On the Relation Between Truth Maintenance and Autoepistemic Logic

Michael Reinfrank
ZFE F 2 INF 22
Siemens AG
Otto-Hahn-Ring 6
8000 Munich 83, FRG

Oskar Dressier
ZFE F 2 INF 22
Siemens AG
Otto-Hahn-Ring 6
8000 Munich 83, FRG

Gerd Brewka
F3HIS
GMD
Postfach 1240
5205 St. Augustin, FRG

Abstract

In a recent paper, Konolige has introduced a new version of autoepistemic logic (AEL), which is based on a strong notion of groundedness. We show that it is well-suited for formalizing the concept of justified belief in a non-monotonic truth maintenance system (TMS). If we consider the justifications of a TMS as formulae of the form $La \wedge \neg Lb \supset c$, it computes the set of non-modal atoms of a strongly grounded AEL-extension. It is shown that a variant of Dressler's encoding of non-monotonic justifications in an assumption-based TMS is correct, and thus also inherits the AEL semantics We argue that more work is needed to come to a better understanding of backtracking routines and so-called nogood inferences, which are identified as sources of ungrounded conclusions. These results contribute to bridging the gap between theory and implementation in the field of non-monotonic reasoning

1 Introduction

1.1 The Problem

Despite their importance in AI problem solving, non-monotonic truth maintenance systems still lack sufficiently well-understood logical foundations Existing logical characterizations of truth maintenance suffer from one or several of the following problems:

- they do not correspond exactly to what a TMS actually does, as for example NML-I [McDermott, 1980].
- they only consider the "easy" monotonic case, as for example Reiter and de Kleer's prime implicant theory for assumption-based truth maintenance (ATMS) [Reiter and de Kleer, 1987]. Recent work on non-monotonic justifications for the ATMS, (de Kleer, 1986b] and [Dressier, 19881, underlines the need for non-monotonicity in truth maintenance,
- they are based on new, specialized non-monotonic logics whose sole purpose is to characterize a TMS, as for example Brown's logic ofjustified belief [Brown, 1988]. This is an interesting path to pursue, but it yields no insights into the relation between truth maintenance and the existing families of non-monotonic logics.

The problem addressed in the present paper is to establish a link between non-monotonic truth maintenance and autoepistemic logic, that is, a "standard" non monotonic logic.

L2 Analysis of the Problem

An analysis of the problem shows that a logical theory of non-monotonic truth maintenance is hard to design because of the following characteristics of a TMS:

- it is finite and logically incomplete
- it is "brave" in the sense that it may adopt one of multiple, mutually incompatible belief states,
- it has a very strong and, in particular, inherently global notion of a belief state being grounded
- nonmonotonic justifications are asymmetric: disbelief in the nonmonotonic antecedents can justify belief in the consequent, but disbelief in the consequent cannot justify belief in the non-monotonic antecedents.

1.3 The Approach

We pursue a two-step bottom-up approach. First, we specify non-monotonic truth maintenance in a direct, non-logical theory. Then, the specification, together with the corresponding formal framework, is used to draw a connection to autoepistemic logic The theory is a derivative of Jon Doyle's reasoned assumptions LDoyle, 19831, and is discussed in [Reinfrank and Freitag, 1988]. It is built on the concept of a non-circular proof being valid relative to a current coherent belief state. To this end, a justification is considered as a rule of the form $<a|b \rightarrow c>$, where a is a monotonic antecedent and b a nonmonotonic antecedent for the conclusion c (In general, we consider sets of antecedents). A justification is valid in a given set S if a€S and bεS A set of such justifications is called a non-monotonic formal system (NMFS). An extension of an NMFS is a set S such that everv element of S has a non-circular proof using only valid justifications, and where the conclusion of every valid justification belongs to S.

One can easily verify that extensions correspond exactly to the IN/OUT labellings of a non-monotonic justification-based TMS [Doyle, 1979, Goodwin, 19871. With some minor extensions, NMFS-theorv is also sufficient to describe an assumption-based TMS (deKleer, 1986a).

1.4 The Solution

The key to a logical theory of truth maintenance now is to relate a justification $<a|b \rightarrow c>$ to a self-referential formula in autoepistemic logic (AEL), $La \wedge \neg Lb \supset c$, where L is an introspective modal operator, and La reads as "a is believed". This formula is essentially in AEL normal form [Konolige, 1988a), and therefore the AEL-transform of a set of justifications can be used as a basis for so-called strongly grounded AEL-extensions. Notice that the scope of modal operators and the right-hand side of the implication are restricted to non-modal atoms. Given an NMFS J, the strongly grounded AEL-extensions of its transform JAEL correspond to the extensions of J, and hence to TMS-labellings, in the sense that a TMS labels exactly the non-modal atoms of a strongly grounded AEL-extension IN.

We show that a variant of Dressler's encoding of non-monotonic justifications in an ATMS [Dressier, 19881 is correct. Therefore it also inherits the AEL semantics. Dependency-directed backtracking and so-called nogood inferences are identified as a source of ungrounded conclusions.

2 A Direct Theory of Truth Maintenance

We consider a countable set V of propositional atoms. A TMS works with finite subsets of V. A *justification* is a rule p $<A|B \rightarrow c>$, where A and B are finite sets of atoms, and c is a single atom. (We usually omit the set parentheses when enumerating the members of A and B.) p is *valid* in a set $S \subseteq V$ (or *S-valid)* iff $A \subseteq S$ and $B \cap S = \emptyset$. Note that a justification of the form $<\emptyset|\emptyset \rightarrow c>$ is valid in every set S. It is called a *premise justification,* and c a *premise.* A *non-monotonic formal system* (NMFS) is a finite set of justifications.

Def. 2.1: Let J be an NMFS, $S \subseteq V$, and $q \in V$. A *J-proof for q valid in S* is a sequence (q1,q2,- q*n)* vvith the following properties:

(1) $q_n = q$
(2) $\forall q_i: q_i \in S$
(3) $\forall q_i: \exists <A|B \rightarrow q_i> \in J:\quad A \subseteq \{q_1, \ldots q_{i-1}\} \wedge B \cap S = \emptyset$

In other words, a J-proof is a non-circular sequence of applications of S-valid justifications to intermediate conclusions.

Def. 2.2: Let J be an NMFS, $S \subseteq V$. S is *J-closed* iff $\forall <A|B \rightarrow c> \in J: A \subseteq S \wedge B \cap S = \emptyset \Rightarrow c \in S$.

Def. 2.3: Let J be an NMFS, $S \subseteq V$. S is *J-grounded* iff every $q \in S$ has a J proof valid in S.

Sometimes we need to consider sets that are only grounded in a substantially weaker sense.

Def 2.4 Let J be an NMFS, $S \subseteq V$ S is *locally J-grounded* iffevery $q \in S$ has an S-valid justification.

Def. 2.5: Let J be an NMFS, $S \subseteq V$. S is a *J-extension* iff

(1) S is J-grounded and
(2) S is J-closed

An extension thus has the property that it contains an atom if and only if a proof - valid in that extension - can be found for the atom. As we have shown in IReinfrank, 19871, a justification-based TMS computes exactly the extensions of its justifications. We can easily verify that NMFS have the following properties:

finiteness : given that J is finite, so are its extensions.

logical incompleteness: consider a propositional language *L* rather than a set V of constants as the domain from which justifications are formulated. $\{<p|p \vee q \rightarrow r>, <\emptyset|\emptyset \rightarrow p>\}$ has the extension {p,r}, since a TMS fails to make the logically valid inference pVq from p.

braveness: $\{<\emptyset|p\rightarrow q>, <\emptyset|q\rightarrow p>\}$ has two extensions: {p}, {q}.

strong, global notion of groundedness: $\{<p|\emptyset\rightarrow p>, <\emptyset|p\rightarrow q>\}$ has exactly one extension {q} Notice that {p} is closed, locally grounded, that is, p is the consequence of a {p}-valid justification, and it is minimal in that respect. But it is not grounded.

asymmetry of justifications: the only admissible extension of $\{<\emptyset|q\rightarrow p>\}$ is {p}. It does not yield an extension {q}, where p is disbelieved and q believed, "backward" justified by disbelief in p.

The usual approach to achieving a higher degree of logical completeness, unless full responsibility for logical inferences is left to the problem solver, is to use sets of justifications to partially encode the meaning of connectives, as, e.g., in [McAllester 1980, de Kleer 1986b, Dressier 1988]. We do not consider such techniques in the present paper.

NMF'S theory provides a direct yet implementation independent specification of justification-based truth-maintenance. It can be easily extended to non-monotonic rules with variables [Reinfrank and Freitag, 1988]. The following results on NMFS are needed in order to formally prove their equivalence to the particular subclass of autoepistemic theories to be introduced later.

Lemma 2.6: Let J be an NMFS, E be a .J-extension, $J^*\subseteq J$ be the set of all E-valid justifications. Then $E=\{c: <A|B\rightarrow c>\in J^*\}$.

Lemma 2.7: Let J be an NMFS, E a J-extension. There is no proper subset $E'\subsetneq E$ such that E' is a J-extension.

3 Autoepistemic Logic and Truth Maintenance

3.1 Strongly Grounded AEL-Extensions

AEL IMoore, 1985] is a cleaned-up version of McDermott and Doyle's first shot at a modal non-monotonic logic [McDermott and Doyle, 1980] In this section, we consider an AEL language based on a propositional logic. In [1988a], Konolige develops the concept of a strongly grounded AEL-extension of a given base set. It is meant to formalize AEL-extensions in which a formula p always has a derivation independent of Lp. That is, Lp itself can only be derived from p, and hence strongly grounded AEL-extensions are a candidate for a logic of justified as opposed to simple belief. The definition of strong groundedness is partly syntactical, since the base set is assumed to be in a particular normal form.

Def. 3.1 (Konolige): An AEL formula p is in *normal form* *iff* $\rho\equiv\neg L\alpha\vee L\beta_1\vee\ldots\vee L\beta_n\vee\gamma$, where all of α,β_i, and y are non-modal formulae. Any of the disjuncts, except for y, rnay be absent.

For a set S of formulae, let $LS=\{L\rho: \rho\in S\}$, $\neg LS'=\{\neg L\rho: \rho\notin S\}$, and So be the set of non-modal formulae contained in S.

Def. 3.2 (Konolige): T is an *AEL-extension* of A iff

$$T=\{\rho: A\cup LT\cup\neg LT'\models\rho\}$$

Def. 3.3 (Konolige): Let A be in normal form, T an AEL-extension of A. Let A* be those formulae in A for which βi€T, for all Lpj. T is *strongly grounded* iff

$$T=\{\rho: A^*\cup LA^*\cup\neg LT_0'\models_{SS}\rho\}$$

The sign $\models_{SS}$ here means that the modal index of autoepistemic valuations, which consist of an ordinary propositional interpretation and a set of beliefs, is restricted to stable sets. For details, see [Konolige, 1988al. We use a slightly different definition of strong groundedness as the one given there. This modification is necessary both for Konolige's results on the relation between strongly grounded AEL-extensions and default logic, and for our purposes here. It is due to Konolige [1988b).

A weaker notion of groundedness only requires minimalitv w.r.t. non-modal formulae. It is equivalent to

Def. 3.4 (Konolige): Let T be an AEL-extension of A, which is not necessarily in normal form T is *moderately grounded* in A iff

$$T=\{\rho: A\cup LA\cup\neg LT_0'\models_{SS}\rho\}$$

Simple AEL-extensions are called *weakly grounded.* AEL-extensions, and hence also strongly grounded extensions, are uniquely determined by the set of non-modal formulae contained in them, their so-called *kernel.*

Lemma 3.5 (Konolige): If two AEL-extensions agree on their non-modal formulae, they are the same.

3.2 The Translation from NMFS to AEL

We transform justifications into AEL-formulae as follows.

Def. 3.6: Let $<a_1,\ldots,a_m|b_1,\ldots,b_n\rightarrow c>$ be a justification. Its *AEL-transform* is $\neg L(a_1\wedge\ldots\wedge a_m)\vee Lb_1\vee\ldots\vee Lb_n\vee c$.

Since, in AEL, $La_1\wedge\ldots\wedge La_m$ is equivalent to $L(a_1\wedge\ldots\wedge a_m)$, and using the definition of $\supset$, the AEL-transform of a justification can be rewritten as

$$La_1\wedge\ldots\wedge La_m\wedge\neg Lb_1\wedge\ldots\wedge\neg Lb_n\supset c,$$

where all of ai, bj, and c are non-modal atoms from V. For a premise $<\emptyset|\emptyset\rightarrow c>$, we get the atom c. Let JAEL be the set of transformed justifications of an NMFS J. Given this particular form of a base set, and by view of Lemma 3.5, the corresponding strongly grounded AEL-extensions are uniquely determined by the set of non-modal atoms contained in them. It is exactly this atomic kernel that is computed by a TMS.

For a set S of formulae, let At(S) be the set of propositional atoms in S, and Th(S) be the set of propositional logic consequences of S.

Theorem 3.7: Let J be an NMFS with AEL--transform JAEL

(1) Suppose T is a strongly grounded AEL-extension of JAEL Then At(To) is a J-extension.
(2) Conversely, let E be a J-extension. Then Th(E) is the kernel of a strongly grounded AEL-extension T of JAEL

3.3 Discussion

It is important to note that the theorem no longer holds if the justifications are formed from a full-fledged propositional language rather than from a set of propositional atoms. The AEL base set {p, $Lp\wedge\neg L(p\vee q)\supset r$} has obviously no extension that contains r, since from p we get $p\vee q$, and hence $\neg L(p\vee q)$ cannot be assumed. But {p,r} is a $\{<p|(p\vee q)\rightarrow r>,<\emptyset|\emptyset\rightarrow p>\}$-extension.

Also, the condition of strong groundedness is necessary, since $\{Lp\supset p, \neg Lp\supset q\}$ has a moderately grounded AEL-extension containing p, which is not strongly grounded. The corresponding NMFS has only {q} as an extension.

This poses the question as to NMFS-counterparts to weakly or moderately grounded AEL-extensions. In some independent piece of work and cast in a quite different formalism, Fujiwara and Honiden [1989] show that weakly grounded AEL-extensions correspond to closed and *locally* grounded J extensions, in the same sense as in theorem 3.7 In NMFS-theory, it is easy to show that in the monotonic case minimality of closed sets is sufficient to guarantee global groundedness. For a non-monotonic NMFS, however, even minimality of locally grounded sets is insufficient, as the example above shows. Similarlv, the minimization involved in the definition of moderately grounded AEL-extensions is too weak to capture the notion of grounded belief used in the TMS-world.

An additional filter is required for moderately grounded extensions. The definition of strong groundedness provides exactly that filter. (Note that it is related to a property of NMFS extensions given in Lemma 2.6.) As we elaborate in the long paper, this extra condition is similar in spirit to *stability* conditions in model-preference theories for default logic [Etherington, 1988] and logic programming [Elkan, 1989].

To our opinion, all of this may eventually shed some new light on the relation between proof theory and model theory in non-monotonic reasoning. The concept of a finite, non-circular proof which is valid relative to an overall coherent state of belief simply has some proof theoretic flavor which is hard to represent independently in the model theory.

3.4 Semantical Considerations

Suppose we are given two justifications $<a|b\rightarrow c>$ and $<\emptyset|\emptyset\rightarrow a>$. The corresponding AEL-base set is $\{a, La\wedge\neg Lb\supset c\}$. Figure 1 shows the usual graphical representation of justifications. A TMS-labelling procedure proceeds as follows: a is a premise, so it is necessarily believed and labelled IN. In AEL, this corresponds to the K45 [Konolige, 1988al inference step a/La. Since there is no valid justification for b, b is labelled OUT, which reflects the AEL-assumption Lb. Now the justification for c has become valid, so it must be the case that c holds, and hence it is also labelled IN That is, a

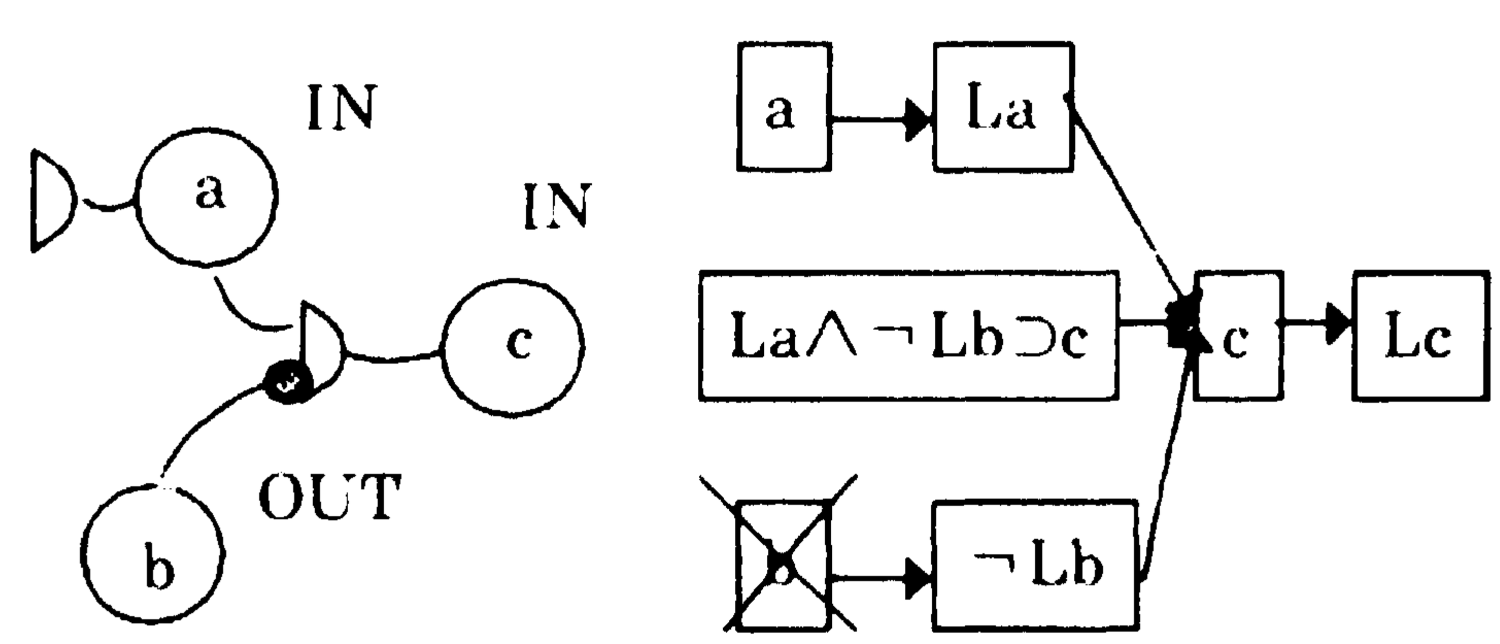

Figure 1: TMS-labeling and AEL-inferences

justification gets the following semantic interpretation, if a is believed (IN the database) and b is not believed (OUT of the database) then it must be the case that c is true and, consequently, must also be believed (IN). Note that this interpretation is different from the usual reading for a justification, where the direct and only conclusion is the self-belief that c must be IN the database, without referring to the world in consideration.

A proper distinction between ¬La and ¬a is crucial to an understanding of truth maintenance. While there is no a priori preference for either of a and ¬a, and hence it might be the case than none of them is believed, ¬La is always preferred to La. I.e., a TMS only adopts the self-belief La if it is forced to do so because it follows from a itself. Otherwise it jumps to the assumption ¬La. As we will see in chapter 5, a confusion of a proposition being OUT and its negation being IN may lead to peculiar consequences.

3.5 Default Logic

AEL and default logic IReiter, 1980] are essentially equivalent [Konolige, 1988a], in that the kernels of strongly grounded AEL-extensions correspond to extensions of default theories, where an AEL formula $La \wedge \neg Lb \supset c$ is translated to a default $a{:}M\neg b/c$. (Since default logic does not preview an empty M-part, dummy conditions are introduced if needed). Theorem 3.7 thus yields a relationship between NMFS and default logic as a corollary.

Corollary 3.8: Let J be an NMFS and Δ_J be the default theory constructed from JAEL

(1) Suppose T is a default logic extension of Δ_J. Then At(T) is a J-extension.
(2) Conversely, let E be a J-extension. Then Th(E) is a default logic extension of Δ_J.

This relationship is of some interest in its own right, because default logic has been given a semantics [Etherington, 1988] in terms of *stable, maximally preferred sets of models.* We show in the long paper [Reinfrank and Dressier, 1988] that the computations performed by a TMS can be regarded as operations on a condensed representation of such model sets.

4 Assumption-Based Systems

Theorem 3 7 begs the question as to a related theorem for a non-monotonic ATMS [Dressier, 1988]. A minor extension to NMFS-theory is sufficient to model an ATMS. We must allow for *assumptions* from a given set $\alpha \subseteq V$ to be used in J-proofs.

Def. 4.1: Same conditions as in Def. 2.1, and $\alpha \subseteq V$. A *J-proof relative to* a is a J-proof as in Def 2.1., condition (3) replaced by:

$$(3')\ \forall q_i: \underline{q_i \in \alpha \vee} \exists <A|B \rightarrow q_i> \in J: A \subseteq \{q_1, \ldots q_{i-1}\} \wedge B \cap S = \emptyset$$

The definitions of *J-grounded relative to* a and *J-extension of a* then are straightforward generalizations of Def. 2.3 and 2.5. An ATMS works with monotonic justifications of the form $<A|\emptyset \rightarrow c>$. Jextensions for a monotonic NMFS always exist and are unique.

Lemma 4.2: Let J be a monotonic NMFS, $\alpha \subseteq V$. There is one and only one J-extension of a. We write $J_{ext}(a)$.

Given a distinguished set $Ass \subseteq V$, an ATMS simultaneously computes all $J_{ext}(a)_y$ $\alpha \in 2^{Ass}$. To integrate non-monotonic justifications into the essentially monotonic ATMS-world, explicit Out-atoms are introduced. For a given V, let OutV = {Out x: $x \in V$}. Out-atoms may not occur as the consequent of a justification.

Def. 4.3: Let $<a|b \rightarrow c>$ be a non-monotonic justification. Its (monotonic) ATMS-transform is $<a, \text{Out } b|\emptyset \rightarrow c>$.

Obviously, there is a one-to-one correspondence between non-monotonic formal svstems over V and those monotonic formal systems over VuOutV that do not permit Out-atoms as consequents. Non-monotonicity in an ATMS then is achieved by manipulating assumption sets. A *J-µ-extension* E of an assumption set a is an ordinary extension of a maximal augmentation of a with a set (3 of Out-atoms. Out x is added to the basis $\alpha \cup \beta$ if and only if x€E.

Def. 4.4: Let J be a monotonic NMFS over $V \cup OutV$, $\alpha \subseteq V$. E is a *J-µ-extension of* α iff $\exists \beta \subseteq OutV$. $E = J_{ext}(\alpha \cup \beta) \wedge (\forall x \in V: \text{Out } x \in \beta \Leftrightarrow x \notin E)$.

It is important to note that, unlike in [Dressier, 1988], J-p-extensions do not involve any so-called nogood inferences. It is exactly these nogood inferences that lead to problems of ungroundedness in an NMATMS, cf next section.

Lemma 4.5: Let J be a non-monotonic NMFS over V, JATMS its ATMS-transform over $V \cup OutV$. Let $\alpha \subseteq V$ be an assumption set.

(1) Suppose E is a J-extension of a and let $\beta = \{\text{Out } x: x \notin E\}$. Then $E \cup \beta$ is a JATMSP" extension of a, with basis $\alpha \cup \beta$.

(2) Conversely, let S be a J_{ATMS}-extension of α. Then $S \cap V$ is a J-extension of a.

It is easy to see that every J-proof relative to a can be simulated with a JATMS-Peoof relative to $\alpha \cup \{Out\ x: x \notin E\}$. Conversely, we can construct a J-proof for every ordinary atom in a J_{ATMS}-μ-extension. Since the transform is bidirectional, Lemma 4.2 also works in the opposite direction (i.e. starting from a monotonic NMFS over VuOutV) and yields a one-to-one correspondence between u-extensions for the NMATMS and strongly grounded AEL-extensions in the sense of Theorem 3.7 as a corollary.

5 Backtracking and Nogood Inferences

Consider an extended language including justifications of the form $<a|b \rightarrow \bot>$, where $\bot$ stands for falsity. Let a be a premise. Since there is no v\ay to derive b in the basic TMS-machinery, we get an inconsistent extension $\{a, \bot\}$. A dependency-directed backtracking routine, which is triggered whenever $\bot$ becomes IN, would in this case introduce an additional justification $<a|\emptyset \rightarrow b>$ Similarly, a nogood inference rule in the NMATMS would i n $<a|\emptyset \rightarrow b>$ o m $<a,\ Out\ b|\emptyset \rightarrow \bot>$ This yields $\{a,b\}$ as the atomic kernel of an AEL-extension. It is strongly grounded in the extended base set. However, from the original base set $\{a, La \wedge \neg Lb \supset \bot\}$, we can onlv infer La and $\neg La \vee Lb$, and hence get the ungrounded assumption Lb, independently of b.

In general, dependency-directed backtracking in a TMS and nogood inference rules in an ATMS can be considered as justification-generating rules of the form

$$\rho \wedge \neg Lq \supset \bot \;/\; \rho \supset q,$$

with approriate control restrictions. This leads to modified base sets with new conclusions that are possibly not strongly grounded w.r.t. the original base set. A peculiar problem arises in an NMATMS due to justifications of the form $<x,\ Out\ x|\emptyset \rightarrow \bot>$, which are used to prevent x and Out x from being assumed simultaneously. This technique appears to be reasonable for de Kleer's negated assumptions fde Kleer, 1988]. However, in the case of genuinely non-monotonic Out-assumptions, together with nogood inferences, it leads to a confusion of inconsistency and incoherence, that is, non-existence of any extension. This difference also suggests to use negated assumptions rather than Out-assumptions to encode logical connectives with sets of justifications.

Applied to the incoherent set $\{<Out\ p|\emptyset \rightarrow p>\}$, for example, a simple nogood inference step yields $\{p\}$ as an extension. The corresponding AEL-base set has no extension at all, neither consistent nor inconsistent. More work is needed to come to a better understanding of the properties of backtracking and nogood inferences. [Morris, 19881 seems to be a first step in the right direction.

6 Related Work

Using the same translation as we do, but described in a quite different formalism, Fujiwara and Honiden 11989] prove a relation, similar to the one presented in theorem 3.7, between weakly grounded AEL-extensions and what we call closed and locally grounded NMFS-extensions.

Elkan 11989] has recently proposed a different translation, $a \wedge \neg Lb \supset c$ (a instead of La). For the simple form of an AEL-language in consideration, it turns out that the weakly grounded extensions of the resulting AEL base sets correspond to the strongly grounded extensions using our translation.

I.e., three groups of researchers have come up independently of each other with related ideas and results, though formulated in superficially quite different formalisms. We consider this as evidence that we are on the right track, and that it was getting to be time to put non-monotonic truth maintenance on a sound logical basis. What is unique to our approach, compared to (Fujiwara and Honiden, 1989] and [Elkan, 19891, is the development of NMFS as a direct theory of truth maintenance, as well as the treatment of ATMSs.

Our research is similar in spirit to the work by Horty, Thomason, and Touretzky on logical theories of inheritance, cf. [Thomason and Horty, 1988]. They also pursue a bottom-up approach using intermediate direct formalizations. The situation in inheritance theory is comparable to the one in truth maintenance in that, in both fields, procedural realizations and network-based concepts were the starting point. We are currently working on a characterization of truth maintenance in a four-valued logic, which is related to Thomason's approach.

7 Conclusions

We have established a relationship between non-monotonic truth maintenance and non-monotonic logics. This provides the technical fundament for an alternative view of truth maintenance as inference in a non-monotonic calculus, in addition to the traditional view as an efficient mechanism for search and caching. We also have developed a direct theory of truth maintenance and presented a correct encoding of non-monotonic justifications in an assumption-based TMS. Our results contribute to closing an important theory-implementation gap, after TMSs and non-monotonic logics have been coexisting for more than a decade. They open the way for further research into the foundations of truth maintenance.

8 Proofs

Complete proofs, including an independent proof for Corollary 3.8, for the claims made in this paper are contained in [Reinfrank and Dressier, 1988].

Acknowledgements

Our work benefited from discussions with Frank Zetsche and the members of the Advanced Reasoning Methods Group.

References

[Brown, 1988] Allen L. Brown, Jr.: Logics of Justified Belief *Proceedings ECAI-88,* pp. 507-512, 1988.

[de Kleer, 1986al Johan de Kleer: An Assumption-Based TMS. *Artificial Intelligence 28,* pp. 127-162, 1986.

[de Kleer, 1986b] Johan de Kleer: Extending the ATMS. *Artificial Intelligence 28,* pp. 163-196, 1986.

[de Kleer, 1988] Johan de Kleer: A General Labelling Algorithm for the ATMS. *Proceedings AAAI-88,* pp. 188-192, 1988

IDoyle, 1979] Jon Doyle: A Truth Maintenance System. *Artificial Intelligence 12,* pp. 231-272, 1979.

[Doyle, 1983] Jon Doyle: Some Theories of Reasoned Assumptions. Tech report CMU-CS-83-125, 1983.

[Dressier, 1988] Oskar Dressier: An Extended Basic ATMS. *Proceedings of the 2nd Intl. Workshop on Non-Monotonic Reasoning.* Springer LNCS 346, pp.143-163,1988.

[Elkan, 1989] Charles Elkan: A Rational Reconstruction of Nonmonotonic TMSs. Submitted for publication.

[Etherington, 1988] David Etherington: *Reasoning with Incomplete Information.* Pitman 1988.

[F'ujiwara and Honiden, 1989] Yasushi Fujiwara and Shinichi Honiden: Relating the TMS to Autoepistemic Logic. *Proceedings IJCAI-89,* this volume, 1989.

[Goodwin, 1987] James W. Goodwin: A Theory and System for Non-Monotonic Reasoning. PhD Dissertation, University of Linkoping.

[Konolige, 1988a] Kurt Konolige: On the Relation between Default and Autoepistemic Logic. *Artificial Intelligence 35,* pp. 343-382,1988.

[Konolige, 1988b] Kurt Konolige: personal communication.

[McAllester, 1980] David McAllester: An Outlook on Truth Maintenance. MIT AI-Lab Memo 551.

[McDermott and Doyle, 19801 Drew McDermott and Jon Doyle: Non-Monotonic Logic I. *Artificial Intelligence 13,* pp. 41-72, 1980.

[Moore, 1985] Robert Moore: Semantical Considerations on Non-Monotonic Logic. *Artificial Intelligence 25,* pp. 75-94, 1985.

[Morris, 1988] Paul H. Morris: Autoepistemic Stable Closures and Contradiction Resolution. *Proceedings of the 2nd Intl. Workshop on Non-Monotonic Reasoning.* Springer LNCS 346, pp. 60-73, 1988

[Reinfrank, 1987] Michael Reinfrank: Lecture Notes on Truth Maintenance, Part I: Fundamentals. RKL Memo 87-04, Linkoping University (Sweden), 1987.

[Reinfrank and Freitag, 1988] Michael Reinfrank and Hartmut Freitag: Rules and Justifications, A Uniform Approach to Reason Maintenance and Non-Monotonic Inference. *Proceedings FGCS-88,* pp. 439-446, Tokyo, 1988.

[Reinfrank and Dressier, 1988] Michael Reinfrank and Oskar Dressier: On the Relation between Truth Maintenance and Nonmonotonic Logics. Siemens Report INF 2 ARM 11 88, 1988

[Reiter 80| Raymund Reiter: A Logic for Default Reasoning. *Artificial Intelligence 13,* pp. 81-132, 1980.

[Reiter and de Kleer, 1987] Raymund Reiter and Johan de Kleer: Foundations of Assumption-Based Truth Maintenance. *Proceedings AAAI-87,* pp. 183-188,1988.

[Thomason and Horty, 1988] Richmond Thomason and John Horty: Logics for Inheritance Theory. *Proceedings of the 2nd Intl. Workshop on Non-Monotonic Reasoning.* Springer LNCS 346, pp. 220-237,1988.

ON THE RELATION BETWEEN AUTOEPISTEMIC LOGIC AND CIRCUMSCRIPTION
Preliminary Report

Kurt Konolige*
Artificial Intelligence Center *and*
Center for the Study of Language and Information
SRI International, Menlo Park, California 94025/USA

Abstract

Circumscription on the one hand and autoepistemic and default logics on the other seem to have quite different characteristics as formal systems, which makes it difficult to compare them as formalizations of defeasible connmonsense reasoning. In this paper we accomplish two tasks: (1) we extend the original semantics of autoepistemic logic to a language which includes variables quantified into the context of the autoepistemic operator, and (2) we show that a certain class of autoepistemic theories in the extended language has a minimal-model semantics corresponding to circumscription. We conclude that all of the first-order consequences of parallel predicate circumscription can be obtained from this class of autoepistemic theories. The correspondence we construct also sheds light on the problematic treatment of equality in circumscription.

1 Introduction

The relations between the major nonmonotonic logic formalisms of AI — default logic, autoepistemic logic, and circumscription — is of some importance, since all of these logics have been proposed as formalisms for various types of commonsense reasoning. The basic formal equivalence of default and autoepistemic logic has already been shown (see [Konolige, 1987]), but the relation between circumscription and default or autoepistemic logic remains obscure. Mostly this is a consequence of the different foundations of these logics: circumscription is based on a minimal-model semantics (see [Lifschitz, 1985]), while the others use more proof-theoretic techniques (default logic [Reiter, 1980]) or an epistemic operator (autoepistemic logic [Moore, 1985]).

In trying to express autoepistemic or default, logic in circumscription, researchers have found the basic problem to be that a minimal-model or even prefered-model semantics simply does not have the capability of representing the requisite proof-theoretic or epistemic concepts (see [Shoham, 1987]). We agree with this assessment, and say nothing further about it here.

On the other hand, there have been several results on expressing circumscription in default logic. These results are summarized in [Etherington, 1986]; they apply to the restricted case of predicate circumscription with no fixed predicates and with a finite, fixed domain.

From a model-theoretic point of view, the predicate circumscription Circum(,4; *P; Z)* of a first-order sentence *A* picks out those models of *A* in which the extension of the predicate *P* is minimal. The comparison is across models with the same domain and denotation function, but which might differ in the extensions of the predicates *Z*. All predicates other than *P* and *Z* are *fixed,* that is, cannot vary in a comparison of models. It was recently shown (see [de Kleer and Konolige, 1989]) that fixed predicates are inessential in predicate circumscription, that is, there is a simple translation from any circumscription with fixed predicates to one without. Hence fixed predicates no longer present an obstacle to representing circumscriptions in default or autoepistemic ogic.

The problem of finite domains remains, however. In this paper we provide a solution to this problem, by first extending autoepistemic logic to a language which allows quantifying into the epistemic operator, and then showing that a certain class of autoepistemic theories, the MIN= theories, express all of the first-order consequences of predicate circumscription.

2 Semantics of Quantifying-in

Autoepistemic (AE) logic was defined by [Moore, 1985] as a formal account of an agent reasoning about her own beliefs. The agent's beliefs are assumed to be a set of sentences in some logical language augmented by a modal operator *L.* As originally defined, and extended in [Konolige, 1987], its language does not permit variables quantified outside of a modal operator to appear inside. In this section we further extend AE logic to deal with quantifying-in.

2.1 Logical preliminaries

We begin with a language *C* for expressing self-belief, and introduce valuations of *C.* The treatment generally

*This research was supported by the Office of Naval Research under Contract No. N00014-85-C-0251, by subcontract from Stanford University under the Defense Advanced Research Projects Administration under Contract No. N00039-84-C-0211, and by a gift from the System Development Foundation.