BILATTICES AND MODAL OPERATORS

Matthew L. Ginsberg
Computer Science Department
Stanford University
Stanford, California 94306

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, and show that this definition generalizes both Kripke's possible worlds approach and Moore's autoepistemic logic. Extensions to causal and temporal reasoning are also discussed.

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 [6,10,8, and others]. The semantics assigned to a particular modal operator are usually determined using a scheme due to Kripke [7] that is based on the notion of possible worlds linked by an accessibility relation. Moore, however, needs to define his own semantics in [8] 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 argument that we generalize their constructions is delayed until Sections 5 and 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. Sections 5 and 6 also contain our fundamental unifying results, showing that first-order logic, Kripke's work and Moore's construction are all special cases of our general approach.

Concluding remarks and suggestions for future work are the topic of Section 7. 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.

Proofs of theorems will appear elsewhere.

2 Mathematical preliminaries

In [5], 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 [5] 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 [5] and have recently been explored by Fitting as well [2].

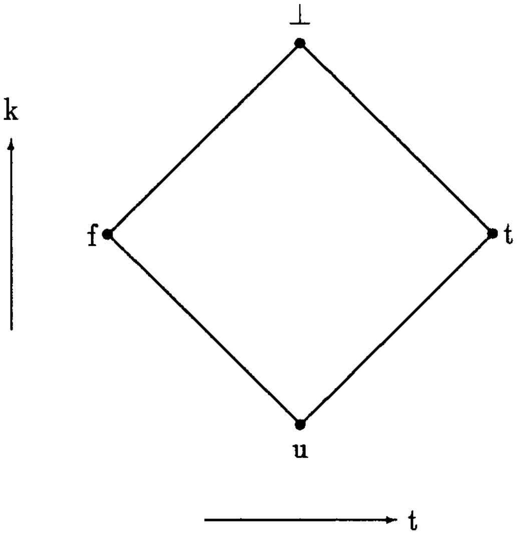

Figure 1: F, the smallest nontrivial bilattice

The two partial orders associated with a bilattice are denoted by $\leq_t$ and $\leq_k$. The partial order $\leq_t$ reflects how certain we are that some sentence is valid, and corresponds to the lattice operations $\wedge$ and $\vee$. The partial order $\leq_k$ 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 $\leq_t$ increases from left to right on the page, and the partial order $\leq_k$ 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 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 [5], bilattices are also discussed that correspond to assumption-based truth maintenance systems (ATMS's) and default reasoning; in [4], 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 $\phi : W \rightarrow B$.*

Given two truth assignments ϕ and ψ, we will say that ϕ is an extension *of ψ, writing $\phi \geq_k \psi$, if $\phi(p) \geq_k \psi(p)$ for every sentence $p \in W$.*

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 replaces the usual idea of a deductively closed set of sentences with that of a deductively 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

$$\phi(\wedge_i p_i) = \wedge_i \phi(p_i) \tag{1}$$

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 (1) 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 see that these notions generalize the conventional ones, we have:

Proposition 2.4 *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

$$\phi(\neg p) \geq_k \neg\phi(p). \tag{2}$$

The reason for this is that the third condition gives us that $\phi(\neg\neg p) = \phi(p)$, so that (2), 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 (2), 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. It is also not hard to see that Definition 2.3 is not altered if we include similar clauses involving other logical connectives. Disjunction is one of the functions describing the bilattice of truth values; implication can be described by taking

$$\supset(x, y) = \neg x \vee y.$$

Quantifiers can be handled similarly. We will view $\forall x.p(x)$ as shorthand for the set of all instantiations of $p(x)$, and take the associated bilattice function to be the infinitary version of $\wedge$ (recall that the t-lattice of any bilattice is complete). The quantifier $\exists$ is similarly related to $\vee$.

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

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

Definition 3.2 *A truth assignment ϕ will be called* prestable *if*

$$\phi(f(p_i)) \geq_k f(\phi(p_i)) \tag{3}$$

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 (2) and other expressions like it is to realize that negation plays two distinct roles in the approach we have proposed. The first role is as a unary function from our bilattice of truth values to itself; the second is as an 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 3.3 *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* $p_1, \ldots, p_n \in W_M$, $f(p_1, \ldots, p_n) \in W_M$.

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.

As we will see in Section 5, these notions lead to a natural generalization of Definition 3.2.

4 Examples and characterization results

4.1 Kripke-style modal operators

The usual description of modal operators is originally due to Kripke [7], 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 K_a by saying that $K_a(p)$ 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 semantics of the modal operator K_a is now that

$$\phi(K_a p, w) = \bigwedge_{w'} \phi(p, w'), \tag{4}$$

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 determined by the requirement that the truth values assigned by ϕ satisfy (4) in addition to the usual restrictions associated with the classical logical connectives.

Of course, (4) bears a striking resemblance to our earlier equation (3) 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 (4) by fixing an accessibility relation a and defining the modal operator (i.e., bilattice function) given by:

$$K_a(g)(w) = \bigwedge_{w'} g(w'), \tag{5}$$

where the conjunction, as in (4), is taken over all worlds w' accessible from w.

4.2 Autoepistemic reasoning

Kripke is not the only author to consider modal operators. Moore, for example, formalizes in [8] 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 $\wedge$, $\neg$, $+$, 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 (5), we have

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

It will be more convenient if rewrite this as

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

where c_j is given by

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

and

$$\pi_{ij}(g)(k) = \begin{cases} g(i), & \text{if } k = j; \\ u, & \text{otherwise.} \end{cases} \tag{6}$$

Roughly speaking, πij projects the world i onto the world j, and c_j indicates that no world at all is accessible from any world other than j.

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*

$$L(x) = \bigwedge\{y | y \geq_k (x \vee u) \text{ and } y \cdot \neg y = u\} \tag{7}$$

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 (6) 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 (6).

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 $s : B \rightarrow F$ and $i : F \rightarrow B$ such that $\pi = is$.

Lemma 4.7 *The various π_{jk} appearing in (6) 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 shown in [5] 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 every 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

$$\phi(p) \in \{(t,t),(t,f),(f,t),(f,f)\}. \tag{8}$$

What characterizes the four bilattice points appearing in (8) 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*

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

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.*

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.

To see that Definition 5.7 generalizes both first-order reasoning and Kripke's work, let K be a collection of modal operators of the form defined by Kripke. We now have:

Theorem 5.8 *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 $S = \phi^{-1}(t) \cap W_K$.*

To draw a connection between this approach and the work on autoepistemic reasoning we need the following definition, repeated from [8], where Moore credits Stalnaker [11] with the idea:

Definition 5.9 *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 5.10 *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 2.4 and 5.8.

6 Groundedness conditions

Unfortunately, as discussed in [8], 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

$$\neg Lb \supset \neg b, \tag{9}$$

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 5.9). 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:*

$$T = \mathrm{cl}(S \cup LT \cup \neg L\overline{T}), \tag{10}$$

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 (10) allows us to conclude from p in (9) 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 and ψ a stable extension of ϕ. We will say that ψ is* grounded *if*

$$\psi \leq_k \phi + \sum_f \psi_f, \tag{11}$$

where the sum is over all nondeductive modal operators f.

Proposition 6.3 *Let $S \subseteq W_L$ be a set of sentences, and set*

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

Then the grounded stable extensions of ϕ_S are in natural correspondence with the autoepistemic expansions of S.

In addition, Theorem 5.8 continues to hold because the sum in (11) is over nondeductive modal operators only, and the Kripke operators are excluded.

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 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.

7.1 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 [3], 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 [9], it is argued that the power of Gärdenfors's approach lies not in the probabilistic reasoning it uses, but in 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[1] and the Gärdenfors construction then reduces to 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.[2] 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.[3]

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

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

In the framework we have developed, this expression immediately assigns a semantics to causality operator $>$.

7.2 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

[1] The close relationship of this bilattice to de Kleer's work on assumption-based truth maintenance [1] has led to this being called the ATMS bilattice in [5].

[2] 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.

[3] 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.

write

$$\mathtt{holds}(p, t). \tag{12}$$

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 (12) 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

$$\mathtt{holds}(\mathtt{clear}(b), t) \wedge \mathtt{occurs}(\mathtt{move}(b, l), t) \supset \mathtt{holds}(\mathtt{loc}(b, l), t + 1), \tag{13}$$

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 (13) in the compact form:

$$\mathtt{clear}(b) \wedge \mathtt{move}(b, l) \supset \Delta\mathtt{loc}(b, l). \tag{14}$$

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 (14) and the flexibility resulting from the fact that reification is not needed by this approach.

7.3 Further work

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 suggest novel ways in which it can be used to describe causal and temporal reasoning, and then to leave the hard work for others.

Acknowledgement

This work has been supported by the Rockwell Palo Alto Laboratory, by General Dynamics, and by NSF under grant number DCR-8620059. I would like to thank Ken Fertig, Vladimir Lifschitz, David Smith, and especially Fred Linton for many enlightening discussions.

References

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

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

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

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

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

[6] 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.

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

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

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

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

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