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.

follows and extends [Konolige, 1987]

Let $\mathcal{L}_0$ be a first-order language with equality and functional terms. The normal formation rules for formulas of first-order languages hold. A *sentence* of $\mathcal{L}_0$ is a formula with no free variables; an *atom* is a sentence of the form *P(t1,* • • • , tn). We extend $\mathcal{L}_0$ by adding a unary modal operator *L;* the extended language is called $\mathcal{L}$. $\mathcal{L}$ can be defined recursively as containing all the formation rules of $\mathcal{L}_0$, plus the following:

If ϕ is a formula of $\mathcal{L}_0$, then so is $L\phi$. (1)

An expression $L\phi$ is a *modal atom.* Sentences and atoms of $\mathcal{L}_0$ are called *ordinary.* Note that nestings such as $LL\phi$ are not allowed; we consider only a single level of nesting here. Because the argument of a modal operator can contain free variables, there may be quantifying into the scope of a modal atom, e.g., $\exists x LPx$ is a sentence. Often we will use a subscript "0" to indicate a subset of ordinary sentences, e.g., $\Gamma_0 = \Gamma \cap \mathcal{L}_0$.

Let us restrict ourselves for the moment to modal atoms which do not contain free variables. From the point of view of first-order valuations, the modal atoms $L\phi$ are simply nilary predicates. Our intended interpretation of these atoms is that ϕ is an element of the belief set of the agent. So we will consider valuations of $\mathcal{L}$ to be standard first-order valuations, with the addition of a belief set *T.* The atoms $L\phi$ are interpreted as true or false depending on whether ϕ is in T. To distinguish these valuations, we will sometimes call them *L valuations.*

The interaction of the interpretation of *L* with first-order valuations is often a delicate matter, and so a perspicuous terminology for talking about *L* valuations is necessary. In particular, it is often useful to decouple the interpretation of modal and ordinary atoms. First-order valuations are built upon the truthvalues of atoms: for ordinary atoms, truthvalues are given by a structure *(U,v,R),* where *v* is a denotational mapping from terms to elements of the universe U, and R. is a set of relations over *U,* one for each predicate. We will refer to any such structure as an *ordinary index,* and denote it with the symbol I. Modal atoms are given a truthvalue by a belief set T, which is called a *modal index.* Note that, because modal operators are not nested, only the ordinary sentences of the modal index *V* are important.

The truthvalue of any sentence in $\mathcal{L}$ can be determined by the normal rules for first-order valuations, given an ordinary and modal index. We write $\models_{I,\Gamma} \phi$ if a valuation (/, T) satisfies ϕ. The valuation rule for modal atoms can be written as

$$\models_{I,\Gamma} L\phi \quad \text{if and only if} \quad \phi \in \Gamma \; . \qquad (2)$$

A valuation that makes a every member of a set of sentences true is called a *model* of the set. A sentence that is true in every member of a class of valuations is called *valid* with respect to the class.

We use Cn(X) to mean the first-order consequences of a first-order set of sentences *X.*

2.2 Autoepistemic Extensions

In [Konolige, 1987], we informally defined an extension of a set of sentences *A* as those consequences of *A* which an agent, should believe. The formal counterpart is given by:

DEFINITION 2.1 *Any set of sentences T which salysfi.e.s the equation*

$$T = \{\phi \mid A \models_{T_0} \phi\}$$

is an autoepistemic extension *of A.*

This is a fixed-point equation for a belief set /', and is a candidate for the belief set of an ideal introspective agent with premises .4. It differs slightly from the original definition of [Konolige, 1987] in that the modal index consists of the ordinary part (the *kernel* of 7^1), this suffices because the language of T does not include nested modal operators.

Note that we are considering all models of *A* in which the interpretation of $L\phi$ is the belief set of the agent itself, that is, the valuations we consider all have a modal index that is the belief set of the agent, following Moore, we call such valuations *autoepistemic* (or *AE)* valuations.

2.3 Quantifying-in

We would like to extend the language of autoepistemic logic to include variables which are quantified outside, the scope of the modal operator, but can appear inside, e.g., $\exists x.LPx$. The problem is that it is not obvious how to extend the semantics of the logic to deal with these "quantified-in" expressions. Recall that the belief set T is a set of sentences that form the beliefs of an agent.. To interpret $L\phi$, we simply ask whether the expression øis in T. But with the quantified-in language, we must also be able to interpret $L\phi(x)$, whe$\phi(x)$ is the proposition that the individual *x* has the property ϕ. In order to construct a propositional expression whose meaning is $\phi(x)$, we must have some way of refering to individuals in the domain.

The simplest, scheme for reference to domain elements is to use the denotation map *v* already present in the first-order mterpretion /. In place of the valuation rule for modal atoms given above, we use:

$$\models_{I,\Gamma} L\phi(x) \quad \text{iff} \quad \text{for some term / such that} \quad (3)$$
$$v(t) = x,\ \phi(t) \in \Gamma.$$

That is, we say that *<ø>(x)* is believed if *x* has a name / such $\phi(t)$ is believed.

Given this addition to the valuation rule, we can use the definition of autoepistemic extension above for the case in which *A* contains quantified-in expressions. However, we will make one technical stipulation that will be useful in later developments, to insure that there are enough "free" names in the language $\mathcal{L}$:

The language $\mathcal{L}$ contains a count ably infinite set of constants C which cannot be used in the premise set A of an extension. (4)

A given individual *x* may have none or many names in a model, a circumstance which leads to some interesting behavior of the fixed-point equation of Definition 2.1. To explore some of these, we first make the observation that the revised definition of satisfiability for modal atoms does not perturb the extensions of any set *A* that contains no quantified-in expressions.

EXAMPLE 2861 Let *A* = *{Pa}.* There is a single extension T of .4, with To - ('n(Po). Therefore, we know that *LPa* and ¬*LPb* are in *T.* By the valuation rule for modal atoms (3), $\exists x.LPx$ will be true in (I,To) if there is some individual *x* such that *x* — *v(a).* Every interpretation / has some such individual, and hence $\exists x.LPx$ is true in every (/,7b) model *of A,* and hence in 7\

Another interesting sentence contained in *T* is s : $\forall x.x{\neq}a \supset \neg LPx.$ To see why this is so, let *x* be an individual with ,$x{\neq}v(a)$. Since *Pa* is the only ground occurrence of the predicate in 7b, it must be the case that *LPx* is false in any model (7,7b) of *A*} and hence s is an element of *T.*

On the other hand, consider a similar sentence .$' : $\forall x.x = b \supset \neg LPx.$ It might be suspected that *s'* is a member of 7^1, but this is not the case. For although *x* is the denotation of/;, it may also be the denotation of *a* in some first-order interpretation, and for (his interpretation, *LPx* will be true. So *s'* will not be true in all (/,7b) models of A

This example highlights a curious situation that occurs when knowledge of properties of individuals hinges on having a name for that individual: the epistemic operator expresses knowledge of the intension of a term. Let, us take P to be the property of being rich, *a* to be the mayor, and *b* to be the former police chief. We have proof thai the mayor is rich (LPa) and no evidence that the former police chief is $(\neg LPb)$. These are statements about the *intension* of the terms *a* and /;, that is, the mayor, *whoever he is,* is rich. On the other hand, the expression *LPx* when *x* is a quantihed-in variable says that we know *Pc* to be true for some intensional concept r whose denotation is *x.* Now if we were to know that a particular individual *x* is the former police chief, we still cannot, say that we have no evidence that *x* is rich, because *x* may also be the mayor.

Another consequence of the intensional nature of the epistemic operator is that even though a universal statement may be true in an AE valuation, its substitution instances may not. Consider the valuation (7,T), where $\models_I\ a{=}b \wedge \forall x.Px \supset x{=}a$ and r = Cn(Pa). We must have $\models_{I,\Gamma} \forall x.(LPx \vee \neg Px)$, because if $x = v(a)$, *LPa* is true; and if not, ¬*Px* is true. However, the substitution instance *LPbV* ¬*Pb* is not true in this valuation: *LPb* is not a member of *V,* and *Pb* is true in 7 because *a—b.*

Finally, we note that the Barcan formula $\forall xLPx \supset L\forall xPx$ is true in every AE valuation, while the converse $L\forall xPx \supset \forall xLPx$ may he false. The reason for the latter is that even though every individual *x* has the property P, some individuals may not be given a name in the AE valuation, and so *LPx* will he false.

This scheme for extending the semantics of L valuations to the quantified-in case is similar to that proposed for the simple epistemic operators in [Konolige, 1984]. It differs from the approach of [Levesque, 1982, Levesque, 1987] in that it is based on the intension of terms rather than their denotation. Nevertheless there are many points of similarity between the two approaches that we have not investigated.

3 MIN Theories

So far we have only looked at extensions of sets of first-order sentences. The normal definition of extension (2.1) earned over to the extended language, with only a change in the valuation rule for modal atoms, and a slight restriction on the constants appearing in the premise set,. We now examine a class of first-order sentences that we call MIN theories. Any such theory has the form

$$W \bigcup_i \{\forall x(\neg LP_ix \supset \neg P_ix)\}\ , \qquad (5)$$

where *W* is a *finite* set of first-order sentences and the P, are a sequence of predicates. We write M(W; P1, .. . P_n) to indicate the MIN theory of *W* over the predicates P_t.

The idea behind MIN theories is to select AE valuations in which every individual not known to have the property *Pi* does not have this property, i.e., to minimize the extension of each Pi.

EXAMPLE 3.1 Let *W* = {Fa}, and let *S* - $\mathrm{Cn}(\forall x.Px \equiv x{=}a)$. Every valuation *(l,S)* which satisfies M(IT; P) makes *Px* true for *x* — *v(a),* and false for every other *x.* Thus, if we define *T* by

$$T = \{\phi | \mathrm{M}(W;P) \models_S \phi\}\ ,$$

it is clear that 7b — *S,* and hence *T* is an extension of M(IT; P). In fact it is the only extension.

Let IT - *{Pu* V P/;}, and as before let 5 = $\mathrm{Cn}(\forall x.Px$ = *x—a).* Again we can show that every valuation *(l, S)* satisfying *M(W,P)* satisfies $\forall x.Px$ = *x=a,* and so *S* is the kernel of an AE extension of M(W';P). In this case the extension is not unique, there is another one whose kernel is Cn(Vx\Px,* = *x.—h).* The sentences which these two extension have in common are all first-order consequences of $(\forall x.Px \equiv x = a) \vee (\forall x.Px \equiv x = b)$.

Let *W* - $\{\exists xPx\}$, and let $S = \mathrm{Cn}(\forall x.Px \equiv x{=}a)$. Again we can show that every valuation (7,5) satisfying M(IT;P) satisfies Vx.Pa: = *z~a,* and so *S* generates an AE extension 7'. But the choice of the constant a was arbitrary, and we can use any other constant in defining *S.* fience there are an infinite number of extensions of M(W;P); the sentences they have in common are the first-order consequences of $\exists!xPx$.

These examples are very suggestive of a correspondence between MIN theories and the minimal models of *W*. However, there is one essential point of difference. A model / of *W* is minimal in P_t if there is no other model *with, the same universe and denotation function* whose extensions of /\ are properly included in those of *I.* (Note that we are assuming here that the extensions of all predicates oilier than the P_{ti} can vary across compared models. In the next, subsection we consider the case *of fixed* predicates.) Comparisons are not made between models with different domains and different denotation functions, and hence choosing only the minimal models of *W* will not lead to any conclusions about the equality relation or the size of the domain not already apparent in *W* (this point lias been noted in [Etherington and Reifer, 1984]). On the other hand, this is not true of the extensions of MIN theories: an extension can contain conclusions not present in \V about equality anions, terms. For example, the set W = *{¬Pa)* *a — b}* does not have *a* = *b* among its fust-order consequences, yet. the single extension of M(W, *P)* does. To get the correspondence between MIN theories and minimal models correct, we need to fix the interpretation of equality in the former.

3.1 Fixing predicates

We extend MIN theories by adding a set of predicates, the *fixed* predicates, to the original definition. A MIN theory now has the form

$$\begin{aligned} W \quad & \bigcup_i \{\forall x(\neg LP_i x \supset \neg P_i x)\} \\ & \bigcup_i \{\forall x(\neg LQ_j x \supset \neg Q_j x)\} \qquad (6) \\ & \bigcup_i \{\forall x(\neg L\neg Q_j x \supset Q_j x)\}\ , \end{aligned}$$

where the Q_i are a sequence of predicates. We write *M(W; P1, . . .* P_n*',Q1,..... Qm)* to indicate the M1N theory of *W* over the predicates P_i with *Qj* fixed.

Note that, to fix Q, both *Q* and its negation ¬Q are minimized. In general this will lead to multiple extensions in which various combinations of *Qx* and ¬*Qx* hold for each individual *x.*

EXAMPLE 3.2 Let *W* = $\{\neg Pa \supset Qa\}$. The M1N theory *M(W',P\Q)* has two classes of extensions: one class contains *{¬Pa,Qa},* while the other contains *{¬Qa,Pa}.* Thus the minimization of *P* does not force the acceptance of *Qa* in every extension.
The presence of a fixed *Q* actually creates an infinite number of extensions because of the presence of the countable set *C* of constants in L We will consider extensions to be equivalent if they differ only in sentences containing these constants; in this case, there are just two nonequivalent extensions.

The equality predicate can be fixed, just as any other predicate, and it is the class of MIN theories M(W; *P\ —)* that we consider in relating extensions to minimal models: call these MIN= theories. We now develop the result that a first-order sentence is true in the P-minimal models of *W* just in case it is true in every extension of W*(W;P-=).*

3.2 Parameter models

A first-order interpretation in which every individual *x* is denoted by some term is called a *parameter* interpretation. Herbrand interpretations are one type of parameter interpretation, in which every term denotes itself. Parameter interpretations are more general than Herbrand interpretations, since in the former two terms can refer to the same individual.

Just as Herbrand interpretations are a sufficient semantics for universal prenex sentences, so too parameter interpretations suffice for sets of first-order sentences. By "suffice" we mean that any such set *W* has a model if and only if it has a parameter model. Note that this statement is not true in general if *W* can contain members of *C:* for example the set $\{\exists x \neg Px, Pc_1, Pc_2, \ldots\}$, in which *Pci* is asserted for every constant $c_i \in C$, has a model but no C-parameter model.

A second interesting property of parameter models is that the *P*minimal parameter models of a. finite set W are sufficient, for minimal entailment, as we now show.[1]

PROPOSITION 3.1 *Let* ϕ *he a sentence of.*$\mathcal{L}_0$ *—C (the base language, with out the constants C), and W a set of sentences of the same language. Then* $\phi \in \mathcal{L}_0$ *is true in all P-minimal countable models of Wif and only if it is true in all P-minimal parameter models.*

Proof. Suppose $W + \phi$ has aP-minimal countable model /. Let the constants C' be those members of *C* not mentioned by <ϕ. We are *free* to construct a model *V* that is the same as /, but in which all elements of / are denoted by one of *C.* *V* must be minimal; if it were not, then there is another model /" with the same denotation function and universe as I', but with a smaller extension of *P.* We can convert I" into a model with the same denotation function as /, which must then be less than /, a contradiction.
In the converse direction, if $W + \phi$ has a minimal parameter model, it obviously has a minimal countable model.

Finally, as the next proposition shows, parameter models are the only ones we need consider in forming extensions of MIN= theories.

PROPosITION 3.2 *Any interpreta11on* (/, T) *is a model of a* MIN= *theory M(W; P; —) only if I is a parameter model.*

Proof. Suppose / isn't a parameter model. Then there is some element *e* of the universe of *1* such that *c* is not denoted by any term. Thus both *-Lx =y* and ¬*L x#y* are true for *x=e,* and this leads to a contradiction in *M(W;P;=).*

3.3 The main theorem

For any P-mininial parameter model / of *W,* we call the *P-diagram* of / the set of ground literals in *P* and = that are true in I. We first, show that the diagram of / picks out a unique extension of M(W;P;=), for which / is a model.

PROPOSITION 3.3 *Let D be the diagram of a P-minimal parameter model I of W. Then (I, D) is a model of some extension of M(W; P;* =).

Proof. This is a sketch of the proof. Let

$$S = \{\phi \in \mathcal{L}_0 | \mathrm{M}(W; P; =) \models_D \phi\}.$$

We first show that the restriction of *S* to ground *P* and equality literals is exactly the set *D.* Note that *D* is complete with respect to equality literals: for all terms *a* and 6, either *a=b* or *a#b* is in D, but not both. From the fixing of equality in *M(W;P;=),* all of these are also contained in 5. *D* is also complete

Note: we will often use a single minimized predicate *P* in propositions in the rest of this paper; the extensions to multiple predicates is obvious.

with respect to P-literals. From the sentence $\forall x.\neg LPx \supset \neg Px$, S contains all negative instances of the P-literals of D. It also must contain all positive instances; if it did not, we could construct a parameter model of W whose P-extension is less than that of I, a contradiction. Finally, S is consistent because $\langle I, D\rangle$ is a model of $\mathsf{M}(W;P;=)$.

Consider the equation

$$T = \{\phi | \mathsf{M}(W;P;=) \models_S \phi\}.$$

Since S and D agree on P and $=$, the T_0 must be equal to S. Hence T is an extension of $\mathsf{M}(W;P;=)$, and $\langle I, D\rangle$ is a model of it.

Example 3.3 This example is from David Kueker (as reproduced in [Perlis, 1986]). Let W be the set $\{Pa,\ \forall x.Px \equiv Ps(x),\ \forall x.a \neq s(x),\ s(x)=s(y) \supset x=y\}$. A P-minimal parameter model I satisfying this is the Herbrand model with Px true for a and all terms of the form $s(s(s(\ldots s(a)\ldots)))$, and false for all other terms. I is isomorphic to the natural numbers, hence the sentence $\forall x.Px \supset ((\neg\exists y.x=sy) \supset x=a))$ is satisfied by I. In fact, this sentence is in every extension of the MIN$^=$ theory of W.

We now prove a suitable converse of the above proposition, namely, that the models of every extension of $\mathsf{M}(W;P;=)$ are minimal in P.

Proposition 3.4 *Let T be an extension of the theory $\mathsf{M}(W;P;=)$. If $\langle I, T_0\rangle$ is a model of $\mathsf{M}(W;P;=)$, then I is a P-minimal parameter model of W.*

Proof. This is a proof sketch. First we show that T_0 is complete with respect to equality and P-literals. For equality, if T_0 doesn't contain $a=b$, then it must contain $a\neq b$ by the equality-fixing sentence of $\mathsf{M}(W;P;=)$. For P-literals, if T_0 doesn't contain Pa, it must contain $\neg Pa$ by the P-minimizing sentence of $\mathsf{M}(W;P;=)$.

From Proposition 3.2 we know that I must be a parameter model. Suppose I is not P-minimal, and I' is a similar parameter model of W less than I. Then $\langle I', T_0\rangle$ is a model of $\mathsf{M}(W;P;=)$, because the equality sentences are satisfied, and the P-minimizing sentence also is. But then the fixed-point equation for T does not hold, since T_0 is complete with respect to P-literals, and I' does not satisfy T_0.

By Proposition 3.3 above, every P-minimal parameter model of *W* satisfies the kernel of some extension of *M(W;P=).* Conversely, by Proposition 3.4 above, the kernel of every extension of *M(W;P;=)* are the sentences true of some class of P-minirnal parameter models of *W.* Given the sufficiency of parameter models for minimal entailment (Proposition 3.1), we have the following theorem.

THEOREM 3.5

The first-order sentences S true in every extension ofM(W; P;=) *are exactly those sentence true in the P-minimal models of W.*

Remarks. From the above theorem we can clarify the relationship between predicate circumscription and autoepistemic logic. The semantics of the second-order circumscription schema Circum$(A;P;Z)$, with every predicate other than P in the tuple of varying predicates Z, is given by the P-minimal models of A. Thus, the first-order consequences of a predicate circumscription of this form are exactly the sentences common to all extensions of a corresponding MIN$^=$ theory $\mathsf{M}(W;P;=)$, holding equality fixed. This result could be extended to the parallel circumscription of a tuple of predicates in an obvious way.

From the results of [de Kleer and Konolige, 1989], we know that fixed predicates are inessential, and that any circumscription involving fixed predicates can always be reduced to one without. For example, the circumscription Circum$(A;P;Z)$ with $Q \notin Z$ is equivalent to Circum$(A \wedge Q'=\neg Q;P,Q,Q';Z)$, where Q' is a new predicate constant. The corresponding construct for MIN$^=$ theories is to fix Q using $\mathsf{M}(W;P;=,Q)$ (note that there is no need to introduce a new predicate constant for $\neg Q$). Hence the first-order consequences of parallel predicate circumscription with fixed predicates are given by the corresponding MIN$^=$ theory.

On the other hand, there is no way to translate in general from autoepistemic theories to circumscription. The basic problem (noted in [Imielinski, 1987]) is that circumscription cannot distinguish between the epistemic idea of knowing Px, and the simple truth of Px. Thus, even without the complications of quantifying-in, there is no adequate circumscriptive translation for sentences such as $LPa \wedge \neg L\neg Qa \supset Qa$. Further, even in the restricted MIN$^=$ theories which have a natural correspondence to circumscription, AE logic is finer-grained than circumscription. Each extension of a MIN$^=$ theory yields a set of sentences true in an equivalence class of some minimal model of A, not all of them.

In the next section we use Theorem 3.5 to explore some issues of reasoning about equality in circumscription.

4 Reasoning about Equality

In defining P-minimal interpretations, we have specified that two interpretations must have the same domain and denotation function in order to be comparable. This corresponds to predicate circumscription with a fixed interpretation of terms.

Example 4.1 Consider a simple abnormality theory (see [McCarthy, 1986]), with $W = \{\forall x.Px \wedge \neg ab(x) \supset Qx,\ Pa,\ \neg Qb\}$ (this is a variation of an example in [Perlis, 1986]). We would expect Qa to be a consequence of Circum$(W;ab;Q)$, but it is not. The reason is that there are ab-minimal models of W in which b and a refer to the same individual, and $\neg Qa$ is true.

This example is typical of the way in which circumscription handles equality: it does not allow any new conclusions about equality, because the denotations of terms are fixed across comparable models. However, it is also possible to compare models with different denotation functions; the corresponding predicate circumscription allows functions to vary, as well as predicates (see [Lifschitz, 1984]).

EXAMPLE 4.2 Consider a simple abnormality theory with $W = \{\forall x.Px \wedge \neg ab(x) \supset Qx,\ ab(a),\ ab(b)\}$. In this case, $a{=}b$ is a consequence of $\mathrm{Circum}(W; ab; Q, a, b)$. The reason is that, if we allow interpretations with different denotations for a and b to be comparable, the interpretations in which a and b are identical are obviously minimal in ab.

From the above example, it seems that allowing terms to vary leads to the danger of unexpected identification of terms, at least if we do not have axioms that explicitly say that differing terms refer to different individuals We would like to treat equality among terms somewhat in between the two extremes of fixed and varying denotations: to remain agnostic about the equivalence of terms, but still be able to draw basic default conclusions.

The $\mathrm{MIN}^{=}$ theories, because of their relation to /■'-minimal models, always leave the denotations of terms fixed, and so fall prey to the same problems with equality as circumscription with fixed terms. However, we can relax the restriction on denotations by using M1N theories, without the sentences fixing equality.

EXAMPLE 4.3 Redoing the previous examples, let $W = \{\forall x.Px \wedge \neg ab(x) \supset Qx,\ Pa,\ \neg Qb\}$. There is one extension of $\mathrm{M}(W; ab)$, whose kernel is $\mathrm{Cn}(W,\ Ax.Px \supset x{=}a,\ a{\neq}b\}$. Here the lack of equality fixation leads to the conclusion that a and b are different individuals.

For the abnormality theory $W = \{\forall x.Px \wedge \neg ab(x) \supset Qx,\ ab(a),\ ab(b)\}$, on the other hand, the extensions of $\mathrm{M}(W; ab; =)$ and $\mathrm{M}(W; ab)$ are the same: getting rid of the equality-fixing sentences does not lead to the identification of a and b.

MIN theories without equality fixation are thus intermediate between a fixed and varying interpretation of equality, and seem to be the right level of variation for commonsense reasoning in abnormality theories.

5 Conclusion

We have extended the language and semantics of autoepistemic logic in a natural way to the case of quantified-in variables. By looking at a class of A10 theories, the $\mathrm{MIN}^{=}$ theories, we showed that all of the first-order consequences of predicate circumscription could be expressed in a simple way in autoepistemic logic. This is the first result on the relationship of these two logics for the case of nonfmite domains. The results have been used to shed some light on the treatment of equality in commonsense reasoning.

References

[Etherington and Reiter, 1984] R. E. Mercer D. W. Etherington and R. Reiter. On the adequacy of predicate circumscription for closed-world reason ing. In *AAA! Workshop on Non-Mono!ante Reasoning.* American Association for Artificial Intelligence, 1984.

de Kleer and Konolige, 1989] Johan de Kleer and Kurt Konolige. Eliminating the fixed predicates from a circumscription. *Artificial Intelligence,* page to appear, 1989.

[Etherington, 1986] David William Etherington. *Reasoning with Incomplete Information: Investigations of Non-Monotonia Reasoning.* PhD thesis, University of British Columbia, Vancouver, British Columbia, 1986.

[Imielinski, 1987] T. Imielinski. Results on translating defaults to circumscription. *Artificial Intelligence,* 32(1):131-146, 1987.

[Konolige, 1984] Kurt Konolige. *A Deduction Model of Belief and its Logics.* PhD thesis, Stanford University, 1984.

[Konolige, 1987] Kurt Konolige. On the relation between default logic and autoepistemic theories *Artificial Intelligence,* 35(3):343-382, 1987.

[Levesque, 1982] Hector J. Levesque. A formal treatment of incomplete knowledge bases. Technical Re port 614, Fairchild Artificial Intelligence Laboratory, Palo Alto, California, 1982.

[Levesque, 1987] Hector J. Levesque. All 1 know; an abridged report. In *Proceedings of the American Association of Artificial Intelligence.* Seattle, Washington, 1987.

[Lifschitz, 1984] Vladimir Lifschitz. Some results on circumscription. In *AAAI Workshop on Non-Monotonic Reasoning,* 1984.

[Lifschitz, 1985] Vladimir Lifschitz. Computing circumscription. In *Proceedings of the International Joint Conference on Artificial Intelligence,* pages 121 127, Los Angeles, California, 1985.

[McCarthy, 1986] John McCarthy. Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence,* 28, 1986.

[Moore, 1985] Robert C. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence,* 25(1), 1985.

[Perlis, 1986] Donald Perlis. On the consistency of commonsense reasoning. *Computational Intelligence,* 2, 1986.

[Reiter, 1980] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence,* 13(1-2), 1980.

[Shoham, 1987] Yoav Shoh am. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence.* MIT Press, Cambridge, Massachusetss, 1987.