Nonmonotonic Logic II: Nonmonotonic Modal Theories

DREW MCDERMOTT

Yale University, New Haven, Connecticut

ABSTRACT Traditional logics suffer from the "monotonicity problem": new axioms never invalidate old theorems One way to get rid of this problem is to extend traditional modal logic in the following way The operator M (usually read "possible") is extended so that Mp is true whenever p is consistent with the theory Then any theorem of this form may be invalidated if $\sim p$ is added as an axiom This extension results in nonmonotonic versions of the systems T, S4, and S5 These systems are complete in that a theorem is provable in a theory based on one of them just if it is true in all "noncommittal" models of that theory, where a noncommittal model is one in which as many things are possible as possible Nonmonotonic S4 is probably the most interesting of the three, since it is stronger than ordinary S4 but has all the usual inferential machinery of S4 There is a straightforward proof procedure for the sentential subset of nonmonotonic S4.

This approach to nonmonotonic logic may be applied to several problems in knowledge representation for artificial intelligence Its main advantages over competing approaches are that it factors out problems of resource limitations and allows the symbol M to appear in any context, since M is a meaningful part of the language

Categories and Subject Descriptors: F.4 1 [**Mathematical Logic and Formal Languages**]. Mathematical Logic—*mechanical theorem proving*, I 2 3 [**Artificial Intelligence**] Deduction and Theorem Proving—*nonmonotonic reasoning and belief revision*, I 2 4 [**Artificial Intelligence**]. Knowledge Representation Formalisms and Methods

General Terms. Theory

Additional Key Words and Phrases: modal logic, completeness theories, proof procedures

1. *Introduction*

In a previous paper [8], Jon Doyle and I investigated the properties of a nonmonotonic predicate calculus. The word "nonmonotonic" refers to first-order theories in which new axioms can wipe out old theorems. The way we accomplished this was by using the inference rule,

$$\text{infer} \quad Mp \quad \text{from the inability to infer} \quad \sim p,$$

where M is an operator (read "consistent") which forms formulas out of other formulas. In this rule, Mp is to mean "p is consistent with the theory." Then M is used in proper axioms like,

$$(\forall X)(\text{BIRD}(X) \wedge M\ \text{CAN-FLY}(X) \supset \text{CAN-FLY}(X))$$

which is intended to capture the idea, "Most birds can fly." I will recapitulate exactly how this rule works shortly.

This work was supported by the National Science Foundation under Contract MCS 78-03599

Author's address Department of Computer Science, Yale University, P O Box 2158, New Haven, CT 06520

The main problem with our treatment was that it provided a very weak semantics. For example, even though $\sim M \sim p$ (abbreviated Lp) might plausibly be expected to mean "p is provable," there was not actually any relation between the truth values of p and Lp. The reason the semantics was so weak was that there were no other inference rules or axioms about L and M.

In this paper I will rectify this by supplying some rules and axioms. These are just the rules and axioms of "standard" modal logic (i.e., modal logic with Kripke semantics). With this addition we can get a clean semantics for nonmonotonic logic. This will give us completeness and some other nice results. At the end I will make some comments on applications and comparisons with related work.

2. *First-Order Modal Theories*

For our purposes the standard treatment of modal logic is not quite adequate. In contrast to nonmodal predicate calculus, where is it common to add nonlogical axioms to the bare logical machinery, modal logicians usually stick to simple systems like T, S4, S5, etc. [6]. For artificial-intelligence applications we need more meat on the bones, including a vocabulary of domain-dependent symbols and lots of (somewhat banal) axioms like, "The block is on the table."

So I define a first-order modal theory as follows. First we need a language Lang, defined in the usual way: there is a supply (possibly countably infinite) of predicate symbols, constant symbols, function symbols, and variable symbols. A *term* is a constant symbol, a predicate symbol, or an expression $f(t_1, \ldots, t_n)$, where f is a function symbol and $t_1, \ldots, t_n$ are terms. I use the symbol Trm to refer to the set of all terms. An *atomic formula* is an expression $P(t_1, \ldots, t_n)$, where P is a predicate symbol and $t_1, \ldots, t_n$ are terms. A *formula* is either an atomic formula; an expression $\sim p$, where p is a formula; an expression $p \supset q$, where p and q are formulas; an expression Mp, where p is a formula; or an expression $(\forall v)p$, where v is a variable and p is a formula. I use the symbol Lang to refer to the set of all formulas. We have the usual abbreviations: $p \vee q$ for $(\sim p) \supset q$; $p \wedge q$ for $\sim((\sim p) \vee (\sim q))$; $p \leftrightarrow q$ for $(p \supset q) \wedge (q \supset p)$; $(\exists v)p$ for $\sim(v)\sim p$; and Lp for $\sim M \sim p$.

Metanotation. Throughout most of the paper the letters p, q, and r denote formulas of the object language. The letter v denotes variables. The letters C, D, and E denote variable-free formulas. These may be subscripted. All set-theoretic apparatus that you see is metalinguistic. In Section 7 a more casual notation will be adopted, and some set notation will occur in the object language. Everywhere in the paper, if a metaexpression denoting a finite set occurs in a formula, it stands for the conjunction of its elements; if it is empty, it stands for some tautology.

Now we define a *first-order modal theory* as a bunch of *proper axioms* plus logical axioms and inference rules. (Sometimes I will use the term "theory" to mean just a set of axioms or even just a set of proper axioms, the rest of the machinery being derivable from the context.)

We get different kinds of modal theory by varying the logical axioms. The logical axioms will always include the tautologies of the first-order predicate calculus. In addition, each theory will contain all instances of various subsets of the following axiom schemata:

AS1: $Lp \supset p$
AS2: $L(p \supset q) \supset (Lp \supset Lq)$
AS3: $((\forall v)Lp) \supset L(\forall v)p$
AS4: $Lp \supset LLp$
AS5: $Mp \supset LMp$

The inference rules are always

Modus Ponens (MP): $p, p \supset q \vdash q$
Universal Generalization (UG): $p \vdash (\forall v)p$
Necessitation (Nec): $p \vdash Lp$

We get various traditional modal systems by taking particular subsets [6]. The weakest system is called T and contains (instances of) axiom schemata AS1 and AS2. Adding AS4 gives S4; adding AS4 and AS5 gives S5. Adding AS3 to T gives T + BF; adding AS3 to S4 gives S4 + BF. (Adding AS3 to S5 does nothing, since AS3 is inferable from the other schemata.) In this paper I will always include AS3 and will refer to T + BF and S4 + BF as just T and S4.

These axiom schemata and inference rules are intended to be a plausible account of the logic of "is-consistent." (The account is incomplete until the nonmonotonic rule is given.) So the logic is supposed to describe provability in itself. The basis for this attempt is the axioms of predicate calculus and the traditional inference rules UG and MP. These things are necessary to get started. AS1 to AS4 and Nec are aimed at capturing the properties of provability. In particular, it seems essential to the concept of provability that something proven be provable, and this is what Nec says.

AS1 says that everything provable is true. This may seem optimistic. But what I am trying to get it to mean is: If p is necessarily true when the proper axioms are, then it is true when the proper axioms are. At any rate, it is difficult to visualize any other way of relating provability and truth.

AS2 describes the operation of the rule MP; that is, where MP allows you to infer q from $p \supset q$ and p, AS2 *says* that this is allowed. AS3, the "Barcan formula," describes how UG works in much the way that AS2 describes how MP works. That is, if every instance of a formula is provable, then its universal closure is provable.

AS4 and AS5 describe the theory in a more global way. AS4 says that p is provable only if it is provably provable. That is, the concept of proof is, in a sense, "public."

AS5 makes a more breathtaking assertion, that p is *un*provable only if it is provably unprovable. At first glance this is quite implausible, since in most interesting systems it makes the the concept of "proven" totally undecidable. This assertion would not be worth making, except that in nonmonotonic systems it is true.

The reason why I study a variety of modal systems is that they are all closely related, and no one is obviously better than the others.

3. *Nonmonotonic Inference*

We now make the system nonmonotonic by adding the rule of

Possibilitation (Pos): "(Can't infer $\sim p$) $\vdash Mp$"

The result is a first-order nonmonotonic modal theory.

I put scare quotes around the rule because, of course, it is ill-formed. An inference rule states a relation between sets of formulas. It is meaningless to insert a comment about inference in the left-hand side. Besides, even if we could, this would only make the definition of "infer" circular.

We must find a more elaborate definition of nonmonotonic inferability. As in [8], we define

$$\text{set1} \mid\!\sim \text{set2}$$

to mean

$$\text{set2} \subseteq \text{TH}(\text{set1}),$$

where

$$\mathrm{TH}(A) = \mathrm{Lang} \cap (\bigcap \{X : X = \mathrm{NM}_A(X)\}),$$

the intersection of all *fixed points* of A, or the whole language Lang if there are no fixed points.

NM_A is defined as

$$\mathrm{NM}_A(B) = \mathrm{Th}(A \cup \mathrm{As}_A(B))$$

where

$$\mathrm{Th}(S) = \{p : S \vdash p\}$$

and

$$\mathrm{As}_A(B) = \{Mq : q \text{ is a statement (i.e., has no free variables) and } {\sim}q \notin B\} - \mathrm{Th}(A).$$

Note that if q is not in $\mathrm{Th}(A \cup B)$, $M{\sim}q$ is in $\mathrm{As}_A(B)$ because ${\sim}{\sim}q$ is not in $\mathrm{Th}(A \cup B)$.

In other words, $\mathrm{Th}(A)$ is the set of theorems of the modal theory with proper axioms A. $\mathrm{As}_A(B)$ is the set of assumptions allowed by B in the modal theory with proper axioms A. $\mathrm{NM}_A(B)$ is the set of theorems of the modal theory with proper axioms $A \cup \mathrm{As}_A(B)$. A fixed point of NM_A is a set X such that $X = \mathrm{NM}_A(X)$. Such a fixed point is a set X containing A and a large set of assumptions $\mathrm{As}_A(X)$, such that no assumption Mp in X is wiped out by ${\sim}p$ being provable from X, and every other element of X has a proof from assumptions and axioms. $\mathrm{TH}(A)$ is the set of all formulas that are in all fixed points. I will continue to talk as if these formulas were "derived by use of the rule Pos," but this just means that they are in the set of theorems defined by this fixed-point construction.

Notice that if there are no fixed points, every formula in the language Lang is provable. Also, if Lang is a fixed point, then, as we proved in [8, Th. 4], it is the only fixed point. Either way, $\mathrm{TH}(A) = \mathrm{Lang}$, and the theory is said to be *inconsistent*.

Except for Th now referring to modal provability, these are the same definitions Doyle and I used in our earlier paper [8]. That paper contained several examples, all of which still work in this "stronger" logic. In what follows, the symbol $\vdash$ denotes inferability in a monotonic system under discussion. When it is necessary to be specific, I will use $\vdash_{\mathrm{T}}$, $\vdash_{\mathrm{S4}}$, or $\vdash_{\mathrm{S5}}$ to denote inferability in a particular kind of theory. The term "consistent" will be modified with the prefixes T-, S4- and S5- in an analogous way. Similarly, all of the symbols Th, NM, TH, and As were defined with respect to an unspecified relation $\vdash$. I will prefix these symbols too with T, S4, or S5 when necessary. For example, S4-TH(A) is defined to be

$$\mathrm{Lang} \cap (\bigcap \{X : X = \text{S4-NM}_A(X)\}).$$

An example will show that this is actually easier to think about than it looks. Consider the S4-based theory with proper axioms

$(\forall X)(\mathrm{BIRD}(X) \wedge M\ \text{CAN-FLY}(X)$
$\quad \supset \text{CAN-FLY}(X))$
$(\forall X)(\mathrm{OSTRICH}(X) \supset (\mathrm{BIRD}(X) \wedge {\sim}\text{CAN-FLY}(X)))$
BIRD(FRED)
OSTRICH(OZZIE)

This theory has one fixed point, which contains the following formulas in addition to the axioms:

BIRD(OZZIE)
~CAN-FLY(OZZIE)
M CAN-FLY(FRED)
CAN-FLY(FRED)

The first two of these follow by predicate calculus. The third follows because ~CAN-FLY(FRED) is *not* a member of the fixed point. In other words, *M* CAN-FLY(FRED) is in $\mathrm{As}_{\mathrm{theory}}$(fixed-point). So by the first proper axiom, CAN-FLY(FRED) is in the fixed point as well.

Of course, I have not proven that this is a fixed point, or the only fixed point. (As I will show below, if we had based the system on S5, there would be other fixed points.) In general, it is undecidable whether a formula appears in any fixed point or all of them. A computer program using nonmonotonic inference must always be prepared to withdraw some conclusion (cf. [2]).

It might appear that we do not need to worry about the distinction between T, S4, and S5, since we have the following theorem.

THEOREM 1. *All closed instances of AS1, AS2, AS3, AS4, and AS5 are derivable from PC plus Nec and Pos, no matter what other proper axioms are present.*

PROOF. Let A be a set of proper axioms plus the axioms of PC. If A has no fixed points, the theorem is obvious. Otherwise, let X be any fixed point of A.

(a) *Proof of AS1.* For every formula p, either p is in X or it is not. If it is, then by the statement calculus (SC), $Lp \supset p$ is in X. If p is not in X, $M{\sim}p$ is in X. So ${\sim}M{\sim}p \supset p$ is in X, but this is just $Lp \supset p$. Either way, this instance of AS1 is in X, so every instance is in every fixed point.

(b) *Proof of AS2.* If q is in X, then by Nec, Lq is in X; so, by SC, $L(p \supset q) \supset (Lp \supset Lq)$ is in X. If p is not in X, then $M{\sim}p$ is in X, so ${\sim}M{\sim}p \supset Lq$ is in X, so by SC, $L(p \supset q) \supset (Lp \supset Lq)$ is in X. If p is in X and q is not in X, then $p \supset q$ is not in X. So $M{\sim}(p \supset q)$ is in X. So by SC, $L(p \supset q) \supset (Lp \supset Lq)$ is in X. In every case, this instance of AS2 is in X, so every instance is in every fixed point.

(c) *Proof of AS4.* Very similar. If p is in X, then LLp is in X by two applications of Nec. Otherwise, $M{\sim}p$ is in X.

(d) *Proof of AS5.* Very similar. If ${\sim}p$ is not in X, then Mp is in X, and, by Nec, so is LMp. If ${\sim}p$ is in X, then $L{\sim}p$ is too, by Nec, and so is ${\sim}Mp$. (This last step requires the theorem $L{\sim}p \supset {\sim}Mp$, which follows from AS1 and AS2 [1, 6].)

(e) *Proof of AS3.* This follows from the others (see [6, p. 145]. Q.E.D.

Unfortunately, in nonmonotonic systems it is not enough to show that a formula is a theorem for it to be dispensable as an axiom. It can actually happen that p is in TH(A) but TH($A \cup \{p\}$) differs from TH(A). It is shown in [8, Th. 8] that some consistent theories have inconsistent subtheories. In a case like that, every p is in TH(A) (where A is inconsistent), but for some such p, $A \cup \{p\}$ is consistent. So, even though p is a theorem of A, $A \cup \{p\}$ gives different theorems.

Inconsistency does not have to be involved. Consider the theory T1 lacking Nec and AS1–AS5 (i.e., the logic of our earlier paper), with the proper axioms

$$MC \supset {\sim}D, \qquad MD \supset {\sim}C.$$

(The letters C, D, and E are propositional constants.) This theory has two fixed points, one with MC and ${\sim}D$, the other with MD and ${\sim}C$ [8].

Now add the axiom

$${\sim}C \supset {\sim}M{\sim}E$$

to get the theory T2. In every fixed point of this theory, E is not a theorem. (This may be verified by the procedure given in [8].) So $M{\sim}E$ is a theorem; so C is a theorem; so ${\sim}MD$ is a theorem. So the fixed point with MD no longer exists. So ${\sim}D$ is a theorem of T2.

Notice that ${\sim}M{\sim}E \supset E$ is a theorem of T2 by the argument in Theorem 1. If we add ${\sim}M{\sim}E \supset E$ to the axioms to get theory T3, then E is a theorem in the fixed point with MD and ${\sim}C$; so $M{\sim}E$ is blocked, and both fixed points come back to life. So ${\sim}D$ is not a theorem of T3.

The best we can do is to prove that adding a theorem as an axiom only makes some theorems go away. That is, every theorem of $A \cup$ {axiom} is a theorem of A. This follows because every fixed point of the original theory is a fixed point of the enlarged theory, just not vice versa.

The unhappy consequence of all this is that to show that a formula is a useful lemma of a nonmonotonic system, one must show that adding it to the axioms does not change the provable theorems. As we have just seen, this is not true for AS1–AS5; we must make them axioms.

In spite of the anticlimax, it was actually possible to derive these strong axiom schemata from nothing, using Nec and Pos. This raises the fear that the Pos rule is too strong; the nonmonotonic logics might all be inconsistent. Later I will show that they are not.

4. *Semantics of Monotonic Modal Theories*

Before giving the semantics for nonmonotonic logic, I must give the semantics for first-order modal theories.

A *modal interpretation* is a tuple $\langle W, \text{alt}, D, V \rangle$. W is a set of possible worlds, and alt is a reflexive relation on W, called the *alternativeness relation*. w_1 alt w_2 means w_2 is an alternative to w_1, that is, w_2 is possible with respect to w_1. D is a domain of objects, not empty and perhaps uncountable. Let Trm aug D be the set of terms obtained by adding D to the set of constants. Let Lang aug D be the language Lang, obtained by using Trm aug D instead of Trm in the definition of Lang.

V is a function from (Lang aug D) × W to {0, 1} which gives the truth value of every expression in the language in every possible world. We have the following constraints on V:

$V({\sim}p, w) = 1$	iff	$V(p, w) = 0$;
$V(p \supset q, w) = 1$	iff	$V(p, w) = 0$ or $V(q, w) = 1$;
$V((\forall v)p, w) = 1$	iff	$V(subst(d, v, p), w) = 1$ for all d in Trm aug D (subst(d, v, p) is the result of substituting term d for variable v in formula p);
$V(Mp, w) = 1$	iff	$V(p, u) = 1$ for some u such that w alt u.

I will use the abbreviation $V(p) = x$ to mean $V(p, w) = x$ for all w in W. If $V(p, w_1) \neq V(p, w_2)$, then $V(p) = -1$. If $V(p, w) = 1$, p is *true in world* w, else *false*. If $V(p) = 1$, then p is *true in* V, else *not true*. A set of formulas S is true in a world or interpretation if all its elements are. (I will usually just use the letter V to refer to an interpretation, writing W_V, alt_V, and D_V when I need to refer to W, alt, and D.)

If alt_V is transitive, then V is an S4-*interpretation*. If alt_V is transitive and symmetric, then V is an S5-*interpretation*. In any case, it is a *T-interpretation*.

A *modal model* of a first-order modal theory is a modal interpretation V such that $V(p) = 1$ for every proper axiom p of the theory. As before, we can distinguish T-modal models, S4-modal models, and S5-modal models.

In this section I will prove the completeness of modal theories, that is, that a formula is true in all modal models of a theory iff it is provable in that theory. This result is not particularly unexpected and has gone unproven so far only because logicians' interests have lain elsewhere. The proof is a little tedious, so you may want to just read Theorem 8 and proceed to the next section.

First, some definitions.

Define L^nS, where n is a nonnegative integer and S is a set of formulas, as follows: L^0S is S. $L^{i+1}S$ is $(L^iS) \cup \{Lq : q \in L^iS\}$.

Define $L^\omega S$ to be the union of all L^iS: $\bigcup\{L^iS : i \geq 0\}$.

Define $s1 \vdash_{PC} s2$ to mean, "$s2$ is inferable from $s1$ using predicate calculus alone" (i.e., using MP and UG but not Nec).

Define $s1 \vdash_{Nec} s2$ to mean, "$s2$ is inferable from $s1$ using predicate calculus and Nec."

In this section T means all instances of axiom schemata AS1–AS3.

LEMMA 2. *If $T \subseteq A$ and $A \vdash_{PC} p$, then $L^1A \vdash_T Lp$.*

PROOF. By induction on the number of applications of MP and UG. If there are no applications of MP and UG, then the theorem is obvious. Otherwise, assume it works for i applications. Let Prf be a proof with $i + 1$ applications.

Case 1. The last application is an application of MP to q and $q \supset r$. By the induction hypothesis, there is a proof of Lq and $L(q \supset r)$ from L^1A. But since $T \subseteq A$, $L(q \supset r) \supset (Lq \supset Lr)$ is in A and hence L^1A. So, by PC there is a proof of Lr from L^1A.

Case 2. The last application is an application of UG to q, using variable v. By the induction hypothesis, there is a proof of Lq from L^1A. So, using UG, there is a proof of $(\forall v)Lq$. But since $T \subseteq A$, $((\forall v)Lq) \supset L(\forall v)q$ is in A and hence L^1A. So, by PC there is a proof of $L(\forall v)q$ from L^1A. Q.E.D.

LEMMA 3. *If $T \subseteq A$, then $A \vdash_{Nec} p$ iff there is a finite subset S of $L^\omega A$ such that $S \vdash_{PC} A$. That is, there is a T proof of p from A iff there is a proof from S that does not use necessitation.*

PROOF. *If.* Let S be such a subset. Then $A \vdash_{Nec} S$ by repeated applications of Nec. So $A \vdash_{Nec} p$.

Only if. Assume $A \vdash_{Nec} p$. I will prove $L^\omega A \vdash_{PC} p$ by induction on the number of applications of Nec in the shortest proof of p from A. If a proof contains no applications of Nec, then the theorem is obvious. Otherwise, assume it is true for all proofs with i or fewer applications. Let Prf be a proof with $i + 1$ applications. The last application is to a formula q. By the induction hypothesis, there is a proof Prf′ of q from $L^\omega A$ that does not use Nec. Hence, by Lemma 2, there is a proof of Lq from $L^1L^\omega A$. But $L^1L^\omega A$ is the same as $L^\omega A$. Q.E.D.

LEMMA 4. *V is a modal model of A iff V is a modal model of $L^\omega A$.*

PROOF. *If.* Obvious.

Only if. Let V be a modal model of A. We will prove it is a modal model of L^iA, for all $i \geq 0$, by induction on i. For $i = 0$ it is obvious. Assume that V is a modal model of L^iA. Now consider Lp, where $p \in L^iA$. Let w_0 be in W_V. Since $V(p) = 1$, $V(p, w_1) = 1$ for all w_1 such that $w_0 \text{ alt}_V w_1$. So $V(Lp, w_0) = 1$. So $V(q, w) = 1$ for all w in W_V and q in $L^{i+1}A$. Q.E.D.

LEMMA 5. *If $A \vdash_{Nec} p$ and $T \subseteq A$, then p is true in all modal models of A.*

PROOF. Assume $A \vdash_{\text{Nec}} p$. Let $Supp(p)$ be some finite subset of $L^{\omega}(A - T)$ such that $Supp(p) \cup L^{\omega}T \vdash_{\text{PC}} p$. (This is guaranteed to exist by Lemma 3.) By the deduction theorem for PC, $L^{\omega}T \vdash_{\text{PC}} Supp(p) \supset p$. But then $\vdash_{\text{T}} Supp(p) \supset p$. So every modal model of $Supp(p)$ is a model of p. But every modal model of A is a modal model of $L^{\omega}A$ (by Lemma 4), hence of $Supp(p)$. So p is true in every modal model of A. Q.E.D.

LEMMA 6. *If X is a (T-, S4-, S5-) consistent set of formulas, then there is a (T-, S4-, S5-) modal interpretation U_X with a world w_X such that for all q in X, $U_X(q, w_X) = 1$.*

PROOF. This is essentially the same as [6, Th. 4, p. 169]. The only difference is that the theorem given there is for a consistent single formula rather than a set. But the proof does not depend on the finiteness of the formula. It is a Henkin proof that extends T (or S4 or S5) plus the formula to a structure of maximal consistent sets, each corresponding to a world of the desired model. The first step is to extend the given formula to a maximal (PC-) consistent set; this step can be taken just as easily for an arbitrary consistent set of formulas. After that the proof proceeds as before. Q.E.D.

LEMMA 7. *If $A \nvdash_{(T,S4,S5)} p$, there is a (T-, S4-, S5-) modal model V of A and a world w_0 in W_V such that $V(p, w_0) = 0$.*

PROOF. If $A \nvdash_{(\text{T,S4,S5})} p$, then $L^{\omega}A \nvdash_{(\text{T,S4,S5})} p$ (since $A \vdash_{\text{Nec}} L^{\omega}A$). So the set $X = L^{\omega}A \cup \{\sim p\}$ is (T-, S4-, S5-) consistent. By Lemma 6 there is a (T-, S4-, S5-) interpretation U_X with world w_X (call it w_0) such that for all q in X, $U_X(q, w_0) = 1$. Let V be the subinterpretation of U_X obtained by discarding all worlds not accessible in a finite number of alt links from w_0. It remains a legal interpretation, since if Mq was true in a world before, the adjacent alternative world that made it true is still there. V is a model of $L^{\omega}A$, since a world i alt links from w_0 will have $L^{\omega}A$ true by virtue of the truth of $L^iL^{\omega}A$ in w_0. By Lemma 4, V is a model of A, with $V(\sim p, w_0) = 1$, and hence $V(p, w_0) = 0$. Q.E.D.

THEOREM 8 (COMPLETENESS OF MODAL THEORIES). *p is true in all modal models of the set of formulas A iff $A \vdash p$.*

PROOF. *If.* Lemma 5.

Only if. Contrapositive of Lemma 7. Q.E.D.

5. *Semantics of Nonmonotonic Modal Theories*

Simple modal models are not adequate for nonmonotonic systems. For example, with no proper axioms, M COLOR(BLOCK1, RED) is a theorem. That is, M COLOR(BLOCK1, RED) $\in$ TH($\varnothing$). Clearly, though, it is not a theorem of S4, as reflected by the existence of modal models in which M COLOR(BLOCK1, RED) is false, and hence $L\sim$COLOR(BLOCK1, RED) is true.

As usual in logical studies, our goal is to strengthen a logic by ruling some of its models out. The usual way of doing this is to change the rules of semantic interpretation so that fewer states of affairs qualify as making the desired formulas false. Unfortunately, this local approach is not going to work.

In fact, to find anything like models for a nonmonotonic logic, we cannot get away with tinkering with interpretation rules but must do considerable violence to the entire classical notion of model. A model traditionally captures the structure of the meanings of statements in a language by saying what states of affairs would make various expressions true. Originally, it was taken for granted that the truth value of an expression depended on the truth values of its subexpressions. This intuition led philosophers like Quine to question whether any non-truth-functional semantics could possibly make sense. Twenty years ago, a revolution led by Saul Kripke changed intuition. The modal models he introduced (essentially those I described in the previous section) specify truth values only indirectly, through possible worlds. This led to so many insights into the semantic structure of modal logics that it was obviously valuable.

If we are to have a semantic analysis of nonmonotonic logics, we must attempt another wrenching of existing intuitions. This is because even the most bizarre Kripke-type model imposes certain properties that fail for nonmonotonic systems. One of these is the property of *semantic locality*. In "standard" model theory, if S1 is the set of all models of theory T1, and S2 is the set of all models of theory T2, then S1 ∩ S2 is the set of all models of T1 ∪ T2. That is, to add constraints to a set of formulas is just to throw away the models not compatible with the constraints.

Semantic locality must fail for nonmonotonic logic. For example, let T1 be $\{MC \supset C\}$ and T2 contain $\sim C$. Our notion of model must be such that all models of T1 have C true and all models of T2 have C false. But this does not mean that T1 ∪ T2 has no models.

As with Kripke's alteration to classical semantic notions, the alterations we make to accommodate nonmonotonic logic will ultimately be justified by whether they lead to technical results and new insights. Some promising samples are presented in what follows. For another way to adapt the classical definition of model, see [7].

My approach begins with the definition of the (T-, S4-, S5-) *accidentals* of V with respect to theory A:

$$\text{(T-, S4-, S5-) acc}(V, A) = \{Mp : p \text{ is a statement in Lang}, V(Mp) = 1, \text{ and some model } V' \text{ of } A \text{ has } V'(Mp) \neq 1\}.$$

In other words, the accidentals are the possible statements which do not have to be.

Now the new kind of model can be defined.

A *noncommittal model* V of a theory A is a modal model of A such that Mp is true in all worlds of V whenever p is true in any world of any model of $A \cup \text{acc}(V, A)$.

Actually, this defines three kinds of noncommittal models, prefixed with T, S4, and S5, as usual, for each kind of modal model and its accompanying kind of accidental. For example, consider the S4-based theory with axiom $MC \supset C$. For any model of this theory to be noncommittal, it must have MC true, because *some* S4-model of this theory has it true. Hence, C is true in all noncommittal models and is a theorem. The reason I call this set of models "noncommittal" is that it excludes models with unfounded necessities like $L{\sim}C$ in any world. Something is allowed to be necessary only if its necessity is a logical consequence of the accidental possibilities. This gives us nontrivial logical truths of the form Mp, something previous modal logics have lacked. (In what follows, I will (somewhat glibly) refer to this as the property of "having as many things possible as possible," but this does not imply that noncommittal models literally have a higher count of possibilities. For one thing, having MC true means that LMC is also true, and hence $M{\sim}MC$ is ruled out.)

Do these models shed light on nonmonotonic logics? I will discuss this question at some length in the rest of the paper. First, I will make the technical foundation secure by proving the following theorem.

THEOREM 9 *V is a (T-, S4-, S5-) modal model of a fixed point of NM_A iff V is a (T-, S4-, S5-) noncommittal model of A.*

From this completeness will follow easily.

PROOF. In this proof I drop the prefixes, it being understood that throughout the proof phrases like "modal model" mean "T-modal model" or one of the other models.

Only if. Assume V is a modal model of S, $\text{NM}_A(S) = S$. Let V' be a modal model of $S \cup \text{acc}(V, S)$ with $V'(p, w) = 1$. Then by Theorem 8, $S \cup \text{acc}(V, S) \nvdash \sim p$. Since S is a fixed point of NM_A, $Mp \in S$, and $V(Mp) = 1$. So V is noncommittal.

If. Let V be a noncommittal model of A. I will prove that $\text{Th}(A \cup \text{acc}(V, A))$ is a fixed point of NM_A, from which it will follow that V is a model of the fixed point $\text{Th}(A \cup \text{acc}(V, A))$, since V is a model of $A \cup \text{acc}(V, A)$.

Here goes. We want to show that

$$\text{NM}_A(\text{Th}(A \cup \text{acc}(V, A))) = \text{Th}(A \cup \text{acc}(V, A)).$$

This will follow from

$$\text{As}_A(\text{Th}(A \cup \text{acc}(V, A))) = \text{acc}(V, A).$$

First, to show that $\text{acc}(V, A) \subseteq \text{As}_A(\text{Th}(A \cup \text{acc}(V, A)))$, let Mp be an element of $\text{acc}(V, A)$. Then

$$A \cup \text{acc}(V, A) \nvdash \sim p.$$

So

$$Mp \in \text{As}_A(\text{Th}(A \cup \text{acc}(V, A))).$$

Second, to show that $\text{As}_A(\text{Th}(A \cup \text{acc}(V, A))) \subseteq \text{acc}(V, A)$, assume that $Mp \in \text{As}_A(\text{Th}(A \cup \text{acc}(V, A)))$. By definition of As_A, $\sim p \notin \text{Th}(A \cup \text{acc}(V, A))$ and $Mp \notin \text{Th}(A)$. Then by Theorem 8 there is a model V' of $A \cup \text{acc}(V, A)$ with a world w_0 such that $V'(\sim p, w_0) = 0$, or $V'(p, w_0) = 1$. So, since V is noncommittal, $V(Mp) = 1$. But $Mp \notin \text{Th}(A)$, so some model V' of A has $V'(Mp) \neq 1$. Therefore $Mp \in \text{acc}(V, A)$. Q.E.D.

A corollary is

THEOREM 10. *p is true in all (T-, S4- S5-) noncommittal models of A iff $A \vdash_{(T, S4, S5)} p$.*

PROOF. Fairly obvious. Given Theorem 9, this proof follows the proof of [8, Th.1 and 2]. Q.E.D.

In the proof of Theorem 10, Theorem 9 plays the same role the definition of truth played in the earlier paper. In that paper we had no way of defining truth without mentioning fixed points of theories. This led to rather unrevealing semantics. Now we are in a better position. Intuitively, a noncommittal model is one in which as many things are possible as possible. The second occurrence of "possible" here is a "metalevel above" the first and makes sense only if you are willing to contemplate the totality of models of a theory. The totality is of ordinary models, so there is no circularity, but there is an uncomfortable kind of holism. Just as in our weaker logic, there is no way to define the meaning of individual expressions by reference only to the meanings of their parts. Put more positively, we have succeeded in avoiding the semantic locality condition.

Pat Hayes has made an important observation about formal semantics. When people are first confronted by semantic rules like those at the beginning of Section 4, they are likely to be unimpressed. It looks as if formal semantics is merely translating formulas from a formal language into a very similar informal one. One is tempted to identify the interpretation with this trivial rewrite system. In fact, the rules merely *describe* the flesh-and-blood things that can serve as interpretations; they allow an infinite variety of different interpretations. One thing they allow is the entire real world. In fact, the real world would have to be a model of the database of a robot which was accurate in its beliefs. This is a substantial strength of classical semantics, which should serve as a test for any proposed departure.

Could the real world be a model of a nonmonotonic theory? The same question comes up in evaluating Kripke's earlier revision of classical semantics. Here the presence of all those possible worlds requires us to translate Hayes's observation to the following:

> A robot is accurate in all its beliefs if and only if the real world is one of the possible worlds in a modal model of the robot's database

In this context I use "modal" loosely, intending to include things like tense logics as well. Then one application of this principle might be: "A robot believes 'Always *p*' truly if and only if *p* is true in the present moment, all past moments, and all future moments." Here past and future states of this world fill the role of possible worlds [11].

Let us try to adapt this for the nonmonotonic case. We want the real world to be one of the possible worlds in a *noncommittal* model of the robot's database. Then everything the robot believes must be true in that model. Now an interesting issue comes up regarding what we mean by "belief." For first-order and modal theories we unconsciously equated belief with theoremhood. This is a routine idealization of the notion of belief; the problems it engenders (such as the impossibility of a person's not believing some consequence of Peano's postulates) are not relevant here (see [5]). For nonmonotonic logic we have a choice of idealizations. We can have a *cautious* robot that believes only the theorems of some nonmonotonic theory, or a *brave* robot that believes all the formulas in some fixed point of such a theory. These are both idealizations, because a finite robot can have explored only a finite part of either set. In the brave case we must imagine that the robot has a way of choosing among formulas that would in the limit choose exactly the formulas of a fixed point.

Either way, the real world could be (one world of) a noncommittal model of what the robot believes. But the brave case leads to a more intuitive description of the semantics. We want Mp to mean, "p is true in some possible world consistent with what the robot believes." If the robot's beliefs are identified with a fixed point of a theory, then this is correct: everything not ruled out by the fixed point is consistent with it and hence is consistent in the noncommittal model (since a noncommittal model just *is* a model of a fixed point).

In the cautious case we do not have such an intuitive characterization. Consider, for example, the theory T1 with proper axioms

$$MC \supset {\sim}D, \qquad MD \supset {\sim}C.$$

Assume that some robot cautiously believes T1, and he is accurate in his beliefs: the real world is (one world of) a noncommittal model of T1. Since the theorems of T1 are taken as what the robot believes, then both C and D are consistent with what it believes, because T1 has neither ${\sim}C$ nor ${\sim}D$ as a theorem. However, only one of MC and MD is true in the real world. So for cautious robots, *Mp cannot* be taken as meaning, "p is true in some possible world consistent with what the robot believes."

The upshot is that the semantics I have described is more satisfying if robots using nonmonotonic logics are thought of as seeking a stable fixed point rather than a set of theorems. In fact, this is the way most practical programs operate, especially the TMS of [2]. Reiter [10] has also advocated concentrating on the property of "arguability," that is, presence in some fixed point [8], rather than provability. One reason for this emphasis is that it is cheaper to bet on one fixed point (debugging when necessary) than to try to reason about all of them. Another is that all fixed points are not alike; as I will discuss in Section 7, it is not worth worrying about "unlikely" fixed points. And another is supplied by my semantic argument, that M means what it claims to mean only if belief is construed as acceptance of a fixed point.

On the other hand, the concept of theorem, as Doyle and I have defined it, is still important and should not be discarded. That is, the intersection of all fixed points of a theory is just as important as the union. For one thing, this intersection has the property that it is closed under ordinary monotonic deduction, so that everything that can be deduced from it monotonically is a theorem. The union does not have this property and will in general be inconsistent.

An even more important reason is that arguability can be a very weak property of a formula. A theory may have many fixed points, and, to put it crudely, a formula is more interesting the more fixed points it appears in. (The "lottery paradox" example in Section 7 illustrates this.) If a formula appears in one fixed point out of an infinite number, it is not clear how interesting it is. I do not have a theory of measure of sets of fixed points, but at least one case is clear: a formula that appears in *every* fixed point is definitely interesting.

The next theorem will make this point quite clear.

THEOREM 11. *If there is an S5-modal model of proper axioms A, then there is an S5-noncommittal model of A.*

PROOF. First of all, notice that the definition of "S5-noncommittal" has a particularly nice equivalent: An S5-modal model V of A is an S5-noncommittal model of A iff there is no S5-modal model V' of A such that $\mathrm{acc}(V, A) \subset \mathrm{acc}(V', A)$. The "only if" case is obvious, even for T and S4. To prove the "if" case, assume there is no such model. Then if some formula p is true in any world of any S5-model V' of $A \cup \mathrm{acc}(V, A)$, it will be an S5-accidental of the largest connected piece of V' containing that world, since in a connected S5-model the truth of p in one world is sufficient for the truth of Mp in all worlds. (The largest connected piece is obtained by throwing away all worlds that cannot be reached from the given world through the $\mathrm{alt}_{V'}$ relation.) This piece has at least the same accidentals as the whole model, so just let V' name the piece. By construction, $\mathrm{acc}(V, A) \subseteq \mathrm{acc}(V', A)$, but not a proper subset, so p must be an accidental of V.

Another equivalent statement is: V is an S5-noncommittal model of A iff for all statements Lp such that $V(Lp) = 1$ and all models V' of $A \cup \mathrm{acc}(V, A)$, $V'(p) = 1$.

Now let U be an S5-modal model of A. Define R_0 to be $\mathrm{acc}(U, A)$. Let $q_1, q_2, \ldots$ be an enumeration of all the statements in Lang. Define R_{i+1} as follows:

(i) If some modal model V of $R_i \cup A$ has $V(q_{i+1}, w) = 1$ for some w in W_V, then let R_{i+1} be $R_i \cup \{Mq_{i+1}\}$.
(ii) Otherwise, every modal model of $R_i \cup A$ has $V(q_{i+1}) = 0$, so let $R_{i+1} = R_i$.

Now, by Theorem 8, if every finite subset of a set has a modal model, then the set itself has one. So there is a model U^* of $A \cup (\bigcup\{R_i : i = 0, 1, \ldots\})$. U^* is S5-noncommittal, by the following argument. Assume U^* is not S5-noncommittal. Then there is a q such that for some world w_0 in W_{U^*}, $U^*(Lq, w_0) = 1$, and for some S5-

modal model $U^{*\prime}$ of $A \cup \text{acc}(U^*, A)$ and some w_1 in $W_{U^{*\prime}}$, $U^{*\prime}(q, w_1) = 0$. Now $\sim q$ is q_{i+1} for some $i \geq 0$. That means that $U^{*\prime}$ is an S5-modal model of $A \cup R_i$ with

$$U^{*\prime}(q_{i+1}, w_1) = 1.$$

So

$$Mq_{i+1} \in R_{i+1}.$$

So

$$U^*(Mq_{i+1}) = 1 \qquad \text{and} \qquad U^*(Lq) = 0.$$

This contradicts our assumption that $U^*(Lq, w_0) = 1$. Q.E.D.

One nice consequence of Theorem 11 is that Pos does not overstrengthen the logic to the point of inconsistency, at least not for S5. Unfortunately, we pay a very high price for this property: A nonmonotonic theory based on S5 has no theorems that the corresponding monotonic theory does not have as well.

THEOREM 12. *If* $A \mid\!\sim_{S5} p$, *then* $A \vdash_{S5} p$.

PROOF. Assume $A \nvdash_{S5} p$. Then $A \cup \{M{\sim}p\}$ has an S5-modal model and hence an S5-noncommittal modal model. But then this model is an S5-noncommittal model of A in which p is not true. So by Theorem 10, $A \not\mid\!\sim_{S5} p$. Q.E.D.

For example, consider the apparently straightforward theory with one proper axiom $MC \supset C$, which is usually intended to mean, "C is to be considered true as a default." Surely C should be a theorem of this theory. But it is not. There is a fixed point of this theory with $M{\sim}C$. In S5, from $M{\sim}C$ and $MC \supset C$ you can infer ${\sim}C$. (Use a procedure like that in [6].) The inference of ${\sim}C$ blocks the assumption MC; so the fixed point works.

This is a serious bug of nonmonotonic S5. This is too bad, because it would be nice to have all the ordinary properties of S5 in doing deductions. S4 is harder to work with and has certain arbitrary properties (like fourteen different modalities) that do not appear to mean much to the logic of consistency. A defender of the arguability relation might find this a good reason to stick with S5, arguing that a robot could perfectly well believe C, as intended. But this overlooks the fact that in S5, ${\sim}C$ is just as arguable as C. Surely the logic should draw some distinction between a default and its negation if it is to be a "logic of defaults" at all.

Fortunately, S4 and T do not have this problem. However, they have other problems. For example, consider the theory with proper axiom $LMC \supset {\sim}C$. This theory is inconsistent. To see this, assume that it has a fixed point not equal to Lang. If MC is assumed in this fixed point, then LMC is in it; so ${\sim}C$ is too, and MC is not assumable in it. If MC is not assumed in it, then ${\sim}C$ is not in it, so MC is assumable in it. So there is no fixed point not equal to Lang, and the theory is inconsistent. But this theory is consistent in monotonic S4.

Strange theories like this one raise the fear that nonmonotonic S4 itself is inconsistent. I conjecture that nonmonotonic S4 and T are consistent but have so far been unable to prove that they are. However, if we restrict our attention to sentential subsets, consistency can be proven, as well as other results. So let me turn to that topic now.

6. *A Proof Procedure for Finite Sentential S4-Based Theories*

One of the weird features of nonmonotonic logic is that "provability" is defined without reference to "proof." Although a theorem will have a proof in any given fixed point of NM_A, there is no obvious way to generate, or even describe, each fixed point. Of course, this is not too surprising, since the structure of each fixed point depends on what is not provable from it, and the set of unprovable statements for a first-order theory is not recursively enumerable. As must be clear to everyone by now, using defaults in reasoning is not a simple matter of "common sense," but is computationally impossible to perform without error.

So we have to look for special cases for which things are not so bad. One promising place to start is with the sentential calculus. This turns out to be decidable. In this section I will prove this for the case of theories based on S4. It is no doubt true for T as well (and it is vacuously true for S5, since nonmonotonic S5 has no theorems that S5 does not).

I define a *finite sentential* theory as a theory that contains as proper axioms a finite list of variable-free formulas. The only theorems of such a theory that we will be interested in are the variable-free theorems.

As usual, I will bootstrap my way into the nonmonotonic case from the classical one. For the classical sentential modal calculus, one procedure for deciding the provability of p consists of a systematic search for a model in which p is false. That is, you create a model with a world in which the proper axioms are true and p is false. Then you explore consequences of the value assignments. When modal operators are involved, this may require creating a new possible world or copying value assignments from one world to others. When a disjunction is assigned a true value, you must split the investigation into *branches*, with a different disjunct made true in each branch.

For example, consider the formula

$$(MM \text{ ON(A, TABLE)}) \supset (LM \text{ ON(A, TABLE)}).$$

We begin by assigning it truth value 0 (false). So its antecedent is true and its consequent false. But if *MM* ON(A, TABLE) is true, there is a possible world in which *M* ON(A, TABLE) is true. Continuing in this vein, we arrive at a complete *tableau* like this:

```
W0:  ⟩(MM ON(A, TABLE)) ⊃ (LM ON(A, TABLE))
       1                 0  0

 W1:  M ON(A, TABLE)
      1

     W2:  ON(A, TABLE)
          1

 W3:  M ON(A, TABLE)
      0 0
```

In this diagram the angle bracket ⟩ points to the formula that started off this tableau. Notice how W1 was constructed from the 1-labeling of *MM* ON(A, TABLE) in W0; W2 was constructed from the 1-labeling of *M* ON(A, TABLE) in W1; W3 was constructed from the 0-labeling of *LM* ON(A, TABLE) in W0. A world constructed this way is a *alternative* of the world it is derived from, as are any worlds derived from it as the procedure progresses. (Each world is counted as an alternative of itself.)

At the point shown no further steps are possible, so the tableau is complete. What it demonstrates is that there is an S4-model in which the given formula is false; hence, by completeness, it is not a theorem. Notice that at the last step the labeling of *M* ON(A, TABLE) with 0 was propagated to all occurrences of ON(A, TABLE) that were in worlds alternative to W3, including W3 itself. This would have included any worlds constructed below W3 but does not include W1 and W2; if it had, the formula ON(A, TABLE) would have had a contradictory labeling, and the original formula would have been proved. (In S5, a stronger system, this last step is allowed, and the original formula *is* a theorem.)

For a more complex example, consider a theory with proper axioms:

Ax1: $LC \wedge MD \supset E$

Ax2: C

Ax3: $\sim D \supset MD$

E is a theorem, as the following tableau shows:

```
Branch 0:  W0:  )E     LC ∧ MD
                 0        0

Split Branch 0 on LC ∧ MD:
    Branch 1:  W0:  )E    LC
                     0     0

    Propagate on LC and copy Ax1:
                   W1:  C     C
                        1     0
                        └─────┘
Split Branch 0 the other way:
    Branch 2:  W0:  )E    MD
                     0    0 0

    Split Branch 2 on Ax3:
        Branch 2.1:  W0:  )E    MD     ~D
                           0    0 0    01
                                  └─────┘
    Split Branch 2 the other way:
        Branch 2.2:  W0:  E    MD     MD
                          0    0 0     1

        Propagate on both MD's:
                          W1:  D     D
                               0     1
                               └─────┘
```

In this example the proper axioms are labeled 1 in every world in every branch; I only copy the relevant parts at each step. Every branch eventually has an atomic formula labeled both 1 and 0 (a square bracket connects the two occurrences). A branch in this state is said to be *closed*, else it is *open*. A formula is unprovable if its tableau has an open branch.

To handle nonmonotonic logic, we must change the procedure to search for a noncommittal model. When a formula Lp is labeled 1 in any world, it is now necessary to create a new tableau in which p is labeled 0. If this new tableau is open, p is not provable, so we can label Lp 0, closing its branch. Similarly, if Mp is labeled 0, a new tableau is created with p labeled 1. If this tableau is open, Mp must be labeled 1 in all other tableaux.

The only problem with this procedure is that deciding whether a tableau is open or closed may depend on the states of other tableaux [8]. So you just have to try all combinations of the labels "OPEN" and "CLOSED" applied to the tableaux. A labeling is *admissible* if, after applying the rules in the previous paragraph, tableaux labeled CLOSED have all their branches closed, and tableaux labeled OPEN have at least one open branch apiece.

For example, with proper axioms,

$$MC \vee MD \supset E,$$
$$MC \supset \sim D,$$
$$MD \supset \sim C,$$

E is a theorem, as shown by the following tableau structure:

					Labelings	
Tableau 0:	B0:	W0:	)E 0	$MC \vee MD$ 0 0 0 0 0	CLOSED	CLOSED
Tableau 1:	B0:	W0:	)C 1	MD 0 0	OPEN	CLOSED
Tableau 2:	B0:	W0:	)D 1	MC 0 0	CLOSED	OPEN

Here tableau 1 exists because MC is labeled 0 in tableaux 0 and 2; it is created by labeling C 1. Tableau 2 exists because MD is labeled 0 in tableaux 0 and 1; it is created by labeling D 1. There are two admissible labelings. If tableau 1 is labeled OPEN, then tableau 2 is legitimately labeled CLOSED, because the OPEN label on tableau 1 entitles us to label MC 1 in all other tableaux. Similarly, tableau 2 may be labeled OPEN and tableau 1 CLOSED. Either way, tableau 0 is CLOSED, because one of MC or MD will be labeled 1.

Here is a summary of the proof procedure. To test provability of a variable-free formula p in a theory with proper axioms A, create a tableau with one branch, containing a world in which p is labeled 0 and every axiom in A is labeled 1. (This is the *tableau for p in A*.) Then repeatedly apply these rules:

1. (Truth-functional propagation). If the label on a formula implies labels on its subparts, then label the subparts accordingly If $q \supset r$ is labeled 0, label q 1 and r 0. If $\sim q$ is labeled 0 or 1, label q 1 or 0, respectively. Such labelings apply throughout the world the labelings take place in
2. (Possibility propagation) If Mq is labeled 1 in some world, create a new alternative world in which q is labeled 1 and every axiom in A is labeled 1
3. (Necessity propagation) If Mq is labeled 0 in some world, label p 0 in all alternative worlds
4. (Disjunction splitting) If $q \supset r$ is labeled 1 in some world, split the branch it occurs in into two branches, each a copy of all the worlds, formulas, and labels of the original branch, except that one has q labeled 0 in that world, and the other has r labeled 1 in that world
5. (Repetition elimination) If all the labelings in a world are duplicated in a world of which it is an alternative, delete it. (See [6, p. 111] for an explanation of this rule.)
6. (Consistency testing) If Mq is labeled 0 in some world, create a new tableau with one branch, containing a world in which $\sim q$ is labeled 0 and every axiom in A is labeled 1. This will be the *tableau for $\sim q$ in A* If there is already a tableau for $\sim q$, use that

These rules are repeatedly applied until they do not change anything. (Rules 2, 4, and 6 may be applied to a given formula only once.) Then every possible labeling of tableaux as OPEN or CLOSED is tried. If the tableau for $\sim q$ is labeled OPEN, then Mq must be labeled 1 in all tableaux. If there is an admissible labeling in which the original tableau is labeled OPEN, then the original formula is not a theorem; otherwise it is.

This is an adaptation of the familiar procedure for modal logic, described in [6]. There are two changes. To handle modal theories, I have specified that whenever a world is created, all proper axioms are labeled 1 there. This is the same as proving in S4 that $LA \supset p$. Clearly, every model of S4 that falsifies this is a modal model of A

that falsifies p, and vice versa, so the modified procedure constructs a modal theory that falsifies p if it is possible.

The other change is rule 6, which specifies the creation of new tableaux to test the consistency of subformulas. It is not so obvious that this adaptation is correct. It is not hard to see that the procedure still always halts: there are only a finite number of subformulas of the axioms and the formula being tested, so only a finite number of tableaux can be created, and the procedure halts for all of them.

It is also true that the procedure decides provability, as the next two theorems show.

THEOREM 13. *For each S4-noncommittal model V of finite sentential theory A such that $V(p) \neq 1$, there is an admissible labeling of the tableau structure for a variable-free formula p in A such that the tableau for p is OPEN.*

PROOF. By Theorem 9, V is an S4-modal model of some fixed point S of A. If a tableau in the structure is for q, label it CLOSED if $q \in S$, else OPEN.

Consider one of the tableaux labeled CLOSED, say the tableau for r in A. Since $r \in S$, there must be some minimal set of elements $X = \{Mr_1, \ldots, Mr_n\}$ such that $X \subseteq \mathrm{As}_A(S)$ and $A \cup X \vdash r$. If $X = \varnothing$, then the tableau for r is closed no matter how the other tableaux are labeled. Otherwise, adding X to A in each world will close every branch of the tableau. But adding Mr_j to a branch will create a new world with r_j labeled 1; if this closes the branch, there must be a proof of $\sim r_j$ from A plus a set Y of formulas $\{\sim q_i\}$ such that each q_i occurs in some Mq_i labeled 0 in a world to which this one is alternative. So there will be a tableau for each $\sim q_i$. But at least one of these tableaux will be labeled OPEN, since otherwise Y would be a subset of S and there would be a proof of $\sim r_j$ from $A \cup S$, and hence $\sim r_j$ would be in S and Mr_j would not be in $\mathrm{As}_A(S)$. Labeling one tableau OPEN makes the corresponding branch of the tableau for r closed. Similarly, every other branch of the tableau for r has an Mr_j that makes it closed, so the whole tableau for r will be closed.

Now consider a tableau labeled OPEN, say the tableau for r in A. One of its branches must be open, or else there must be a proof of r from some set of assumptions $\{Mq_i\}$, where the tableau for each $\sim q_i$ is labeled OPEN. But then r would be in S, because each $\sim q_i$ would not be in S.

The labeling is therefore admissible. By construction, p is in S if and only if the tableau for p is labeled CLOSED. Q.E.D.

THEOREM 14. *For each admissible labeling of the tableau structure for p in A such that the tableau for p is OPEN, there is a noncommittal model V of A such that $V(p) \neq 1$.*

PROOF. I will construct a fixed point not containing p, and hence a model in which p is not true, from an admissible labeling. Let R_0 be $\{Mq : \text{the tableau for } \sim q \text{ is labeled OPEN}\}$. Let $q_1, q_2, \ldots$ be an enumeration of all the variable-free statements of Lang, with the property that if q_j contains $\sim Mq_i$ as a subexpression, then $i < j$. Now define S_i to be S4-Th$(A \cup R_i)$ and R_{i+1} to be

$$R_{i+1} = \begin{cases} R_i & \text{if } S_i \vdash_{S4} \sim q_{i+1}, \\ R_i \cup \{Mq_{i+1}\} & \text{otherwise.} \end{cases}$$

The first thing to show is that for all i, $R_i \nvdash_{S4} p$. I will show this by induction. It is obviously true for R_0, since the tableau-construction procedure works for ordinary S4. Now assume that $R_i \nvdash_{S4} p$, and consider R_{i+1}. Clearly, if $S_i \vdash_{S4} \sim q_{i+1}$, the lemma is obvious, so assume $S_i \nvdash_{S4} \sim q_{i+1}$. To get a contradiction, assume that $R_i \cup \{Mq_{i+1}\} \vdash_{S4} p$. This means that augmenting A with the elements of R_i plus Mq_{i+1} would close

p's tableau. But the only effect of labeling Mq_{i+1} 1 is (by rule 2) to create several new worlds in various branches of the tableau (already augmented by adding the elements of R_i). Consider one such world, and call it w^*. If the branch b that w^* is on is closed as a result, then in the tableau for p there must be formulas $p_1, p_2, \ldots, p_m$ which are labeled 1 in w^* such that

$$\{p_1, \ldots, p_m\} \vdash_{S4} \sim q_{i+1}.$$

These p_k can come from three sources: some are axioms of A; some are derived from formulas $\sim M\sim p_k$ labeled 1 in the original tableau in some world to which w^* is alternative; and some are derived from formulas $\sim M\sim p_k$ which are subexpressions of elements of R_i.

Clearly, for each p_k from the first source, $R_i \vdash_{S4} p_k$. For each element of the second kind there will be a tableau for p_k (rule 6). This tableau must be labeled CLOSED, or the branch b would have been closed. So $R_0 \vdash_{S4} p_k$ for each such p_k and hence $R_i \vdash_{S4} p_k$. Finally, for elements in the third class, either $R_i \vdash_{S4} p_k$ or $M\sim p_k \in R_i$ by construction of R_i, given the subexpression property that I specified for the ordering $q_1, q_2, \ldots$ of formulas. But if $M\sim p_k \in R_i$, then the branch b would have been closed as soon as $\sim M\sim p_k$ got labeled 1; so $R_i \vdash_{S4} p_k$ in this case, too.

In all cases, $R_i \vdash_{S4} p_k$. So $R_i \vdash_{S4} \{p_1, \ldots, p_m\}$. But then $R_i \vdash_{S4} \sim q_{i+1}$, which is a contradiction. So $R_{i+1} \nvdash_{S4} p$.

Now define R to be $\bigcup\{R_i : i \geq 0\}$ − S4-Th(A) and S to be $\bigcup\{S_i : i \geq 0\}$. Clearly, S = S4-Th($A \cup R$). I will now show that S is a fixed point of S4-NM$_A$. This will follow if I show that S4-As$_A$(S) = R. First, let p be a statement such that $Mp \in$ S4-As$_A$(S). So $\sim p \notin$ S4-Th($A \cup R$) and $Mp \notin$ S4-Th(A). But p is q_{i+1} for some i, so $R_i \nvdash_{S4} \sim p$, and hence $Mp \in R_{i+1}$ and $Mp \in R$. So S4-As$_A$(S) $\subseteq R$.

Second, let p be a statement such that $Mp \in R$. If $A \cup R \vdash_{S4} \sim p$, then $A \cup R \vdash_{S4} \sim Mp$, and some R_i would be inconsistent. But then $R_i \vdash_{S4} p$, which I proved impossible. So $A \cup R \nvdash_{S4} \sim p$, and hence $Mp \in$ S4-As$_A$(S). So $R \subseteq$ S4-As$_A$(S).

So S is the desired fixed point, Furthermore, $p \notin S$, since otherwise some finite subset, and hence some R_i, would entail p. Q.E.D.

We also have the obvious

COROLLARY 15. *Nonmonotonic S4 (i.e., the theory with no proper axioms) is consistent.*

PROOF. Theorem 14 allows us to conclude that some formulas (e.g., Pa) are not theorems of nonmonotonic S4. Q.E.D.

This proof procedure works only for the sentential calculus, but it gives some insights that might be useful in constructing a heuristic prover for first-order theories. Notice, for instance, that labeling Mq or Lq 0 can be thought of as setting up a subgoal of proving q consistent or provable, respectively. What is curious is that in the proof of Lq you are allowed to use the other truth-value assignments in the current tableau; in the proof of Mq you are not.

7. *Applications*

In this section I will survey some practical and not so practical applications of nonmonotonic logic. These do not depend directly on the technical results I have derived so far, except that these results lend legitimacy to the discussion.

First let us look at some more "philosophical" uses of nonmonotonic logic. By using the word "philosophical" I do not mean to imply that they are irrelevant to more hardheaded artificial-intelligence research, but only that their relevance is likely to be indirect: they show the kind of thinking encouraged by nonmonotonic logics.

One application has occurred to several people (Jerry Hobbs, for instance, in a personal communication). This is to make set theory nonmonotonic. Naive set theory is afflicted with paradoxes, which are all due to an overly powerful "comprehension axiom," which states that for every property there is a set of just the objects with that property. The usual first-order approximation to this is the axiom schema,

$$(\exists s)(\forall x)(x \in s \leftrightarrow \mathrm{PROP}(x)),$$

which generates a new axiom for each formula PROP(x) with a free variable x. This allows us to conclude, for instance, that $(\exists s)(\forall x)(x \in s \leftrightarrow \text{color}(x, \text{red}))$, that is, that there is a set of all red things. Unfortunately, it also allows us to conclude that there is a set s such that $(\forall x)(x \in s \leftrightarrow x \notin x)$, from which it follows that $s \in s \leftrightarrow s \notin s$. (This is, of course, Russell's Paradox.)

The existence of this problem has led logicians to devise carefully stated restrictions of the comprehension axiom which apparently do not entail paradoxes. No one has been able to prove that the resulting systems are free of contradictions, and there is reason to believe no one ever will. But apparently they are.

Another approach to the problem would be to turn the theory nonmonotonic by switching to the following version of the comprehension axiom,

$$\mathrm{Pres}\ (\exists s)(\forall x)(x \in s \leftrightarrow \mathrm{PROP}(x)),$$

where Pres p (read "presumably p") is a convenient abbreviation for $Mp \supset p$. Now Russell's Paradox is avoided because $(\exists s)(\forall x)(x \in s \leftrightarrow x \notin x)$ is contradictory.

This approach is unlikely to be of interest to working logicians, because there is absolutely no way to check whether a given set of assumptions is contradictory or not. I have no idea how many fixed points nonmonotonic set theory would have, or even whether there are any.

On the other hand, it may be of interest to philosophers and engineers. As an engineering technique, this is probably just the right way to do set theory; there is no need for a rigorous proof that something is a set, because every property that occurs in the real world does correspond to a set. A robot would need to be able to withdraw its assumption that a property gave rise to a set only if he got involved in a cocktail-party conversation with a mathematician.

Philosophers may find nonmonotonic set theory interesting as casting some light on the question of whether sets really exist (see [3]). One problem with carefully chosen consistent axiomatizations of set theory is that there is more than one, and they give slightly different results. Which is correct? One is tempted to say that the question is meaningless, that set theory is just a formal game played by logicians, except that you can see that some sets do exist (the finite ones, for instance), and it is hard to see what could keep all the others from existing too. The nonmonotonic approach suggests an answer: The sets that exist in all noncommittal models of set theory definitely exist, and so do some others, but we can never know which they are, since we can never know which noncommittal model the real world is in. Whichever it is, all the well-behaved sets you need for mathematics do exist.

A more down-to-earth application is to the "sorites paradox": If you remove one grain of sand from a heap of sand, you still have a heap. But if you continue doing this, you will ultimately get to a single grain. Does that mean that a single grain is a heap? If not, is there some number, say 57,895 grains, below which a bunch of grains are not a heap?

Here is the problem stated in predicate calculus:

$$\sim\text{heap}(0)$$
$$\text{heap}(n) \supset \text{heap}(n + 1)$$
$$\text{heap}(n + 1) \supset \text{heap}(n)$$

Now, by a simple induction, we may infer $(\forall i)\sim\text{heap}(i)$.

This problem is old, but has recently caused problems for people trying to give truth conditions for natural-language sentences. Many words, like "near" and "recent," resist being given precise boundaries. This probably indicates that trying to specify truth conditions for natural language is silly, but let us look at the nonmonotonic solution anyway, which is to keep the first two axioms and replace the third with

$$\text{heap}(n + 1) \supset \text{Pres heap}(n)$$
$$\sim\text{heap}(n) \supset \text{Pres} \sim\text{heap}(n + 1)$$

Now, given Peano's axioms and noncontroversial things like

$$\text{heap}(100000),$$

it is true in all noncommittal models that

$$(\exists j)(j < 100000 \wedge (\forall n)((n \leq j \supset \sim\text{heap}(n)) \wedge (n > j \supset \text{heap}(n))))$$

but the value of j varies from model to model. Here is why it is true in every noncommittal model: 100000 is a heap, so every number bigger must be a heap. Zero is not a heap, so to preserve noncommittalness, $M\sim\text{heap}(1)$ "must" be true, and hence $\sim\text{heap}(1)$ as well. Similarly, heap(99999) "must" be true. Clearly there must be points j and $j + 1$ in the middle with all of the following true:

$\sim\text{heap}(j)$	$\text{heap}(j + 1)$
$L\sim\text{heap}(j)$	$L\ \text{heap}(j + 1)$
$M\ \text{heap}(j) \supset \text{heap}(j)$	
$M\sim\text{heap}(j + 1) \supset \sim\text{heap}(j + 1)$	

There can be only one such crossover, given the second axiom and its contrapositive: $\sim\text{heap}(n + 1) \supset \sim\text{heap}(n)$.

Another "paradox of monotonicity" is the lottery paradox. This paradox arose from investigations into rules of *conclusion* [13], a device in decision theory to allow you to take as true any formula whose probability is close enough to 1. The problem arises in situations such as lotteries involving many bettors. Say the threshold of acceptance is set at probability 0.995. Then in a lottery involving 1000 people, among them your friend Fred, the statement "Fred loses" has probability 0.999, so it will be accepted. But the same argument works for everyone else, so you end up with 1000 conclusions, together contradictory.

The solution appears obvious: use axioms like

$$\text{win(Fred)} \vee \text{win(Frank)} \vee \text{win(Frieda)} \vee \cdots$$
$$(\forall x)(\text{Pres} \sim\text{win}(x))$$

This theory has a thousand fixed points, in only one of which Fred is the winner.

In both of these examples it is important that a system using the logics try to find a stable piece of fixed point rather than accepting only proven things, in other words, that it be brave rather than cautious. For example, given a heap of 9087 things, the system can conclude that taking one away can make a heap, but this is not a theorem (because there is a noncommittal model in which j is 9086 and everything less than 9087 is not a heap).

There is an interesting kind of probability theory lurking in the background here. It seems important that 999 out of 1000 models of the lottery situation have Fred losing; it is a *safe bet* that he loses. Similarly, "most" models of the sorites situation have 9086 being a heap. In cases where a formula appears in only a few models, it seems much less reasonable to try to accept it. So far, nuances like this have not been taken into account by systems like Doyle's TMS (Tenability Maintenance System) [2] for managing nonmonotonic databases.

Another technical problem is that of "exceptions to exceptions." The usual examples of nonmonotonic formulas are formulas like, "A professor has a Ph.D. unless proven otherwise," which is overridden for a particular professor by proving that he or she does not have a Ph.D. But what about rules like this:

> All snakes are not poisonous,
> . . . except those in South America,
> . . . except Andean snakes living above 5000 feet,
> . . . except those kept in greenhouses by the Beitcha tribe,
> . . .

Here the third rule concludes the same thing as the first, but with a different status. If the second rule is satisfied, it overrules the first but not the third.

The solution to this problem is to introduce explicit indications whether a rule is applicable or not. Define the construction

(RULE name var p q $o1$ . . . oN)

to abbreviate

$$\begin{array}{ll} \text{isrule(name)} & \\ \wedge(\forall \text{var})(\,p \supset & \\ & M\ \text{applic(name, var)} \supset q \\ & \wedge \\ & \sim\text{applic}(o1, \text{var}) \\ & \wedge \\ & \sim\text{applic}(o2, \text{var}) \\ & \wedge \cdots) \end{array}$$

where the predicate "applic" is true of a rule and an object if the rule is applicable to the object.

Now we can express our facts about snakes thus:

(RULE R1 x snake(x) ~poisonous(x))
(RULE R2 x SouthAmerican(x) poisonous(x) R1)
(RULE R3 x over5000ft(x) ~poisonous(x) R2)
(RULE R4 x Beitcha(x) poisonous(x) R3)

Now, given a set of facts about a particular snake, these uniquely determine a fixed point in which the snake is either poisonous or not.

I should point out that the reducibility of this exception mechanism to ordinary nonmonotonic logic does not mean that in a real program this is the way to implement exceptions. It might be better to build in RULE, for instance, as a connective the program knows about the same way it knows about "if." For example, the program could look for conclusions of the form ~applic(. . .) and tell the TMS to handle them in a special way.

It is often thought that $(\forall x)$Pres $\sim P(x)$ means "I know all instances of P," since it can be used to deny P-ness to objects that cannot be shown to be P. It does not mean this, as the following argument (due to Robert Moore, in a personal communication) shows.

One would like the following kind of inference to work: Given that "I know what's in this box," "Some cheese is in this box," "Something in this box smells funny," and no knowledge of anything else's presence, infer that "The cheese smells funny." Here is how this looks in our notation:

$$(\forall x)\text{Pres} \sim \text{in}(x, \text{ThisBox})$$
$$\text{in}(\text{TheCheese}, \text{ThisBox})$$
$$(\exists x)(\text{in}(x, \text{ThisBox}) \wedge \text{Stinks}(x))$$

Unfortunately, Stinks(The Cheese) is not a theorem of this system. The problem is that there are plenty of noncommittal models in which the x that smells funny is not the cheese. In those models, Pres ~in(x, ThisBox) is true for this x because M~in(x, ThisBox) is false, or, put another way, I know about *two* things in the box, the cheese and the thing that smells.

One way to solve this problem would be to introduce into the logic a distinction between known and unknown individuals (see [9]). I think a simpler solution to this problem might lie along the following lines. (I believe this is probably a variant of McCarthy 's "circumscription" device [7].) What is missing is some way of inferring that {TheCheese} is the set of all things in the box. So what if we added the axiom schema,

$$\begin{array}{l} (\forall x)\text{Pres} \sim \text{PROP}(x) \\ \supset \\ (\text{RULE inferset } S \\ \quad M(S = \{x : \text{PROP}(x)\}) \\ \quad S = \{x : \text{PROP}(x)\}) \end{array}$$

This attempts to say, "If I know everything with PROP, then a set which could be the set of all PROPs is the set of all PROPs." Now there is a fixed point of the theory in which {TheCheese} = $\{x : \text{in}(x, \text{ThisBox})\}$ is proven. Unfortunately, there is also a fixed point in which {TheCheese, TheMouse} is the set. (We cannot have our axiom $(\forall x)$Pres ~in(x, ThisBox) rule out TheMouse $\in \{x : \text{in}(x, \text{ThisBox})\}$, since from the latter we can infer in(TheMouse, ThisBox).) The rule is allowing us to find more things in the box rather than fewer.

In previous examples of this section, I concluded that fixed-point ambiguity could be tolerated. In those cases I visualized using a rule like heap(n) $\supset$ heap($n - 1$) on demand to infer the heapness of particular pile sizes. A brave robot would just keep using this rule until trouble developed. In the cheese problem, the rule that generates the ambiguity does not so nicely suggest a course of action. If we use it by just trying out candidate sets of all objects in the box, we are not likely to hit on a good fixed point.

We can patch the rule by adding this:

$$\begin{array}{l} (\forall S)(M(S = \{x : \text{PROP}(x)\}) \\ \qquad \supset \\ \qquad (\forall S')(S \subset S' \\ \qquad\qquad \supset \\ \qquad\qquad \sim\text{applic}(\text{inferset}, S'))) \end{array}$$

Now all we need to be able to prove is TheCheese $\neq$ TheMouse, from which it will follow that {TheCheese} $\subset$ {TheCheese, TheMouse}, which will cancel the unwanted use of inferset.

I would not want to guarantee that this will handle every case, but it seems promising. In practice, a program using these axiom schemata could use them in a more specialized way. As soon as a formula of the form $(\forall x)$Pres ~PROP(x) was learned, the system would begin keeping track of objects with property PROP. They would be used in a theorem like this:

$$(\forall x)(\mathrm{PROP}(x) \supset x = a_1 \vee x = a_2 \vee \cdots \vee x = a_n),$$

where $a_1, a_2, \ldots, a_n$ are the objects collected. (Any deductions made from such a theorem would have to be revised, using data dependencies, as new a_i were deduced; see [2]. Compare the THFIND operator of PLANNER [4].)

8. *Conclusions*

The major result of this paper is Theorem 10. This establishes a satisfying notion of semantic interpretation and model for nonmonotonic theories.

Let me contrast this with some other treatments of similar issues. Many researchers in artificial intelligence have resisted the idea of exploring logics such as this one. One common line of argument, due most notably to Terry Winograd [15], is that it is a mistake to study the abstract notion of nonmonotonic provability, since any real program will be constrainted by lack of resources; the real issue is when it is reasonable to stop trying to prove something and start acting on the assumption that it cannot be proved.

It is true that this is an interesting question, and in the long run more interesting and difficult than those I have examined. However, it seems like a waste of our resources to implement programs for default reasoning without giving any thought to what conclusions they ought to come to. For example, a question that comes up in the design of real proof procedures is, "What is the significance of goal interactions?" If "prove p" is a subgoal of one proof effort, and "prove ~p" is a subgoal of another, under what circumstances can you stop working on your original goal? Questions like this must be answered before programs with limited resources are written.

Another criticism made of the particular approach Doyle and I have pursued is that it is at the wrong level: nonmonotonic inference rules are really rules about changing theories, not inference rules within a theory (e.g., [14], and Bundy, in a personal communication). For example, "If x is a bird, you can assume it can fly, unless you can prove otherwise," gets translated into "If ~can-fly(x) is not a theorem of T, you may adopt $T' = T \cup$ {can-fly(x)}." This rule is a theorem about legal and desirable actions taken in a space of theories, which is applied by some kind of executive which is trying to adopt the best theory at any given time.

The problem with this alternative is that it fails to capture one's intuitions about how defaults work, even in some very simple cases. For example, consider a theory with the structure,

$$MC \supset D, \qquad ME \supset {\sim}C.$$

If we translate this into the theory-transformation format, we have (apparently) two possible actions. assuming D and assuming ~C. However, if we do the second, the first becomes impossible; if we do the first and then the second, then we are left believing D "for no reason." (It is not clear what reasons would look like in this system anyway.)

Of course, there is another basic problem, which is that the condition parts of these action rules are not computable. So it is not actually decidable when an action should be taken.

One way around these difficulties is to introduce the notion of an "ideally allowable action," one that would be takable if all nontheorems were known and if all other ideally allowable actions were taken. But this sort of construction has little to do with a realistic theory of theory selection.

An approach along these lines has been worked out by Raymond Reiter [10]. His system is halfway between that of Weyhrauch and those of Doyle and me, in that nonmonotonic rules are thought of as rules for permitting theory extensions, but they are all considered together in a fixed-point construction somewhat like ours. Where we have implications of the form $(p \wedge Mq) \supset r$ and a single inference rule Pos, Reiter allows a theory to contain an arbitrary number of inference rules of the form, "If p is a theorem and q is not, r is." This is written in a notation like this:

$$\frac{p\ M \sim q}{r}.$$

The resulting system has many attractive properties, especially a somewhat more tractable proof theory than mine. There are a few drawbacks to his system. For example, from $(\exists x)Q(x)$ and

$$\frac{M\ P(x)}{P(x)}$$

his system allows the inference of $(\exists x)(P(x) \wedge Q(x))$. This does not seem correct to me, since there may well be millions of unnamed objects that lack property P, one of which is the one with property Q. Notice that if $(\exists x)Q(x)$ is replaced with $Q(a)$, where a is a new constant, we can infer $(\exists x)(P(x) \wedge Q(x))$, which shows that Skolemization [12] changes the meanings of formulas in our logic.

The main difference between the two approaches is whether M is to be included in the language or used as a mere marker in nonmonotonic inference rules, and this is just the question whether a decent semantics can be given to it if it is included. I do not think any conclusive answer to this question has yet been attained, but the preliminary answer supplied by this paper shows that the attempt is worth it. This preliminary answer is that Mp can mean what one intuitively thinks it should mean: that p is consistent with one's current beliefs. This applies only in systems, the "brave" ones, that search for a particular fixed point to try to accept.

ACKNOWLEDGMENTS. The key results of Section 5 were arrived at in collaboration with Jon Doyle. I thank him also for many stimulating conversations on all the topics of this paper. The referee made many useful criticisms. Any remaining mistakes are, of course, my fault.

REFERENCES

1 BOOLOS, G *The Unprovability of Consistency* Cambridge University Press, Cambridge, England, 1979

2 DOYLE, J. A truth maintenance system Memo 521, M I.T. Artificial Intelligence Lab , Cambridge, Mass , 1979

3 GOEDEL, K What is Cantor's continuum problem? *Amer Math Monthly 54* (1947), 515–525 Also in *Philosophy of Mathematics Selected Readings*, P Benacerraf and H. Putnam, Eds , Prentice-Hall, Englewood Cliffs, N J , pp 258–273

4 HEWITT, C Description and theoretical analysis (using schemata) of PLANNER A language for proving theorems and manipulating models in a robot Tech Rep 258, M.I T Artificial Intelligence Lab , Cambridge, Mass., 1972

5 HINTIKKA, J. *Knowledge and Belief.* Cornell University Press, Ithaca, N.Y., 1962
6. HUGHES, G E., AND CRESSWELL, M.J. *An Introduction to Modal Logic* Methuen and Co., London, 1972.
7 MCCARTHY, J Circumscription· A nonmonotonic inference rule *Artif Intell 13,* 1, 2 (1980), 27–40
8. MCDERMOTT, D V , AND DOYLE, J. Non-monotonic logic I. *Artif Intell. 13,* 1, 2 (1980), 41–72
9. MOORE, R. Reasoning about knowledge and action Ph.D. Dissertation, M.I T , Cambridge, Mass , 1979.
10. REITER, R. A logic for default reasoning *Artif Intell 13,* 1, 2 (1980), 81–132.
11 RESCHER, N , AND URQUHART, A *Temporal Logic* Springer-Verlag, New York, 1971.
12 ROBINSON, J.A A machine-oriented logic based on the resolution principle. *J ACM 12,* 1 (Jan 1965), 23–41
13. TUKEY, J. Conclusions vs. decisions. *Technometrics 2* (1960), 423–433.
14 WEYHRAUCH, R Prolegomena to a theory of formal reasoning *Artif. Intell. 13,* 1, 2 (1980), 133–170.
15 WINOGRAD, T Extended inference modes in reasoning by computer systems. *Artif Intell. 13,* 1, 2 (1980), 5–26.

RECEIVED SEPTEMBER 1979, REVISED NOVEMBER 1980, ACCEPTED DECEMBER 1980