88-32 1988

We use partial interpretations where propositions can have the value T, F, or U (for unknown). For a given set Γ of propositions, and the set of partial interpretations in which all members of Γ have the value T, we say that $\Gamma \models \alpha$ iff α has the value T in the minimal ones of those partial interpretations, according to an ordering which is part of the model. A new propositional operator D (for Default) is introduced whereby α may depend on the ordering, and whereby Γ may also constrain the ordering. The resulting definition of semantic entailment is shown to be non-monotonic with respect to the set Γ, and to be adequate for expressing default rules. Its properties are illustrated through a number of examples.

This research was supported by the Swedish Board of Technical Development, Rank Xerox AB, and the Rockwell Corporation. It was completed while the authors was visiting at the Computer Science Department at Stanford University.

Appeared in Methodologies for Intelligent Systems, vol. II, Proc. of ISMIS 1988 conference in Torino, Italy, Z.W. Ras and L. Saitta (eds), North Holland, 1988, pp. 391-397.

1. Approaches to non-monotonic entailment.

The semantic entailment relation $\models$ of ordinary, monotonic logic is monotonic in the sense that if

$$\Gamma \models \alpha$$

and

$$\Gamma \subseteq \Gamma'$$

it follows

$$\Gamma' \models \alpha$$

where of course Γ and Γ' are sets of logical formulae, and α is one such formula. We wish to find a modified definition of $\models$ which is not monotonic in this sense, and which has plausible properties, particularly with respect to the uses of non-monotonic reasoning which are of current great interest in A.I. The definition should allow us to express and perform defeasible reasoning, where default "rules" (expressed as logical formulae) characterize what conclusions one can draw from lack of information about the truth of certain other formulae.

Shoham [Sho87, Sho88] and others have proposed definitions of non-monotonic entailment based on a preference ordering of models, where the ordering is a parameter for the entailment relation itself, and therefore is not specified or constrained by the set Γ of axioms. We are here addressing the other case, where Γ specifies both the models and the preference ordering(s).

Other previous definitions of non-monotonic entailment use auxiliary mechanisms. Circumscription uses minimizations over possible models, and autoepistemic logic relies on "autoepistemic extensions" which are solutions to a particular equation. We believe that simpler and more direct definitions can be obtained by basing the semantics on *partial models.*

Thomason and Horty [TH88] define non-monotonic entailment using minimal partial models in a manner similar to our approach here. The explicit default operator D, which is crucial in our approach, however does not have a counterpart in their paper.

Turner [Tur84] and Ginsberg [Gin86] have also studied non-monotonic logic with unknown as a truth-value, but without defining entailment on that basis.

The present paper gives an introduction to our approach, with the basic definitions and a few examples. Another and more extensive paper [San88] contains more examples and additional detail.

2. Definitions.

Let **V** (vocabulary) be a set of *proposition symbols,* and let **L** be the set of *logic formulae* (= wff) formed from **V** using the ordinary propositional connectives plus the monadic operators L, M, N, and D. Thus if α is a logic formula, then $L\alpha$, $M\alpha$, $N\alpha$, and $D\alpha$ are also logic formulae.

An *interpretation* is a mapping from **V** to the three truth-values T, F, and U (for unknown). The partial order $\sqsubseteq$ is defined over interpretations so that

$$i \sqsubseteq i'$$

iff

$$i(p) = i'(p) \vee i(p) = \mathrm{U}$$

An *aggregate* is a twotuple $\langle J, \leq \rangle$ where J is a non-empty set of interpretations, and the relation $\leq$ over J is a partial order and a strengthening of $\sqsubseteq$ i.e.

$$i \sqsubseteq i' \rightarrow i \leq i'$$

Aggregates will be used similarly to models in modal logic, i.e. if an aggregate $S = \langle J, \leq \rangle$ is given together with a member i of J, and a logical formula α, the *value* of α in the "point" i in S will be defined. If the value is T we shall write

$$S \models_i \alpha$$

The value of α in i in S is defined as follows. If α is a proposition symbol, then $i(\alpha)$ is the value. The value of $\alpha \wedge \beta$, $\alpha \vee \beta$, and $\neg\alpha$ are defined from the values of α and β using Kleene's strong definitions [Kle52]. Thus $\wedge$ and $\vee$ are the *min* and *max* operations in an order where F $\prec$ U $\prec$ T. The values of $\neg\alpha$, $L\alpha$, $M\alpha$, and $N\alpha$ are obtained from the value of α according to the following table:

α	$\neg\alpha$	$L\alpha$	$M\alpha$	$N\alpha$
T	F	T	T	F
U	U	F	T	T
F	T	F	F	F

The value of $D\alpha$, finally, depends not only on i but also on the position of i in J according to the ordering $\leq$, and requires the following auxiliary definition.

Let $i \in J$ in $\langle J, \leq \rangle$. A *preferred path* from i is a sequence of members of J,

$$i_0, i_1, ...$$

where $i_0 = i$, and for all k, $i_k \sqsubseteq i_{k+1}$, $i_k \leq i_{k+1}$, and i_{k+1} is an *immediate* successor of i_k according to both $\sqsubseteq$ and $\leq$. Also the preferred path must continue as long as possible within J.

The value of a formula $D\alpha$ in i in S can now be defined: it shall be T iff α is T in the last interpretation of every preferred path from i, and F otherwise.

If $S = \langle J, \leq \rangle$ is given, the set $Min(S)$ of minimal members of J is defined as the set of all those members of J for which no other member of J is lesser according to $\leq$. The set $Min(S)$ is of course non-empty. Given also the definition of $S \models_i \alpha$, we now define

$$S \models \alpha$$

with respect to the whole of S, and not only one point in S, to hold iff $S \models_i \alpha$ for all members i of $Min(S)$.

An aggregate $S = \langle J, \leq \rangle$ is said to be a *model aggregate* for a set $\Gamma \subseteq \mathbf{L}$ if $S \models \beta$ for all members β of Γ, and in addition S is maximal in that J is as big as possible, and $\leq$ is no stronger than necessary. (More details about this can be found in [San88]). If S is a model aggregate, the members of J will be called *model states,* and we will write

$$\Gamma \rhd S$$

Using these definitions, entailment from formulae to formulae on the form $\Gamma \models \alpha$ (where $\alpha \in \mathbf{L}$) is finally defined as follows:

$$\Gamma \rhd S \rightarrow S \models \alpha$$

i.e. α must be true in all model aggregates for Γ. (If no S satisfies $\Gamma \rhd S$, then $\Gamma \models \alpha$ for all $\alpha \in \mathbf{L}$).

3. Some simple examples.

We illustrate these definitions with some examples of what formulas are entailed from simple sets of axioms.

Example 1. Let the formula a be the only member of Γ, and b be another proposition symbol. Then

$$\Gamma \models Nb$$

which can be understood as "if a is all we know, then we do not know whether b". Formally, we obtain only one model aggregate

$$\langle\{\mathrm{TT}, \mathrm{TF}, \mathrm{TU}\}, \sqsubseteq\rangle$$

where we write TU for the interpretation where a is T, b is U, and similarly for the others. TU is the only minimal model state, and the value of Nb in the interpretation TU is T.

If the formula b is added to Γ as a second axiom, then TT will be the only model state, and Nb is no longer entailed. This shows that the entailment relation is non-monotonic.

Example 2. Let the formula $a \vee b$ be the only member of Γ. This means that we do not know whether a is true, and we do not know whether b is true, but we do know that one of them is true. Then Γ does not entail Na or Nb, but it does entail Ma and Mb. These latter may be mnemonically understood as "maybe a" and "maybe b".

In the model aggregate, we obtain two minimal model states namely TU and UT, since UU is not a model state.

Example 3. The default rule "unless a is known to be false, assume that it is true" is written as

$$Ma \rightarrow Da$$

which can mnemonically be read as "if maybe a then default a". The model aggregate is

$$\langle\{T, F, U\}, \leq\rangle$$

where $\leq$ must be chosen so that $T \leq F$. In this way there will be only one minimal model, namely U, and one preferred path namely $\langle U, T\rangle$.

If Γ is the set consisting only of that rule, then Γ entails Da, which would be read as "by default, a". If $\neg a$ is added as a second axiom, then the model aggregate shrinks so that F is the only model state, Da is no longer entailed, and $\neg a$ is (of course) entailed.

The antecedent Ma is necessary, since the set $\{Da, \neg a\}$ is semantically inconsistent (does not have any model aggregate).

4. Default rules.

In the NME logic defined here, a tentative conclusion α from a set Γ of assumptions is written $\Gamma \models D\alpha$. This is different from earlier approaches, where Γ is said to entail α itself (or α is said to belong to an extension), and where there is no counterpart of the D operator.

The introduction of the D operator is in fact a necessary consequence of the approach we have taken for defining the semantics. With the new notation, we also have to reconsider how default rules are to be written. This is not the occasion for a systematic treatment of this topic, but we shall illustrate some of the key issues using classical examples.

Example 4: The royal elephant. Consider the statements "royal elephants are typically albinos" and "if an elephant is not an albino, then it is typically gray". We represent the proposition "the elephant at hand is a royal one" as the propositions symbol r, and similarly use a for albino and g for gray.

The two rules now have the general form

$$r \leadsto a$$

$$\neg a \leadsto g$$

where the exact interpretation of the $\leadsto$ symbol remains to be specified. In spite of the apparent similarity, there is a significant difference between these two rules, with respect to how the antecedents are defaulted. For the first rule, the default is that the antecedent does not hold, since most elephants are not royal ones, whereas for the second rule the antecedent holds by default. We may of course also encounter rules where the antecedent doesn't default either way.

We have worked through a number of different ways of expressing simple default rules as formulae in the logic defined here, and it appears that a correct formulation of the antecedent's default is important for obtaining the "right" results. In the present example, we would write the rules as follows:

$$M\neg r \rightarrow D\neg r$$

$$Lr \wedge Ma \rightarrow Da$$

$$\neg Lr \wedge M\neg a \rightarrow D\neg a$$

$$L\neg a \wedge Mg \rightarrow Dg$$

The condition $\neg Lr$ in the third axiom is the specificity constraint. Without it, the given axioms plus the axiom r would support both Da and $D\neg a$ in two distinct extensions.

Let us now consider what model aggregates we obtain for these axioms, using the entailment defined above. If we take only the first three axioms, and consider interpretations over the proposition symbols r and a, we obtain the following model aggregate

TT TF FT FF

TU UT UF FU

UU

In the figure, we draw $\sqsubseteq$ relationships as full lines, and $\leq$ relationships which are not at the same time $\sqsubseteq$ relationship as broken arrow pointing from the lesser to the greater model state, i.e. from the more preferred to the less preferred path. As the figure illustrates, the $\leq$ relation in the model aggregate must satisfy the following:

FT $\leq$ TT so that UT satisfies axiom 1
FF $\leq$ TF so that UF satisfies axiom 1
UF $\leq$ TU, one of several choices whereby UU satisfies axiom 1
TT $\leq$ TF, so that TU satisfies axiom 2
FF $\leq$ FT, so that FU satisfies axiom 3
UF $\leq$ UT, one of several choices whereby UU satisfies axiom 3

We see here that UU is the only minimal model state, so the set of the first three axioms entail Nr and Na, meaning we do not know whether the elephant at hand is royal, and whether it is an albino. However with the preference ordering just described, and the few others which are also possible, all preferred paths end in the model state FF, so consequently $D\neg r$ and $D\neg a$ are entailed: the elephant at hand is by default neither royal nor an albino.

If we add the axiom r saying that the elephant is royal, then the extended set of axioms entails Da but of course not $D\neg a$. If the axiom $\neg a$ is also added, whereby we state that the elephant at hand is very exceptional in not being an albino in spite of being royal, then the conclusion Da is lost, but still no contradiction is obtained.

It is also fairly easily verified that if the fourth of the original axioms is included, the model aggregate needs essentially only be extended in those points where a is assigned the value F. Again the set of axioms entails Dg except when overridden by a or Da.

Example 5: Even loop. The classical even loop in non-monotonic logic has the following form [San72]

$$\neg a \rightsquigarrow b$$

$$\neg b \rightsquigarrow a$$

In the NME logic defined above, these two rules (if formulated properly) have a model aggregate with two minimal model states, corresponding to the two extensions in the conventional way of looking at such an example. The model aggregate and thereby the two rules, entail the following formula

$$(a \wedge D\neg b) \vee (b \wedge D\neg a)$$

where each of the two disjuncts represents an "extension".

The details of these examples, as well as a number of additional examples have been presented in the more extensive paper, [San88]

5. Strange properties of NME logic.

The entailment operator defined here behaves in unorthodox manners. Let us however first notice two simple and natural properties which it *does* have:

a) if $\alpha \in \Gamma$ then $\Gamma \models \alpha$

b) modus ponens is sound: if $\Gamma \models \alpha$ and $\Gamma \models \alpha \to \beta$ then $\Gamma \models \beta$

These conventional properties follow directly from the definitions above. Now to the non-conventional properties:

c) if $\Gamma \models \alpha$ and $\Gamma \models \beta$ we can not in general conclude $\Gamma \cup \{\alpha\} \models \beta$

d) if $\Gamma \models \alpha$ and $\Gamma \cup \{\alpha\} \models \beta$, we can not in general conclude $\Gamma \models \beta$

The counterexamples are described in [San88]. These two observations mean that NME does not satisfy cumulative monotony and cumulative transitivity in the sense of Makinson [Mak88].

e) if α is a tautology in the sense of conventional propositional logic, we can not in general conclude $\Gamma \models \alpha$, nor even $\Gamma \models D\alpha$. Consider for example the counterexample where Γ has the proposition Na as its only member, and α is the proposition $a \vee \neg a$.

f) if $\Gamma \models \alpha$ and Γ is consistent (i.e. has a model aggregate), then we can not in general conclude that $\Gamma \cup \{\alpha\}$ is consistent.

6. Summary and discussion.

The proposed definition for non-monotonic entailment based on partial models, behaves as intended in many cases which have been analyzed. It forces us to exert special care with how certain default rules are expressed. In some cases its behavior is surprising, but can be argued to be appropriate, for example when $a \vee b$ entails Ma but not Na.

The counterexamples which establish the unorthodox properties described in the previous section are in some cases quite odd ones, and do not seem to arise naturally when expressing default rules or drawing the more obvious consequences from them. This suggests that the logic defined here maybe is a bit over-expressive, and that a non-monotonic logic with fewer unorthodox properties could be obtained by introducing constraints of some sorts, for example syntactic constraints on the formulae.

References

[Gin86] Matthew L. Ginsberg. Multi-valued logics. In *Proc. Fifth AAAI*, pages 243–247, 1986.

[Kle52] S. Kleene. *Introduction to Metamathematics.* Van Nostrand, 1952.

[Mak88] D. Makinson. General theory of cumulative inference. In Michael Reinfrank, editor, *Non-Monotonic Reasoning and Reason Maintenance*. Springer, 1988.

[San72] Erik Sandewall. An approach to the frame problem, and its implementation. In *Machine Intelligence, Vol. 7*, pages 195–204. Edinburgh University Press, 1972.

[San88] Erik Sandewall. The semantics of non-monotonic entailment defined using partial interpretations. In M. Reinfrank et al., editor, *Non-Monotonic Reasoning*, pages 27–41. Springer-Verlag, 1988. LiTH-IDA-R-88-31, LAIC-IDA-88-TR32.

[Sho87] Yoav Shoham. Non-monotonic logics: Meaning and utility. In *Proc. Tenth IJCAI*, pages 388–393, 1987.

[Sho88] Yoav Shoham. *Reasoning about Change*. MIT Press, 1988.

[TH88] R. Thomason and J. Horty. Logics for inheritance theory. In Michael Reinfrank, editor, *Non-Monotonic Reasoning and Reason Maintenance*. Springer, 1988.

[Tur84] Raymond Turner. *Logics for Artificial Intelligence*. Ellis Horwood, 1984.