'Extensionality in sentence position'. *Journal of Philosophical Logic* 15, pp. 27–54.
—— 1995: 'Singulary extensional connectives: a closer look', typescript.
Kracht, Marcus and Wolter, Frank 1991: 'Properties of independently axiomatizable bimodal logics'. *The Journal of Symbolic Logic* 56, pp. 1469–1485.
Kreisel, Georg 1953: 'On a problem of Henkin's'. *Indagationes Mathematicae* 15, pp. 405–406.
—— 1965: 'Mathematical logic', in Saaty 1965, pp. 95–195.
Löb, Martin 1955: 'Solution to a problem of Leon Henkin'. *The Journal of Symbolic Logic* 20, pp. 115–118.
Lucas, John 1961: 'Minds, machines and Gödel'. *Philosophy* 36, pp. 112–127.
Mönnich, Uwe, ed. 1981: *Aspects of Philosophical Logic*. Dordrecht: Reidel.
Penrose, Roger 1989: *The Emperor's New Mind.* Oxford: Oxford University Press.
—— 1994: *Shadows of the Mind.* Oxford: Oxford University Press.
Putnam, Hilary 1975: *Mind, Language and Reality: Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press.
Reinhardt, William 1986: 'Epistemic theories and the interpretation of Gödel's incompleteness theorems'. *Journal of Philosophical Logic* 15, pp. 427–474.
Saaty, Thomas, ed. 1965: *Lectures on Modern Mathematics*, vol. 3. New York: Wiley.
Salmon, Nathan 1989: 'The logic of what might have been'. *The Philosophical Review* 98, pp. 3–34.
—— 1993: 'This side of paradox'. *Philosophical Topics* 21, pp. 187–197.

Scroggs, Schiller Joe 1951: 'Extensions of the Lewis system S5', *The Journal of Symbolic Logic* 16, pp. 112–120.
Shin, Hyun Song and Williamson, Timothy 1994: 'Representing the knowledge of Turing machines'. *Theory and Decision* 37, pp. 125–146.
Solovay, Robert 1976: 'Provability interpretations of modal logic'. *Israel Journal of Mathematics* 25, pp. 287–304.
Takeuti, Gaisi 1955. 'On the fundamental conjecture of GLC I', *Journal of the Mathematical Society of Japan* 7, pp. 249– 275.
Thomason, Steven 1980: 'Independent propositional modal logics'. *Studia Logic* 39, pp. 143–144.
Van Dalen, D., Lascar, D. and Smiley, T., eds. *Logic Colloquium '80*. Amsterdam: North-Holland.
Williamson Timothy. 1990: *Identity and Discrimination*. Oxford: Blackwell.
—— 1992: 'Inexact knowledge'. *Mind* 101, pp. 217–242.
—— 1998: 'Some computational constraints in epistemic logic', typescript.