Bilattices and Modal Operators

Matthew L. Ginsberg

Computer Science Department
Stanford University
Stanford, California 94305

Abstract

A *bilattice* is a set equipped with two partial orders and a negation operation that inverts one of them while leaving the other unchanged; it has been suggested that the truth values used by inference systems should be chosen from such a structure instead of the two-point set $\{t, f\}$. Given such a choice, we redefine a modal operator to be a function on the bilattice selected. We show that this definition generalizes both Kripke's possible worlds approach and Moore's autoepistemic logic, and also use recent developments in the formalization of logic programs to construct a procedure for determining the truth value of a sentence when the underlying knowledge base involves modal operators of this form.

1 Introduction

Modal operators are used in a variety of ways in AI, including reasoning about knowledge and belief, about time, and applications to nonmonotonic inference [11, 15, 13, and others]. The semantics assigned to a particular modal operator are usually determined using a scheme due to Kripke [12] that is based on the notion of possible worlds linked by an accessibility relation. Moore, however, needs to define his own semantics in [13] in order to establish the desired link between a modal operator of knowledge and existing ideas in nonmonotonic reasoning.

The purpose of this paper is to show that Moore's and Kripke's ideas can be unified into a single approach if we view modal operators not in terms of possible worlds, but as mappings on the truth values assigned to various sentences. Thus the modal operator L, where Lp means, "I know that p," simply assigns the truth value true to Lp if p is known to be true, and assigns Lp the value false if p is either known to be false or is not known to be true or false (i.e., if p is not known to be true).

The approach we are proposing is made possible by the fact that we will work with a formal system that explicitly allows us to label sentences with values other than the conventional ones of true and false. The description in the previous paragraph, for example, implicitly took advantage of a potential label for p that indicated that it was "unknown" in that it was not known to be either true or false.

In Section 2, we discuss the mathematical ideas underlying this approach, where truth values are taken not from the two-point set $\{t, f\}$ but instead from a larger set known as a *bilattice*. Section 3 extends these ideas as we have suggested, formally defining a modal operator to be a function on the elements of the bilattice of truth values.

Section 4 contains a variety of mathematical results. We show that our approach has analogs to the modal operators used by Kripke and by Moore, although the proof that we generalize their constructions is delayed until Section 6. We also present some results regarding modal operators generally, showing that Moore's L operator cannot be expressed in terms of conventional logical connectives but that it, in combination with Kripke-style modal operators, can be used to generate all possible modal operators on any bilattice corresponding to monotonic reasoning. In Section 5, we set the stage for proving that we have generalized Kripke's and Moore's work by extending our definition of inference to truth functions that involve modal operators. Finally, in Section 6 we return our attention to Moore's construction, showing that his notion of groundedness can be translated naturally into our setting. Section 6 also contains our fundamental unifying results, showing that first-order logic, Kripke's work and Moore's construction are all special cases of our general approach.

Section 7 discusses the construction of an automated theorem prover that handles modal operators of the sort we have defined. The principal observation made in this section is that it is possible to extend to our more general setting existing work on *stratification* [1, 7, 6, 5] that can be used to compute the consequences of some autoepistemic theories.

Finally, concluding remarks and suggestions for future work are the topic of Section 8.

One especially promising feature of the work that we will present is that it allows us to define modal operators that accept more than a single sentence as an argument. This may allow us to develop precise formalizations of notions such as causality; applications to temporal reasoning are also discussed.

In the interests of maintaining the continuity of the exposition, all proofs have been relegated to an appendix.

2 Mathematical preliminaries

In [9], a mathematical structure called a *bilattice* was introduced. Essentially, a bilattice is a set equipped with two partial orders and a negation operation that inverts one of them while leaving the other unchanged:

Definition 2.1 *A* bilattice *is a sextuple* $(B, \wedge, \vee, \cdot, +, \neg)$ *such that:*

1. $(B, \wedge, \vee)$ *and* $(B, \cdot, +)$ *are both complete lattices, and*
2. $\neg : B \to B$ *is a mapping with:*
   (a) $\neg^2 = 1$, *and*
   (b) $\neg$ *is a lattice homomorphism from* $(B, \wedge, \vee)$ *to* $(B, \vee, \wedge)$ *and from* $(B, \cdot, +)$ *to itself.*

A bilattice will be called distributive *if the bilattice operations* $\wedge$, $\vee$, $\cdot$ *and* $+$ *distribute with respect to one another.*

It is suggested in [9] that bilattices are natural objects to use in artificial intelligence applications, since the elements of the bilattice can be thought of as "truth values" labelling the statements in a declarative database. The two partial orders represent how much confidence we have in the validity of a particular sentence, and how much information we have about it. These ideas are expanded on considerably in [9] and have recently been explored by Fitting as well [3].

The two partial orders associated with a bilattice are denoted by $\le_t$ and $\le_k$. The partial order $\le_t$ reflects how certain we are that some sentence is valid, and corresponds to the lattice operations $\wedge$ and $\vee$. The partial order $\le_k$ is concerned with the amount of knowledge we have about a proposition, and is associated with the lattice operations $\cdot$ and $+$. The fact that negation inverts the t partial order is the representation of de Morgan's laws in our setting; negation leaves the k partial order unchanged because we know more about a sentence p than about a sentence q if and only if we know more about $\neg p$ than about $\neg q$.

Diagrammatically, we can draw a bilattice so that the partial order $\le_t$ increases from left to right on the page, and the partial order $\le_k$ from the bottom of the page to the top. Thus, for example, the simplest nontrivial bilattice is as depicted in Figure 1. This is the bilattice that corresponds most closely to conventional first-order reasoning. The four elements of this bilattice are used to label sentences that are known to be true (t) or false (f), about which

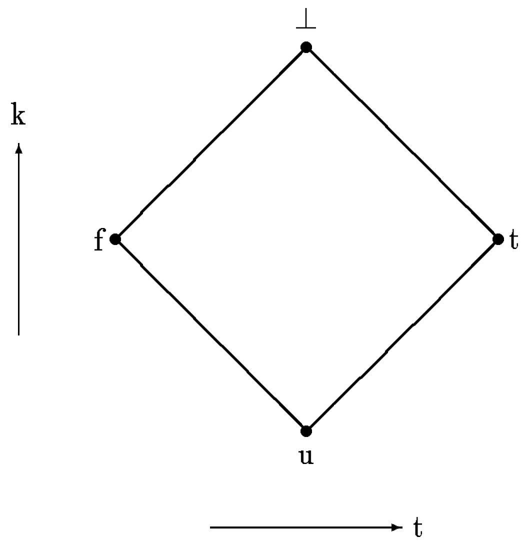

Figure 1: F, the smallest nontrivial bilattice

nothing is known (u), or that are known to be true *and* false ($\perp$). (This last truth value indicates the presence of a contradiction in our declarative database.) Larger bilattices also contain the four distinguished elements t, f, $\perp$ and u, these being the maximal and minimal elements of the two partial orders $\leq_t$ and $\leq_k$.

In [9], bilattices are also discussed that correspond to assumption-based truth maintenance systems (ATMS's) and default reasoning; in [10], this work is extended to a bilattice that can often be used to determine whether or not a given sentence follows from the circumscription axiom.

In the setting suggested by this approach, a declarative database consists of a mapping ϕ from the set W of the well-formed sentences in our logical language into a fixed bilattice B:

Definition 2.2 *A* truth assignment *is a mapping* $\phi : W \to B$.

In practice, most of the sentences in W will be mapped to u (unknown).

Of course, these ideas are of little use without an associated notion of inference. The one we will present is modelled on that used in first-order logic, where one first gives conditions under which a subset $U \subseteq W$ is deductively closed, and then defines the deductive closure of a set S as:

$$\mathrm{cl}(S) \equiv \bigcap \{T | T \supseteq S \text{ and } T \text{ is deductively closed}\}. \tag{1}$$

The multivalued analog of a closed set is a closed truth assignment:

Definition 2.3 *A truth assignment* ϕ *will be called* closed *if:*

1. $\phi(\wedge_i p_i) \geq_k \wedge_i \phi(p_i)$,
2. $\phi(\neg p) = \neg\phi(p)$ *for all* p, *and*

3. If $p \models q$, then $\phi(q) \geq_t \phi(p)$.

The motivation behind this definition is as follows: Condition (1) says that we should know at least as much about a conjunction as we could conclude by conjoining the truth values assigned to the various conjuncts. (2) says that the two views of negation – as a syntactic operator on our language and as a bilattice function – are equivalent, and (3) says that if p entails q, then q should be "at least as true" as p is.

The reason that condition (1) does not read

$$\phi(\wedge_i p_i) = \wedge_i \phi(p_i) \tag{2}$$

can be seen by considering the sentence $p \wedge \neg p$, where the truth value assigned to p is u (unknown). Now condition (2) says that the truth value assigned to $\neg p$ should be u as well (since if we know nothing about p, then we also know nothing about its negation), and the constraint given by (2) would incorrectly lead us to assign u to $p \wedge \neg p$ as well.

The truth value assigned to $p \wedge \neg p$ is instead determined by conditions (2) and (3). Specifically, if q is any sentence with $\phi(q) = t$, then since $q \models \neg(p \wedge \neg p)$, we know from condition (3) that

$$\phi(\neg(p \wedge \neg p)) \geq_t t,$$

so that $\phi(\neg(p \wedge \neg p)) = t$, and therefore by condition (2) that $\phi(p \wedge \neg p) = f$.

To construct an analog to (1), we need a bilattice version of the set-inclusion relation among subsets of W. Here it is:

Definition 2.4 *Let ϕ and ψ be two truth assignments. If $\phi(p) \geq_k \psi(p)$ for all $p \in W$, we will write $\phi \geq_k \psi$ and say that ϕ is an* extension *of ψ.*

The definition given by (1) now generalizes as follows:[1]

$$\mathrm{cl}(\phi) = \prod\{\psi | \psi \geq_k \phi \text{ and } \psi \text{ is closed}\}. \tag{3}$$

The bilattice product operation is the multivalued analog to set intersection; the condition that ψ be an extension of ϕ is the analog to the requirement that $T \supseteq S$ in (1).

To see that this is a reasonable definition, consider once again the four-point bilattice F appearing in Figure 1. If we fix a set of sentences $S \subseteq W$, we can define an associated truth assignment ϕ_S by:

$$\phi_S(p) = \begin{cases} t, & \text{if } p \in S; \\ u, & \text{otherwise.} \end{cases} \tag{4}$$

Essentially, this is the truth assignment that "knows" the sentences in S but nothing else. The following result is proved in [9]:

[1] This is not the definition adopted in [9]. The two definitions are equivalent for distributive bilattices, however.

Proposition 2.5 *Let $S \subseteq W$ be a set of sentences. Then for any sentence p,*

$$\mathrm{cl}(\phi_S)(p) = \begin{cases} \perp, & \text{if } S \text{ is inconsistent;} \\ t, & \text{if } S \text{ is consistent and } S \models p; \\ f, & \text{if } S \text{ is consistent and } S \models \neg p; \\ u, & \text{otherwise.} \end{cases}$$

In other words, once we compute $\mathrm{cl}(\phi_S)$, we can tell immediately whether or not a sentence p or its negation follows from the original set of assumptions S.

Lemma 2.6 *A closed truth assignment ϕ takes the value $\perp$ at some sentence p if and only if ϕ is the constant function $\perp$.*

Corollary 2.7 *A consistent set S of sentences is deductively closed if and only if there is some closed truth assignment ϕ such that $S = \phi^{-1}(t)$.*

3 Modal operators

Let us consider Definition 2.3 once again. What we would like to do at this point is to separate the third condition, which is the only one specifically involving entailment, from the first two.

As a preliminary, note that we can replace the second condition of the definition, that $\phi(\neg p) = \neg\phi(p)$, with the weaker constraint that

$$\phi(\neg p) \geq_k \neg\phi(p). \tag{5}$$

The reason for this is that the third condition gives us that $\phi(\neg\neg p) = \phi(p)$, so that (5), applied to $\neg p$, gives us

$$\phi(p) \geq_k \neg\phi(\neg p),$$

or

$$\neg\phi(p) \geq_k \phi(\neg p).$$

Combining this with (5), we conclude that $\phi(\neg p) = \neg\phi(p)$.

The point of this rewriting is that the first two conditions of Definition 2.3 are now of the same syntactic form. To make this remark more precise, we define a new truth assignment $\phi_\neg$ that assigns $\neg\phi(p)$ to $\neg p$, and assigns the truth value u to any sentence that is not of the form $\neg p$ for some p:

$$\phi_\neg(q) = \begin{cases} \neg[\phi(p)], & \text{if } q \text{ is of the syntactic form } \neg p; \\ u, & \text{otherwise.} \end{cases}$$

Thus, if ϕ assigns t to p, $\phi_\neg$ will assign f to $\neg p$ and u to p. If ϕ assigns u to q, then $\phi_\neg$ will assign u to both $\neg q$ and q. The truth assignment $\phi_\wedge$ is constructed similarly, taking

$$\phi_\wedge(p \wedge q) = \phi(p) \wedge \phi(q). \tag{6}$$

We now have:

Lemma 3.1 *A truth assignment ϕ is closed if and only if it satisfies:*

1. $\phi \geq_k \phi_\wedge$,
2. $\phi \geq_k \phi_\neg$, *and*
3. *If* $p \models q$, *then* $\phi(q) \geq_t \phi(p)$. $\square$

Of course, $\wedge$ and $\neg$ are not the only logical connectives; $\vee$ can also be dealt with in this fashion by defining

$$\phi_\vee(p \vee q) = \phi(p) \vee \phi(q).$$

In a similar way, $\supset$ corresponds to

$$\phi_\supset(p \supset q) = \neg\phi(p) \vee \phi(q).$$

In addition, quantifiers can be handled by introducing infinitary versions of $\wedge$ and $\vee$ (recall that the t-lattice of any bilattice is complete):

$$\begin{aligned} \phi_\forall(\forall x.p(x)) &= \wedge\phi(p(a)) \\ \phi_\exists(\exists x.p(x)) &= \vee\phi(p(a)) \end{aligned}$$

where the conjunction or disjunction is over all instantiations for x.

Lemma 3.2 *A truth assignment ϕ is closed if and only if it satisfies:*

1. $\phi \geq_k \phi_f$ *for* $f \in \{\wedge, \vee, \neg, \supset, \exists, \forall\}$ *and*
2. *If* $p \models q$, *then* $\phi(q) \geq_t \phi(p)$.

Definition 3.3 *A truth assignment ϕ will be called* prestable *if*

$$\phi \geq_k \phi_f \tag{7}$$

for any $f \in \{\wedge, \vee, \neg, \supset, \exists, \forall\}$.

A consequence of this definition is that ϕ is closed if and only if it is prestable and satisfies the final clause of Definition 2.3.

What we have done in (6) and other expressions like it is to realize that conjunction plays two distinct roles in the approach we have proposed. The first role is as a binary function from our bilattice of truth values to itself; the second is as a binary operator on the set W of well-formed sentences in our language.

We will take the view that this is in fact a special case of a much more general phenomenon – bilattice operations can be viewed *in general* as establishing semantic meanings for their syntactic counterparts. These syntactic counterparts are generally referred to as *modal operators*; we define a modal operator to be instead the associated function on the underlying bilattice:

Definition 3.4 *A* modal operator *is any n-ary function from the bilattice B to itself. Given a set M of modal operators, we define the* extended language W_M *to be the smallest set satisfying the following properties:*

1. *$W \subseteq W_M$.*
2. *For any modal operator $f \in M$ and elements $p_1, \ldots, p_n \in W_M$, $f(p_1, \ldots, p_n) \in W_M$.*

For simplicity, we will denote by $\overline{W}$ the extension of W that includes all modal operators of any arity, and if M is the singleton set $\{m\}$, will denote $W_{\{m\}}$ simply by W_m.

The definition of the extended language W_M matches the usual definition in which the set of well-formed formulae is extended in accordance with the introduction of one or more modal operators; when we say that $f(p_1, \ldots, p_n) \in W_M$, we mean only that W_M includes an element of this syntactic form, since the function f obviously cannot be applied directly to the various sentences p_i.

It is also possible to extend Definition 3.3 to our extended language. As a preliminary, we extend (6) to an arbitrary modal operator $f : B^n \to B$:

$$\phi_f(q) = \begin{cases} f(\phi(p_1), \ldots, \phi(p_n)), & \text{if } q \text{ is of the syntactic form } f(p_1, \ldots, p_n); \\ u, & \text{otherwise.} \end{cases}$$

This notation leads to a natural generalization of Definition 3.3; note that the condition (7) that $\phi \geq_k \phi_f$ can be rewritten in the equivalent form

$$\phi(f(p_1, \ldots, p_n)) \geq_k f(\phi(p_1), \ldots, \phi(p_n)). \tag{8}$$

4 Examples and characterization results

4.1 Kripke-style modal operators

The usual description of modal operators is originally due to Kripke [12], and is based on the notion of *possible worlds.*

Roughly speaking, Kripke considers a set of possible worlds in which all of the sentences in the unextended language are assigned truth values of t or f. These possible worlds are related via an *accessibility relation* a, and Kripke defines a modal operator L_a by saying that $L_a(p)$ holds in a particular world w if and only if p holds in all worlds accessible from w.

One way to formalize this (although not the conventional one) is to introduce a function ϕ that takes a sentence p and a world w and returns the truth value of p in w. The condition defining the modal operator L_a is now that

$$\phi(L_a p, w) = \bigwedge_{w'} \phi(p, w'), \tag{9}$$

where the conjunction is taken over the set of all w' with $a(w, w')$ (in other words, the set of all w' accessible from w). The semantics of the modal operator are then determined by the requirement that the truth values assigned by ϕ satisfy (9) in addition to the usual restrictions associated with the classical logical connectives.

Of course, (9) bears a striking resemblance to our earlier equation (8) that was also intended to describe a semantics for modal operators. To make this observation precise, suppose that we denote by S the set of possible worlds appearing in the Kripke construction. F^S, the set of functions from S to F, now inherits a bilattice structure from the set F, where the bilattice operations are computed pointwise and the assignment of the function g to a sentence p means that the truth value taken by p at the world w is given by $g(w)$.

We can now capture the sense of (9) by fixing an accessibility relation a and defining the modal operator (i.e., bilattice function) given by:

$$L_a(g)(w) = \bigwedge_{w'} g(w'), \tag{10}$$

where the conjunction, as in (9), is taken over all worlds w' with $a(w, w')$.

4.2 Autoepistemic reasoning

Of course, Kripke is not the only author to consider modal operators. Moore, for example, formalizes in [13] a modal operator L, where Lp is intended to capture the notion of, "I know that p." This is related to the following unary mapping on the bilattice F:

$$L(x) = \begin{cases} t, & \text{if } x = t \text{ or } x = \bot; \\ f, & \text{otherwise.} \end{cases}$$

We know p if its truth value is either t of $\bot$ (if the latter, we know p to be both true *and* false), and do not know p if its truth value is either f or u.

In Section 5, we will see that once we have extended Definition 2.3 to deal with general modal operators, the redescriptions that we have given of Kripke's and Moore's definitions do indeed generalize this earlier work. Before doing so, however, we develop some general results concerning the form of modal operators on distributive bilattices.

4.3 Characterization results

The principal result of this section is Theorem 4.8, where we show that every modal operator on a distributive bilattice can be expressed in terms of conventional logical connectives, Moore's L operator, and a set of operators that we will call *projections.*

As a preliminary, we have the following:

Proposition 4.1 *The L operator on the bilattice F cannot be written in terms of the existing bilattice functions $\wedge$, $\vee$, $\cdot$, $+$ and $\neg$ defined on F.*

In other words, the L operator is legitimately distinct from those that have already been defined on the bilattice F. This explains why the semantics of autoepistemic logic cannot be captured by the existing methods of first-order reasoning.

Proposition 4.2 *Every modal operator on the bilattice F can be written as a combination of the operators $\wedge$, $\neg$, $+$ and L, and the constant function u.*

As the upshot of the previous proposition was that the L operator cannot be written in terms of the existing logical connectives, the upshot of Proposition 4.2 is that no additional operators are needed, at least on the bilattice F.

For larger bilattices, there are additional possibilities. On the bilattice F^S, for example, there is a modal operator that assigns to the world j the truth values corresponding to the world i (corresponding to the accessibility relation a_{ij} where world i is accessible from world j and no other worlds are accessible at all). In terms of (10), we have

$$L_{a_{ij}}(g)(w) = \begin{cases} g(i), & \text{if } w = j; \\ t, & \text{otherwise.} \end{cases}$$

It will be more convenient if rewrite this as

$$L_{a_{ij}}(g) = \pi_{ij}(g) + c_j,$$

where c_j is given by

$$c_j(g)(w) = \begin{cases} u, & \text{if } w = j; \\ t, & \text{otherwise} \end{cases}$$

and

$$\pi_{ij}(g)(k) = \begin{cases} g(i), & \text{if } k = j; \\ u, & \text{otherwise.} \end{cases} \tag{11}$$

We also make the following definition:

Definition 4.3 *Let B be an arbitrary bilattice. We define a modal operator L on B by taking*

$$L(x) = \bigwedge\{y | y \geq_k (x \vee u) \text{ and } y \cdot \neg y = u\} \tag{12}$$

Lemma 4.4 *L and L^S coincide on F^S.*

Lemma 4.5 *Any modal operator on F^S can be written in terms of $\wedge$, $\neg$, $+$, L, constant functions and the various π_{ij} appearing in (11) for $i, j \in S$.*

In general, of course, the bilattice being used in a particular application may not be of the form F^S for any set S. To deal with these situations, we need to generalize the modal operators π_{ij} appearing in (11).

Definition 4.6 *Let B be a bilattice. A* projection *on B is a bilattice homomorphism that factors through F.*

In other words, π is a projection if an only if there exist bilattice homomorphisms $s : B \to F$ and $i : F \to B$ such that $\pi = is$.

Lemma 4.7 *The various π_{jk} appearing in (11) are projections.*

Theorem 4.8 *Any modal operator on any distributive bilattice can be written in terms of $\wedge$, $\neg$, $+$, L, constant functions and projections.*

It is also shown in [9] that a reasoning system is monotonic if and only if the associated bilattice is distributive. Theorem 4.8 therefore can be used to characterize all modal operators for monotonic inference systems.

5 Inference

In order to demonstrate that the description of modal operators that we have presented does in fact generalize earlier work, we need to extend Definition 2.3 to deal with inference in a wider setting.

Note first that Kripke and Moore view inference very differently. For example, autoepistemic reasoning is nonmonotonic: $\neg Lp$ is a consequence of the empty set $\emptyset$, but not of $\{p\}$. The reason for this is that the truth values assigned to sentences such as Lp or $\neg Lp$ are determined simply by evaluating the results of applying the modal operator L to the truth value of p.

In Kripke's case, things are not so simple. His approach requires that we consider the models of our base theory, and determine *from them* what truth values should be assigned to modal expressions. Thus if K is a modal operator corresponding to an accessibility relation that considers only one possible world W, and such that W is accessible from itself, we can prove

$$Kp \equiv p$$

so that it is possible to conclude p from Kp and $\neg p$ from $\neg Kp$. The analogous conclusions are not sanctioned in Moore's approach – from $\neg Lp$ we cannot conclude that p is actually false, even though p is false in the only model where $\neg Lp$ holds.

To formalize this distinction, we will split the set of modal operators on a particular bilattice B into *deductive* and *nondeductive* subsets. The intention is that we treat the deductive modal operators as Kripke does, determining their truth values by examining models, but treat the nondeductive modal operators simply as defining the truth values of their results.

Definition 5.1 *A modal operator on a bilattice B will be called* deductive *if and only if it commutes with $+$ and $\cdot$. All other modal operators will be called* nondeductive.

Proposition 5.2 *The L operator is nondeductive.*

Proposition 5.3 *Any modal operator on a distributive bilattice B that can be written in terms of $\wedge$, $\neg$, $+$, constant functions and projections is deductive. If B is either finite or isomorphic to F^S for some set S, then any deductive modal operator can be written in this fashion.*

We now need to extend the idea of a model to the bilattice setting. To do this, consider first the four-point bilattice F. It is fairly clear that a truth assignment corresponds to a model if and only if it assigns either t or f to every sentence in W. For larger bilattices such as F^2 corresponding to multiple copies of the four-point bilattice, we want to say that a "model" should label a sentence as being true or false in each *copy* of F. Thus for the bilattice F^2, we require

$$\phi(p) \in \{(t,t),(t,f),(f,t),(f,f)\}. \tag{13}$$

What characterizes the four bilattice points appearing in (13) is that they cannot be extended in the k direction without introducing some element of contradiction into the truth value that they represent. We formalize this as follows:

Definition 5.4 *An element x of a bilattice will be called* complete *if $x \cdot \neg x = u$ but for any element $y >_k x$, $y \cdot \neg y \neq u$.*

We are now in a position to give a definition of a model. Recalling our observation that the truth values assigned to sentences generated by nondeductive modal operators should *not* be determined by appealing to the truth values taken on models, we begin with the following:

Definition 5.5 *A truth assignment ϕ will be called* prestable *if $\phi \geq_k \phi_f$ for every deductive modal operator f.*

This leads to:

Definition 5.6 *A prestable truth assignment ϕ will be called a* model *if $\phi(p)$ is complete for every sentence p. ϕ will be called a* model of ψ *if ϕ is a model and $\phi \geq_k \psi$.*

If $\phi(q) \geq_t \phi(p)$ for every model ϕ, we will say that p entails *q and write $p \models q$.*

Definition 5.7 *A truth assignment ϕ will be called* stable *if and only if it satisfies the following conditions:*

1. *ϕ is prestable, so that*

$$\phi(f(p_1,\ldots,p_n)) \geq_k f(\phi(p_1),\ldots,\phi(p_n))$$

for any deductive modal operator f.

2. *If $p \models q$, then $\phi(q) \geq_t \phi(p)$.*

3. *For any nondeductive modal operator f,*

$$\phi(f(p_1,\ldots,p_n)) = f(\phi(p_1),\ldots,\phi(p_n)).$$

The first two conditions describe the semantics of deductive modal operators, generalizing the notions appearing in Definition 2.3. The final condition makes precise the observation we made at the beginning of this section that the truth values assigned to sentences of the form $f(p_1, \ldots, p_n)$ for nondeductive f be determined simply by evaluating the result of applying the modal operator f to the truth values of the p_i.

Proposition 5.8 *A truth assignment ϕ defined on W is closed if and only if it is the restriction to W of a stable truth assignment defined on $\overline{W}$.*

It follows from this result that we can replace (3) with the following:

Definition 5.9 *Let ϕ be a truth assignment defined on $\overline{W}$. Then the* stable closure *of ϕ will be defined to be*

$$\mathrm{scl}(\phi) = \prod\{\psi | \psi \textit{ is a stable extension of } \phi\}. \tag{14}$$

Combining Corollary 2.7 and Proposition 5.8, we get the following:

Theorem 5.10 *A consistent set of sentences $S \subseteq W$ is deductively closed if and only if there is some stable truth assignment ϕ such that $S = \phi^{-1}(t) \cap W$.* □

To see that (14) generalizes Kripke's work, let K be a collection of modal operators of the form defined by Kripke. Since the Kripke semantics of the operators in K are given by (9) or its bilattice analog (10), we immediately have:

Proposition 5.11 *A truth assignment ϕ defined on W_K is closed under the Kripke semantics if and only if it is the restriction to W_K of a stable truth assignment defined on $\overline{W}$.*

In order to prove an analog to Theorem 5.10, we need to extend Corollary 2.7 to the language W_K. In fact, we generalize Proposition 2.5:

Proposition 5.12 *Let $S \subseteq W_K$ be a set of sentences. Then for any sentence p,*

$$\mathrm{cl}(\phi_S)(p) = \begin{cases} \bot, & \textit{if } S \textit{ is inconsistent;} \\ t, & \textit{if } S \textit{ is consistent and } S \models p; \\ f, & \textit{if } S \textit{ is consistent and } S \models \neg p; \\ u, & \textit{otherwise.} \end{cases}$$

As for Theorem 5.10, we combine this with Proposition 5.11 to get:

Theorem 5.13 *A consistent set of sentences $S \subseteq W_K$ is deductively closed if and only if there is some stable truth assignment ϕ such that $S = \phi^{-1}(t) \cap W_K$.* □

Finally, we would like to draw a connection between this approach and the work on autoepistemic reasoning. To do this, we need the following definition, repeated from [13], where Moore credits Stalnaker [16] with the idea:

Definition 5.14 *A set of sentences $S \subseteq W_L$ will be called a* stable set *if and only if it satisfies the following conditions:*

1. *S is deductively closed,*
2. *If $p \in S$, then $Lp \in S$ as well, and*
3. *If $p \notin S$, then $\neg Lp \in S$.*

Proposition 5.15 *A consistent set of sentences $S \subseteq W_L$ is a stable set if and only if there is some stable truth assignment ϕ such that $S = \phi^{-1}(t) \cap W_L$.*

Note the close resemblance between this result and Theorems 5.10 and 5.13.

6 Groundedness conditions

Unfortunately, as discussed in [13], the closure of an autoepistemic theory is not given by the intersection of the stable sets containing it. As an example, consider Moore's example

$$\neg Lb \supset \neg b, \tag{15}$$

which might be interpreted as, "If I don't know that I have a brother, then I don't have a brother."

If we denote this sentence by p, then there are two minimal stable sets containing p. In the first (the intended one), we have $\neg Lb$, $\neg b$, $L\neg b$, and so on. Here, $\neg Lb$ holds, so that we don't believe that we have a brother, and $\neg b$ holds as a result.

The other minimal stable set containing p contains Lb, and therefore contains b (if it did not, it would contain $\neg Lb$ by virtue of Definition 5.14). In this counterintuitive situation, we know that we have a brother, and conclude from this that we have a brother in order to close our beliefs under the L operator.

In order to distinguish between the two stable sets in this example, Moore makes the following definition:

Definition 6.1 *A set of sentences T will be called an* autoepistemic expansion *of a base theory S if and only if T satisfies the following equation:*

$$T = \mathrm{cl}(S \cup LT \cup \neg L\overline{T}), \tag{16}$$

where $\overline{T}$ is the complement of T in W_L.

Clearly any autoepistemic expansion of a set S is stable.

The appearance of S in (16) allows us to conclude from p in (15) that we do not have a brother. Specifically, if T is to be an autoepistemic expansion of the set $\{p\}$, then T cannot contain Lb unless it contains b. But there is no way for b to be a consequence of sentences in S, LT and $\neg L\overline{T}$.

The bilattice analog to this definition is the following:

Definition 6.2 *Let ϕ be a truth assignment, ψ an extension of ϕ and S a collection of modal operators. We will say that ψ is* grounded for S *if*

$$\psi \leq_k \operatorname{scl}\left(\phi + \sum_{f \in S} \psi_f\right).$$

We will say that ψ is grounded *if it is grounded for the set of nondeductive modal operators.*

We now can finally replace the expression appearing in (3) with the following more general version:

Definition 6.3 *Let ϕ be a truth assignment. Then the* closure *of ϕ, to be denoted* $\operatorname{cl}(\phi)$, *is given by*

$$\operatorname{cl}(\phi) = \prod\{\psi | \psi \textit{ is a grounded stable extension of } \phi\}. \tag{17}$$

To see that the results of the previous sections still hold, we have:

Proposition 6.4 *Let ϕ be a truth assignment defined on $\overline{W}$ that takes the value u outside of W. Then* $\operatorname{cl}(\phi)$ *and* $\operatorname{scl}(\phi)$ *agree on W.*

This suffices to reproduce Theorem 5.10. For Theorem 5.13, we have:

Proposition 6.5 *Let ϕ be a truth assignment defined on $\overline{W}$ that takes the value u outside of W_K. Then* $\operatorname{cl}(\phi)$ *and* $\operatorname{scl}(\phi)$ *agree on W_K.*

With regard to autoepistemic logic, we also have the following:

Proposition 6.6 *Let $S \subseteq W_L$ be a set of sentences. Then for any sentence p,*

$$\operatorname{cl}(\phi_S)(p) = \begin{cases} \bot, & \textit{if } S \textit{ has no autoepistemic expansions and otherwise} - \\ t, & \textit{if } p \textit{ holds in every autoepistemic expansion of } S; \\ f, & \textit{if } \neg p \textit{ holds in every autoepistemic expansion of } S; \\ u, & \textit{otherwise.} \end{cases}$$

An immediate outgrowth of these results is that we can use our description to simultaneously provide a semantics for Kripke-style and autoepistemic modal operators.

7 Implementation

The formal description contained in (17) is, unfortunately, not suitable for use in an automated theorem prover. In the autoepistemic case, however, there is a fairly wide class of databases known as *stratified* [1, 7, 5] for which efficient computational procedures exist. This idea has an analog in our setting.

Definition 7.1 *Let ϕ be a truth assignment. We will say that a sentence $p \in \overline{W}$ is* active *in ϕ provided that either:*

1. *$\phi(p) \neq u$, or*

2. *There is an active sentence q with $q = f(p_1, \ldots, p_n)$ for a modal operator f where $p = p_i$ for some i.*

We will say that ϕ is stratified *if there is a mapping $s : \overline{W} \to Z^+$ that maps our language into the positive integers such that:*

1. *If $q = f(p_1, \ldots, p_n)$ is active in ϕ for a deductive modal operator f, then $s(q) = s(p_i)$ for each i, and*

2. *If $q = f(p_1, \ldots, p_n)$ is active in ϕ for a nondeductive modal operator f, then $s(q) > s(p_i)$ for each i.*

In this case, we will say that s is a stratification *of ϕ.*

The intent of the definition is to capture the sense in which we need to establish the truth value of one sentence before we can attempt to evaluate the truth value of another. Thus, for example, if ϕ assigns the truth value t to the sentence

$$La \supset b,$$

then the definition tells us that $s(La) = s(b)$ and that $s(a) < s(La)$. In other words, the truth values of b and of La need to be determined simultaneously, and they need to be determined after we have decided what truth value will ultimately be assigned to a.

The point of the definition is that if a truth assignment is stratified, we can evaluate the closure by working our way "up" the ordering given by s. The main result of this section indicates that this is indeed possible:

Theorem 7.2 *Let ϕ be a stratified truth assignment, stratified by the function s. This stratification allows us to define a sequence of truth assignments ϕ_i, where*

$$\phi_i(p) = \begin{cases} \phi(p), & \text{if } s(p) \leq i; \\ u, & \text{otherwise.} \end{cases}$$

Now $\phi_i \geq_k \phi_j$ if $i > j$, and $\phi = \sum_0^\infty \phi_i$. In addition, $\mathrm{cl}(\phi_j)(p) = \mathrm{cl}(\phi_{s(p)})(p)$ if $j > s(p)$, and

$$\mathrm{cl}(\phi) = \lim_{i \to \infty} \mathrm{cl}(\phi_i).$$

An examination of the proof of Theorem 7.2 shows that to compute the truth value in $\text{cl}(\phi_{i+1})$ of some sentence p with $s(p) = i + 1$, we should compute the truth values assigned to all of the relevant sentences in $\text{cl}(\phi_i)$, evaluate the truth values of nondeductive functional expressions by applying the appropriate modal operator to the sentences involved, and then compute the truth value of p using conventional logical techniques. This is very like the approach taken by PROLOG, where various auxiliary proof efforts are undertaken in order to compute the value to be assigned to a negated subgoal.

7.1 Any-time declarativism

An interesting property of the approach we are presenting is that it is possible to suspend these auxiliary proof efforts at any time, and to estimate the truth value to be assigned to some sentence in $\text{cl}(\phi_i)$. Having done so, we can continue to apply the modal operator to evaluate the truth value of p; this gives us a notion of "approximate" inference that can be used if computational resources are limited.

As an example, consider the following simple logic program:

```
bird(Tweety).
flies(Tweety) :- bird(Tweety),not(ab(Tweety)).
```

This translates [8] into the autoepistemic theory

$$\texttt{bird(Tweety)} \tag{18}$$

$$\texttt{bird(Tweety)} \wedge \neg L\texttt{ab(Tweety)} \supset \texttt{flies(Tweety)} \tag{19}$$

Suppose that we are interested in determining whether or not Tweety flies. Backward chaining, we discover that in order to determine the truth value assigned to this statement in $\text{cl}(\phi_2)$, we need to decide whether or not Tweety is abnormal, as indicated by $\text{cl}(\phi_1)$. The index always decreases, since the program is stratified.

What is happening here is that we are constructing $\text{cl}(\phi_2)$ *incrementally* from $\text{cl}(\phi_1)$. If our original logic program had more stratification levels, we would be computing $\text{cl}(\phi_n)$ from $\text{cl}(\phi_{n-1})$, which would in turn be computed from $\text{cl}(\phi_{n-2})$ and so on, until we reduced the original query to the problem of determining the truth values assigned to various sentences by $\text{cl}(\phi_1)$, which could be calculated from the truth assignment ϕ_1 itself.

If we wish, we can terminate this computation early by assuming that $\text{cl}(\phi_i) = \phi_i$. In our simple example, where we need to determine the truth value assigned by $\text{cl}(\phi_1)$ to $\texttt{ab(Tweety)}$, we can *approximate* this truth value by $\phi_1[\texttt{ab(Tweety)}] = u$ in order to estimate the impact of (19) on our belief that Tweety can fly. What this approach does is allow us to "jump" to the conclusion that Tweety flies without fully examining the proof tree involved.

An alternative, of course, is to avoid the approximation and to call the theorem prover recursively to evaluate

$$\text{cl}(\phi_1)[\texttt{ab(Tweety)}].$$

In this particular case, there is no information to the effect that Tweety is abnormal, and we draw the same conclusion as in the previous paragraph. However, had our database also included the facts that ostriches are abnormal and that Tweety is an ostrich, there would have been a difference between the two approaches, since the approximate version would have concluded $\neg L$ab(Tweety) without checking the consistency of this assumption.

These ideas can be used to give an "any-time" flavor to the declarative approach that we have described. At any given point in the inference process, the analysis will have proceeded to some stratification level i, and will be attempting to determine the truth values assigned by $\mathrm{cl}(\phi_i)$ to various sentences in our language. If we have exhausted the computational resources allocated to solving the declarative problem in question, we can approximate $\mathrm{cl}(\phi_i)$ by ϕ_i and return an approximate answer immediately.

7.2 Relationship to the standard definition

It is important to realize that the definition of stratification given by Definition 7.1 is somewhat weaker than the usual one. To see why, consider the following logic program:

```
a :- not(b).
a :- b.
```

This translates into the autoepistemic theory

$$\begin{aligned} \neg Lb &\supset a \\ b &\supset a \end{aligned}$$

Although the original program was stratified, the autoepistemic version is not, since we need to have

$$s(Lb) = s(a) = s(b).$$

The reason that the original program was stratified is because Van Gelder assigns to the rule `a :- b` only the condition that

$$s(a) \geq s(b),$$

instead of our stronger requirement that $s(a) = s(b)$. This is a consequence of the fact that the logic programming community views this rule *as a way to draw conclusions about a only.* If we somehow knew that a was false (which is actually impossible given the syntactic restrictions generally placed on logic programs), we would like to be able to conclude that b was false as well, but this conclusion would not be a consequence of the usual interpretation of the logic program.

This is in fact the reason that we view the above program as nonstratified. Upon learning that a is false, we will be forced to simultaneously assign truth values to both b and Lb, and the fact that program is not stratified reflects the difficulties to which this simultaneous assignment may lead.

It is interesting to note, however, that although the program we are considering is not stratified according to our definition, the algorithm corresponding to Theorem 7.2 will nevertheless give the correct answer if it is applied to the corresponding truth assignment. The reason for this is that the sentence $b \supset a$, which potentially effects the truth value of b, is equivalent to

$$\neg a \supset \neg b,$$

and can therefore only have the effect of changing the truth value assigned to b from u to f. Since the modal operator L maps both of these points into the same value (namely f), we are justified in assigning Lb the value f and concluding a from this.

There are other situations where this idea can be applied. Consider the following logic program, which is not stratified according to even the usual definition:

```
a :- not(b).
c :- not(d).
d :- not(c).
```

The truth values to be assigned to c and d cannot be determined, but we are nonetheless entitled to conclude a from $\neg Lb$.

What is happening here is that the interpretation of L as a bilattice function allows us to take advantage of the fact that the eventual response to the query a is insensitive to the question of whether b is assigned the truth value f or u and of whether c or d is true, provided that it is assigned one or the other of these values. This is quite similar to Van Gelder's description of the semantics of logic programs in terms of *alternating fixed points*, where he sanctions all conclusions that can be shown to hold independent of the eventual resolution of declarative loops such as that appearing in the above program. Van Gelder's approach remains distinct from ours, however, since (as mentioned at the beginning of this section) he lacks a true negation operator, and interprets a rule such as `a :- b` as fundamentally "about" a and not about b.

These restrictions on Van Gelder's approach appear to us to be difficulties with his ideas; having accepted them, however, he is able to show that *every* logic program has a well-defined alternating fixed-point semantics; we have no result indicating under precisely what circumstances the algorithm we have presented does indeed compute the closure of an arbitrary truth assignment.

8 Conclusion and future work

The purpose of this paper has been to argue that modal operators are best thought of not in terms of Kripke's possible-worlds construction, but as functions on the bilattice of truth values being used in any particular application. The technical content of the paper has been to show that this approach generalizes both the possible worlds work and Moore's autoepistemic logic, and to show that it is frequently possible to compute the results of a modal theory using slight generalizations of existing ideas from logic programming.

8.1 Extensions

8.1.1 Causality

The real value of our ideas, however, is not in their ability to combine existing notions under a single formal framework, but to extend them. As an immediate example, we have already noted that the work in Sections 4.1 and 4.2 will allow us to define modal operators that combine the features of Kripke's and of Moore's.

More interesting, however, is the fact that our construction does not inherit the possible-worlds construction's limitation to *unary* modal operators. If we write $p > q$ for "p causes q," this suggests that it may be possible to interpret $>$ as a binary modal operator on its arguments.

This idea is lent support by recent work of Gärdenfors [4], where it is suggested that causal and explanatory reasoning can be understood in terms of probabilistic manipulations on the truth values assigned to the sentences whose causal relationship is being investigated. In [14], it is argued that the power of Gärdenfors's approach lies not with the probabilistic reasoning it uses, but with the idea that the causal relationship between p and q can be determined by examining the sets of assumptions needed to guarantee the truth of these two sentences. These assumptions can be recorded in a bilattice-based truth value[2] and the Gärdenfors construction then reduces to the development of a modal operator of the sort we have discussed.

More specifically, suppose that we say that a causes b provided that a and b are both true, and that in the nearest world where a fails, b would fail as well.[3] In order to formalize this, we will suppose that we have identified some set C of ATMS contexts, and that for a given sentence p, we know the truth value of p in each of these contexts, so that our bilattice is in fact given by F^C. We will also assume the existence of a map n that accepts as arguments a set S of contexts and a particular context c and returns the context in S that is nearest to c.[4]

It is now reasonable to say that p causes q in a context c if $\phi(p)(c) = t = \phi(q)(c)$ (i.e., both p and q hold in c), and if q fails in $n(c, \phi(p)^{-1}(f))$, so that q fails in the context nearest to c among those in which p is false. In other words,

$$\phi(p > q)(c) = \phi(p)(c) \wedge \phi(q)(c) \wedge \neg\phi(q)[n(c, \phi(p)^{-1}(f))].$$

In modal terms, we have

$$x > y = x \wedge y \wedge \neg y[n(\cdot, x^{-1}(f))].$$

Note that the general approach we have developed immediately assigns a semantics to this operator – and perhaps an implementation as well.

[2]The close relationship of this bilattice to de Kleer's work on assumption-based truth maintenance [2] has led to this being called the ATMS bilattice in [9].

[3]My intention here is not to argue either in favor of or against this definition, but simply to show that it can be captured within the modal framework we have been discussing. Suffice it to say that definitions such as this are the topic of considerable discussion in the philosophical community.

[4]Once again, we sidestep philosophical issues such as whether or not this nearest context is unique. This is a paper about modal operators, not causality.

8.1.2 Temporal reasoning

Finally, we will sketch a possible application of our ideas to temporal reasoning problems. As with the discussion of causality in the previous section, this work should be viewed as preliminary.

One of the conventional approaches to temporal reasoning involves reifying the sentences in our language, so that in order to say that some sentence p holds at a time t, we actually write

$$\texttt{holds}(p,t). \tag{20}$$

The term *reification* refers to the fact that we have had to make the sentence p an object of our language in order to include it under the scope of the `holds` relation.

It seems more natural instead to treat `holds` as a modal operator, although this raises the problem that we need to deal with the temporal variable appearing in (20) in some way. We will do this by replacing the bilattice B with which we are working with B^T, where T is the set of time points in our temporal language. Thus we label a sentence not with an element of B, but with a *function* that gives its truth value as a function of time. Sentences with no temporal component are labelled by constant functions from T to B.

Having taken this view, how are we to express a causal rule such as

$$\texttt{holds}(\texttt{clear}(b),t) \wedge \texttt{occurs}(\texttt{move}(b,l),t) \supset \texttt{holds}(\texttt{loc}(b,l),t+1), \tag{21}$$

saying that we can relocate a block that is clear in the blocks world by moving it? The difficulty arises because the conclusion of the above rule is temporally delayed relative to the premises.

In order to describe this in our setting, we need to introduce a modal operator that corresponds to temporal delay. Here it is:

$$\Delta(f)(t) = f(t+1).$$

Δ is a modal operator that takes any truth value (i.e., function from T into B) and delays it by one time unit. We can now rewrite (21) in the compact form:

$$\texttt{clear}(b) \wedge \texttt{move}(b,l) \supset \Delta\texttt{loc}(b,l). \tag{22}$$

Note that Δ is deductive.

There are some difficulties with this approach; for example, it is somewhat awkward to describe situations in which the amount of delay varies depending upon features of the situation at time t. It remains to be seen whether these difficulties are offset by the advantages of the simplicity of (22) and the flexibility resulting from the fact that reification is not needed by this approach.

8.2 Further work

The analysis appearing in Sections 8.1.1 and 8.1.2 is clearly extremely preliminary, and much more work needs to be done on these and other applications of the general ideas that this paper has presented.

In addition, work is needed on the implementation issues discussed in Section 7. If a particular database is stratified, the results of that section can indeed be used to compute the closure. But what if the database is *not* stratified? As we have seen, the procedure implicit in Theorem 7.2 can be applied in somewhat wider circumstances than those in the premise of the theorem, and the precise nature of these circumstances needs to be investigated. The idea of any-time declarativism clearly warrants further investigation as well.

In fact, this paper has only begun to investigate the ideas suggested by the approach we have presented. Indeed, this has been our intention – to describe the approach itself, to show that it generalizes a variety of existing notions including Kripke's and Moore's constructions, to argue that it admits computationally efficient implementations, to suggest novel ways in which it can be used to describe causal and temporal reasoning, and then to leave the hard work for others.

A Proofs

Lemma 2.6. *A closed truth assignment ϕ takes the value $\perp$ at some sentence p if and only if ϕ is the constant function $\perp$.*

Proof. We need to show that if $\phi(p) = \perp$, then $\phi(q) = \perp$ for any sentence q. But if $\phi(p) = \perp$, then $\phi(\neg p) = \neg\perp = \perp$, so $\phi(p \wedge \neg p) = \perp$. But $p \wedge \neg p \models q$, so that $\phi(q) \in \{t, \perp\}$.

However, $p \wedge \neg p$ entails $\neg q$ as well, so that $\phi(\neg q) \in \{t, \perp\}$. Combining this with $\phi(\neg q) = \neg\phi(q)$, we get $\phi(q) = \perp$. □

Corollary 2.7. *A consistent set S of sentences is deductively closed if and only if there is some closed truth assignment ϕ such that $S = \phi^{-1}(t)$.*

Proof. If S is deductively closed, then $S = \mathrm{cl}(\phi_S)^{-1}(t)$. For the converse, suppose that $S = \phi^{-1}(t)$ for some closed ϕ. Now if $S \models q$, then there exist $p_1, \ldots, p_n \in S$ such that

$$p_1, \ldots, p_n \models q$$

and therefore

$$\phi(q) \geq_t \phi(p_1 \wedge \ldots \wedge p_n) \geq_k \phi(p_1) \wedge \ldots \wedge \phi(p_n) = t.$$

But since ϕ must avoid $\perp$ by Lemma 2.6, we have that $\phi(p_1 \wedge \ldots \wedge p_n) = t$ and therefore that $\phi(q) = t$, so that $q \in S$. □

Lemma 3.2. *A truth assignment ϕ is closed if and only if it satisfies:*

1. *$\phi \geq_k \phi_f$ for $f \in \{\wedge, \vee, \neg, \supset, \exists, \forall\}$ and*
2. *If $p \models q$, then $\phi(q) \geq_t \phi(p)$.*

Proof. We must only show that a closed truth assignment satisfies the first condition above. For $f = \wedge$ or $f = \neg$, this is an immediate consequence of Lemma 3.1; $f = \vee$ and $f = \supset$ are easy consequences of this. The interesting case is where f is a quantifier.

We know from Corollary 2.7 that if ϕ is a closed truth assignment, then $\phi^{-1}(t)$ is a closed subset of W. The only way for the lemma to fail is if $\phi(p(a)) = t$ for every instantiation of $p(x)$, but $\phi(\forall x.p(x)) = u$. But any model in which $\forall x.p(x)$ fails is one in which some instantiation $p(a)$ also fails. Thus $\forall x.p(x)$ is a consequence of $\phi^{-1}(t)$, $\phi(\forall x.p(x)) = t$ and the proof is complete. $\square$

Proposition 4.1. *The L operator on the bilattice F cannot be written in terms of the existing bilattice functions $\wedge$, $\vee$, $\cdot$, $+$ and $\neg$ defined on F.*

Proof. All of the functions defined on the first-order bilattice distribute with respect to the bilattice operation $\cdot$, so that if g is any modal operator constructed from them, we will have

$$g(u) = g(t \cdot f) = g(t) \cdot g(f).$$

However, $L(u) = f$ and $L(t) \cdot L(f) = t \cdot f = u$. $\square$

Proposition 4.2. *Every modal operator on the bilattice F can be written as a combination of the operators $\wedge$, $\neg$, $+$ and L, and the constant function u.*

Proof. First note that $\vee$ can be written in this form, since

$$x \vee y = \neg(\neg x \wedge \neg y).$$

Next, suppose that we define the functions δ_{xy} by

$$\delta_{xy}(z) = \begin{cases} x, & \text{if } z = y; \\ u, & \text{otherwise.} \end{cases}$$

Now any unary function f can be rewritten as

$$f(z) = \sum_x \delta_{f(x),x}(z),$$

and it follows that the sixty-four unary functions on F are generated by the various δ_{xy}.

The δ_{uy} are all the constant function u; δ_{fy} is the negation of δ_{ty} for each y and $\delta_{\perp y} = \delta_{ty} + \delta_{fy}$. It follows that we must simply write

$$\delta_{ty}(z) = \begin{cases} t, & \text{if } z = y; \\ u, & \text{otherwise} \end{cases} \tag{23}$$

as a combination of the given operators. But $\delta_{tt}(z) = u \vee [\neg(Lz \wedge L\neg z) \wedge z]$, and $\delta_{tf}(z) = \delta_{tt}(\neg z)$. Finally, $\delta_{tu}(z) = u \vee \neg L(Lz \vee \neg z)$ and $\delta_{t\perp}(z) = u \vee (Lz \wedge L\neg z)$.

Binary functions can be handled similarly, using the functions

$$\delta_{xyz}(v, w) = \begin{cases} x, & \text{if } v = y \text{ and } w = z; \\ u, & \text{otherwise.} \end{cases}$$

As before, we need construct only the δ_{tyz}. But $\delta_{tyz}(v, w) = \delta_{ty}(v) \wedge \delta_{tz}(w)$. Functions of higher arity can be handled similarly. $\square$

Lemma 4.4. *L and L^S coincide on F^S.*

Proof. Since (12) can be evaluated component-wise, it suffices to prove the lemma for the first-order bilattice F. The result of evaluating (12) for the various elements of F is:

$$
\begin{aligned}
L(u) &= u \wedge f \wedge t = f \\
L(t) &= t \\
L(f) &= u \wedge f \wedge t = f \\
L(\bot) &= t \qquad \square
\end{aligned}
$$

Lemma 4.5. *Any modal operator on F^S can be written in terms of $\wedge$, $\neg$, $+$, L, constant functions and the various π_{ij} appearing in (11) for $i, j \in S$.*

Proof. As in the proof of Proposition 4.2, we need to show only that we can construct the δ-functions

$$\delta_{tif}(g)(w) = \begin{cases} t, & \text{if } w = i \text{ and } g = f; \\ u, & \text{otherwise.} \end{cases} \tag{24}$$

Now suppose that we define the modified δ-functions given by

$$\delta_{tijx}(g)(w) = \begin{cases} t, & \text{if } w = i \text{ and } g(j) = x; \\ u, & \text{otherwise.} \end{cases} \tag{25}$$

Then the δ-functions appearing in (24) would be given by

$$\delta_{tif}(g)(w) = \bigwedge_{j \in S} \delta_{tij,g(j)}(g)(w).$$

It is therefore sufficient for us to construct the δ-functions appearing in (25). But examining (11), (23) and (25) immediately gives us

$$\delta_{tx}(\pi_{ij}(g)(w)) = \delta_{tijx}(g)(w),$$

from which we can conclude

$$\delta_{tijx} = \pi_{ii} \delta^S_{tx} \pi_{ij}, \tag{26}$$

where $\delta^S_{tx} : F^S \to F^S$ is the result of applying δ_{tx} to each of the components of an element of F^S.

Since the various modal operators appearing in the rewritten version of δ^S_{tx} will simply be $\wedge^S = \wedge$, $\neg^S = \neg$, $+^S = +$, constant functions or $L^S = L$, the proof is complete. $\square$

Lemma 4.7. *The various π_{jk} appearing in (11) are projections.*

Proof. We define $\pi_k : F^S \to F$ by

$$\pi_k(g) = g(k)$$

and $i_j : F \to F^S$ by

$$i_j(x)(w) = \begin{cases} x, & \text{if } w = j; \\ u, & \text{otherwise.} \end{cases}$$

Now it is clear from (11) that $\pi_{jk} = i_j \pi_k$. $\square$

Theorem 4.8. *Any modal operator on any distributive bilattice can be written in terms of* $\wedge$, $\neg$, $+$, L, *constant functions and projections.*

Proof. It is shown in [9] that any distributive bilattice B can be embedded in a bilattice of the form F^S for some S. Suppose, then, that we have embedded B in F^S and extended the given modal operator to all of F^S in an arbitrary way, obtaining the function f by doing so. It follows from the previous two lemmas that f can be written in terms of $\wedge$, $\neg$, $+$, L, constant functions and projections.

There are two things that might prevent this reexpression from applying to the bilattice $B \subseteq F^S$. The first is that the projections used might map B into elements of F^S outside of B itself; the second is that the operator L may be different when viewed as an operator on B and on F^S.

For the first of these, note that we only need in (26) that there be "enough" projections for us to construct the various δ-functions.

In general, suppose that we have a set of injections $\{i_k\}$ from F into B such that any element of B can be written as the sum of elements in the ranges of the various injections. Now for any i_k, let π_{kk} be a bilattice homomorphism from B to the image of F under i_k such that π_{kk} is the identity function on this image. Such bilattice homomorphisms always exist, since they exist on F^S. We now define mappings s_k from B into F by

$$s_k(x) = i_k^{-1}[\pi_{kk}(x)].$$

Now if we define projections π_{jk} by

$$\pi_{jk} = i_j s_k,$$

it is clear that the collection of π_{jk} can be used in (26) instead of the projections defined in (11). Finally, it is clear that we can always produce a set $S = \{i_k\}$ of injections, simply by adding new injections to S until all of B is covered.

It remains to show that the fact that the L operator may return different results on B and on F will not have any impact on the values returned by the rewritten expression. To see this, note that each term in (26) only deals with a subset of B that is isomorphic to F, and that for any embedding of F into B, the effective results of L on B and on F^S will match. □

Proposition 5.2. *The L operator is nondeductive.*

Proof. Unchanged from the proof of Proposition 4.1. □

Proposition A.1 *Let $(L, \wedge, \vee)$ be a distributive lattice. Then if L is either finite or isomorphic to 2^S for some set S, any lattice homomorphism on L can be written in terms of $\wedge$, $\vee$, constant functions and lattice projections (i.e., lattice homomorphisms that factor through the two-point lattice).*

Proof. I am indebted to Fred Linton for bringing this result to my attention (personal communication). □

Proposition 5.3. *Any modal operator on a distributive bilattice B that can be written in terms of $\wedge$, $\neg$, $+$, constant functions and projections is deductive. If B is either finite or* *isomorphic to F^S for some set S, then any deductive modal operator can be written in this fashion.*

Proof. The first part of the proposition is clear. For the second, suppose that we say that an element $x \in B$ is *t-grounded* if $x = x \vee u$, and that x is *f-grounded* if $x = x \wedge u$. It is shown in [9] that any bilattice element can be expressed uniquely as the sum of t-grounded and f-grounded elements, and that this expression is

$$x = (x \vee u) + (x \wedge u).$$

It is also shown that the sets of t-grounded or f-grounded points constitute a lattice with meet and join operations given by $\cdot$ and $+$ respectively. Since $\neg$ maps these two lattices into one another, and $\wedge = \cdot$ on the lattice of t-grounded points, this proposition follows from Proposition A.1. $\square$

Proposition 5.8. *A truth assignment ϕ defined on W is closed if and only if it is the restriction to W of a stable truth assignment defined on $\overline{W}$.*

Proof. Suppose first that ϕ is closed. We extend ϕ to $\overline{W}$ in stages. Provided that ϕ has been defined on each of $p_1, \ldots, p_n$, we take $\phi(f(p_1, \ldots, p_n))$ to be either $f(\phi(p_1), \ldots, \phi(p_n))$ if f is nondeductive or, if f is deductive, to be the k-minimal element of the bilattice that satisfies both the deductive closure condition and the requirement that

$$\phi(f(p_1, \ldots, p_n)) \geq_k f(\phi(p_1, \ldots, p_n)).$$

It is clear that the resulting truth assignment satisfies all of the conditions of Definition 5.7.

That the restriction of ϕ to W is closed if ϕ is stable is an immediate consequence of the fact that the definition of entailment appearing in Definition 5.6 matches the usual one on the unextended language W. $\square$

Proposition 5.11. *A truth assignment ϕ defined on W_K is closed under the Kripke semantics if and only if it is the restriction to W_K of a stable truth assignment defined on $\overline{W}$.*

Proof. The proof is essentially identical to the proof of Proposition 5.8. $\square$

Proposition 5.12. *Let $S \subseteq W_K$ be a set of sentences. Then for any sentence p,*

$$\mathrm{cl}(\phi_S)(p) = \begin{cases} \perp, & \text{if } S \text{ is inconsistent;} \\ t, & \text{if } S \text{ is consistent and } S \models p; \\ f, & \text{if } S \text{ is consistent and } S \models \neg p; \\ u, & \text{otherwise.} \end{cases}$$

Proof. If S is inconsistent, the result is immediate, so suppose otherwise. In this case, the models of S (including the semantic restrictions associated with the Kripke operators) are in natural correspondence with the models of ϕ_S in the sense of Definition 5.6, and the result follows. $\square$

Proposition 5.15. *A consistent set of sentences $S \subseteq W_L$ is a stable set if and only if there is some stable truth assignment ϕ such that $S = \phi^{-1}(t) \cap W_L$.*

Proof. Except for the presence of the L operator, the proof matches that of Theorem 5.10. That the L operator is handled satisfactorily is a consequence of the fact that the definition of stability requires that $\phi(Lp) = L\phi(p)$ for all p. $\square$

Proposition 6.4. *Let ϕ be a truth assignment defined on $\overline{W}$ that takes the value u outside of W. Then $\mathrm{cl}(\phi)$ and $\mathrm{scl}(\phi)$ agree on W.*

Proof. Since ϕ only deals with standard logical connectives, the condition that a stable extension ψ of ϕ be grounded is equivalent to the condition that $\psi \leq_k \mathrm{scl}(\phi)$. This will be true if and only if we take ψ to actually be the closure as given by (14), and it follows that the two definitions agree in this case. $\square$

Proposition 6.5. *Let ϕ be a truth assignment defined on $\overline{W}$ that takes the value u outside of W_K. Then $\mathrm{cl}(\phi)$ and $\mathrm{scl}(\phi)$ agree on W_K.*

Proof. The same as for Proposition 6.4; recall that the modal operators in K are all deductive. $\square$

Proposition 6.6. *Let $S \subseteq W_L$ be a set of sentences. Then for any sentence p,*

$$\mathrm{cl}(\phi_S)(p) = \begin{cases} \perp, & \textit{if } S \textit{ has no autoepistemic expansions and otherwise} - \\ t, & \textit{if } p \textit{ holds in every autoepistemic expansion of } S; \\ f, & \textit{if } \neg p \textit{ holds in every autoepistemic expansion of } S; \\ u, & \textit{otherwise.} \end{cases}$$

Proof. We prove this result by showing that the stable grounded extensions of ϕ_S are in natural correspondence with the autoepistemic expansions of S. Specifically, we will show that T is an autoepistemic expansion of S if and only if there is some stable grounded extension ψ of ϕ_S such that $T = \psi^{-1}(t)$.

First, we know from Lemma 2.6 that we need consider only the truth assignments that avoid $\perp$. In addition, we note that any stable truth assignment defined on the four-point bilattice that avoids $\perp$ can be characterized in terms of the sentences that it maps to t, since it maps the negations of these sentences to f and all other sentences to u.

The condition that an extension ψ of ϕ_S be grounded is now that

$$\psi \leq_k \mathrm{scl}(\phi_S + L\psi). \tag{27}$$

Since ψ is stable and an extension of ϕ_S, we have $\psi \geq_k \phi_S$ and $\psi \geq_k L\psi$, so that

$$\psi \geq_k \phi_S + L\psi$$

and therefore

$$\psi = \mathrm{scl}(\psi) \geq_k \mathrm{scl}(\phi_S + L\psi),$$

and (27) becomes

$$\psi = \mathrm{scl}(\phi_S + L\psi).$$

But this is equivalent to

$$\psi^{-1}(t) = \mathrm{cl}[\phi_S^{-1}(t) \cup (L\psi)^{-1}(t) \cup \neg(L\psi)^{-1}(f)].$$

If we write T for $\psi^{-1}(t)$, this becomes

$$T = \mathrm{cl}[S \cup LT \cup \neg(L\psi)^{-1}(f)]. \tag{28}$$

But $L^{-1}(f) = \{f, u\}$, so that

$$(L\psi)^{-1}(f) = L[\psi^{-1}(u)] \cup L[\psi^{-1}(f)] = L\overline{\psi^{-1}(t)} = L\overline{T},$$

and (28) becomes

$$T = \mathrm{cl}(S \cup LT \cup \neg L\overline{T}). \qquad \square$$

Theorem 7.2. *Let ϕ be a stratified truth assignment, stratified by the function s. This stratification allows us to define a sequence of truth assignments ϕ_i, where*

$$\phi_i(p) = \begin{cases} \phi(p), & \text{if } s(p) \leq i; \\ u, & \text{otherwise.} \end{cases}$$

Now $\phi_i \geq_k \phi_j$ if $i > j$, and $\phi = \sum_0^\infty \phi_i$. In addition, $\mathrm{cl}(\phi_j)(p) = \mathrm{cl}(\phi_{s(p)})(p)$ if $j > s(p)$, and

$$\mathrm{cl}(\phi) = \lim_{i \to \infty} \mathrm{cl}(\phi_i).$$

Proof. The theorem follows fairly easily from the assertion that $\mathrm{cl}(\phi_j)(p) = \mathrm{cl}(\phi_{s(p)})(p)$ if $j > s(p)$. If this were not the case, there must be some sentence q with $j \geq s(q) > s(p)$ such that q and p are logically related and, in fact such that p actually appears as an argument to a modal operator in q. If the top-level modal operator in q is deductive, then we would have $s(p) = s(q)$, so it must be the case that q is nondeductive.

For nondeductive q, it may well be the case that $\phi_{s(q)}$ has a stable extension ψ in which $\psi(p) \neq \phi_{s(p)}(p)$; however, it is clear that any such extension will not be grounded. The argument is unchanged from the discussion at the beginning of Section 6. $\square$

Acknowledgement

This work has been supported by General Dynamics and by NSF under grant number IRI89-12188. I would like to thank Ken Fertig, Vladimir Lifschitz, David Smith, and especially Fred Linton for many enlightening discussions.

References

[1] K. Apt, H. Blair, and A. Walker. Towards a theory of declarative knowledge. In *Proceedings of the 1986 Workshop on Foundations of Deductive Databases and Logic Programming*, 1986.

[2] J. de Kleer. An assumption-based truth maintenance system. *Artificial Intelligence*, 28:127–162, 1986.

[3] M. C. Fitting. Logic programming on a topological bilattice. *Fundamenta Informatica*, 11:209–218, 1988.

[4] P. Gardenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, 1988.

[5] A. V. Gelder. Negation as failure using tight derivations for general logic programs. In *Proceedings of the 1986 Workshop on Foundations of Deductive Databases and Logic Programming*, 1986.

[6] A. V. Gelder. The alternating fixpoint of logic programs with negation: Extended abstract. In *Proceedings of the ACM SIGACT-SIGMOD Symposium on Principles of Database Systems*, 1989.

[7] M. Gelfond. On stratified autoepistemic theories. Technical report, University of Texas at El Paso, 1986.

[8] M. Gelfond. Autoepistemic logic and formalization of common-sense reasoning. In *Proceedings 1988 Non-monotonic Reasoning Workshop*, Munich, 1988. American Association for Artificial Intelligence.

[9] M. L. Ginsberg. Multivalued logics: A uniform approach to reasoning in artificial intelligence. *Computational Intelligence*, 4:265–316, 1988.

[10] M. L. Ginsberg. A circumscriptive theorem prover. *Artificial Intelligence*, 39:209–230, 1989.

[11] J. Y. Halpern and Y. Moses. A guide to the modal logics of knowledge and belief: Preliminary draft. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 480–490, 1985.

[12] S. A. Kripke. Semantical considerations on modal logic. In L. Linsky, editor, *Reference and Modality*, pages 63–72. Oxford University Press, London, 1971.

[13] R. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25:75–94, 1985.

[14] E. Paek and M. L. Ginsberg. A justification-based theory of explanation. Unpublished manuscript, 1989.

[15] Y. Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence.* MIT Press, Cambridge, MA, 1988.

[16] R. Stalnaker. A note on non-monotonic modal logic. Unpublished manuscript.