SRI International

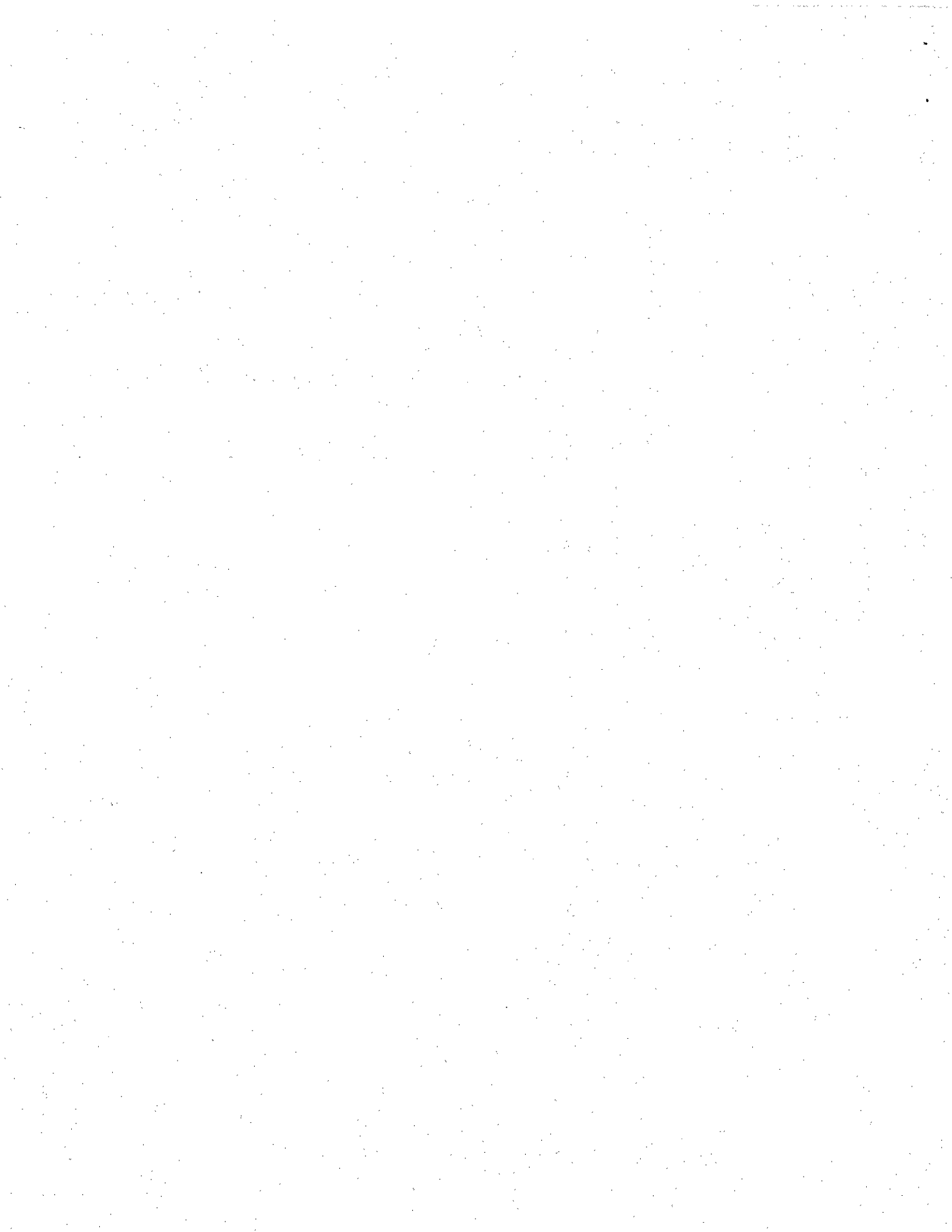# QUANTIFICATION IN AUTOEPISTEMIC LOGIC

Prepared by:

Kurt Konolige
Senior Computer Scientist

Artificial Intelligence Center *and*
Center for the Study of Language and Information

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

Abstract

Quantification in modal logic is interesting from a technical and philosophical standpoint. Here we look at quantification in autoepistemic logic, which is a modal logic of self-knowledge. We propose several different semantics, all based on the idea that having beliefs about an individual amounts to having a belief using a certain type of name for the individual.

Contents

1 Introduction

Autoepistemic (AE) logic [21] is a logic of self-belief, or, more precisely, it is a logic of an agent's knowledge of his own beliefs. Typically, relationships between beliefs and the world are expressed within the logic. For example, to state the proposition "if my brother is rich, then I know it," one uses:

(1) $P(brother) \supset LP(brother)$.

where $P(x)$ is the proposition that x is rich, and L is a modal operator expressing self-knowledge.

The statement (1) is propositional. Introducing full quantification into the language of AE logic presents both a challenge and an opportunity. The challenge is to find a natural semantics for expressions that have quantifying into a modal context: variables that are bound outside a modal operator, but used inside it. The opportunity is that the expressive power of the logic is expanded, and it is possible to use statements of self-knowledge for many representation problems in AI: domain closure, the unique names hypothesis, minimizing class membership, and reasoning about the equality of terms.

The semantics of AE logic is based on a *belief set*, a set of sentences Γ that are taken to be the agent's beliefs. The expression LP has a simple meaning with respect to this set: P is a member of the set. Think of the belief set as a knowledge base containing an agent's information about the world as a set of sentences; among these sentences can be ones, like LP, that refer to the knowledge base itself.

There are many types of knowledge an agent may have about his own beliefs. Some of this knowledge is about the identity of individuals that have certain properties. For example, I may know that someone is rich, without knowing who that someone is, in which case my beliefs would contain something like $\exists x.Px$, and my knowledge of my beliefs would include

(2) $L\exists x.Px$.

Now suppose I have a stronger belief, that there is a particular person who is rich. To express this fact, my knowledge base might contain the expression Pa for some term a that identifies the individual in question for me. In turn, my knowledge of my beliefs would say that I know that I have a belief that a particular individual is rich, and this could be expressed by

(3) $\exists x.LPx$.

The distinction between two statements (2) and (3) is well-known in standard epistemic logics with possible-worlds models. In these logics, an agent's beliefs are represented as a set of possible worlds, intuitively the possible ways the world could be if

the agent's beliefs were true. The sentence (2) is true if in each such possible world, there is some individual with property P; since it could be a different individual in each possible world, the agent doesn't know a particular individual for which P holds. In contrast, (3) is true if in each possible world the same individual has property P, so the agent knows who has the property.

Our treatment of quantification differs from this norm in two respects. First, as noted, we assume an explicitly intensional interpretation of free variables in modal contexts, so that an agent must have a certain kind of name for an individual in order to assert (3). Second, because the logic is self-referential, statements about beliefs can have the effect of defaults. For example, the expression

(4) $\forall x. P(x) \supset LP(x)$

says that if any individual has the property P, the agent will know it. Interpreted as one of the agent's beliefs, (4) says that the agent believes he has full knowledge of the members of the class P. So if the only member of P that the agent has beliefs about is a, the agent will come to believe

(5) $\forall x. P(x) \supset x = a$,

i.e., that there are no other members of P. We call defaults of the form (4) MIN theories, and we will show that they have a close relation to circumscription over countable models [18, 26].

Another aspect of our treatment of quantifying-in is that we can reason about the equality of terms when using defaults. Our semantics recognizes that distinct terms will often denote distinct individuals, but can also draw conclusions about the equality of terms when defaults are present. This is a particularly difficult area for other nonmonotonic logics.

In Section 2 we recall the basic definitions of AE logic without quantifying-in. Some properties of equality in the logic are discussed in Section 3. The language and semantics of quantifying-in are developed in Section 4, then used to discuss the issue of knowing who an individual is in Section 5. We next concentrate on technical results for MIN theories (Section 6) and apply them to various representational problems (Section 7). Section 8 discusses a further refinement of MIN theories in which predicates can be fixed during minimization, and in Section 9 we give the connection between MIN theories with fixed equality and circumscription over countable models. Related research is examined in Section 10.

6

2 Autoepistemic Logic

We first consider autoepistemic logic with equality but without free variables in modal contexts. The treatment generally follows and extends [8].

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 $P(t_1, \cdots, t_n)$. 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:

(6) If ϕ is a sentence of $\mathcal{L}$, then so is $L\phi$.

An expression $L\phi$ is a *modal atom*. Sentences and atoms of $\mathcal{L}_0$ are called *nonmodal*. The argument of a modal operator cannot contain free variables, so there is no quantifying into the scope of a modal atom, e.g., $\exists x LPx$ is a not a sentence. Often we will use a subscript "0" to indicate a subset of nonmodal sentences, e.g., $\Gamma_0 = \Gamma \cap \mathcal{L}_0$. $\mathcal{L}_1$ is the subset of $\mathcal{L}$ containing all sentences with no nested modal operators.

How are atoms of the form $L\phi$ to be interpreted? Since the operator L refers to the beliefs of an agent, it is natural to use a set of sentences, the *belief set*, as an element of the semantics. Let $\Gamma \subseteq \mathcal{L}$ be a set of sentences; we say that a first-order interpretation w *respects* Γ if for all sentences $L\phi$[1]:

(7) $w \models L\phi$ iff $\phi \in \Gamma$,

that is, w treats the sentence $L\phi$ as meaning that ϕ is in the belief set Γ. Logical implication is defined, relative to a particular belief set Γ, as

(8) $A \models_\Gamma \phi$ iff $\forall w.$ such that w respects Γ and is a model
      of A, $w \models \phi$.

Because the intended meaning of $L\phi$ is not just belief, but *self*-belief, the correct belief set to use in defining an agent's beliefs is the very belief set being defined, so that the definition of the belief set becomes circular. Formally, we use a fixed-point equation:

(9) $T = \{\phi \mid A \models_T \phi\}$.

Here the index T is the belief set of the agent, so that the interpretation of $L\phi$ is correctly taken to be self-belief. The fixed-point equation states that an ideal introspective agent should have a belief set in which all beliefs follow logically from

---

[1] If w respects Γ and is a model of Γ then it is an *autoepistemic interpretation* of Γ [21].

the base set A, under the assumption that atoms $L\phi$ in A are treated as referring to the belief set itself. A set of sentences which obeys (9) is called an *extension* of the base set A. Extensions are completely determined by their first-order part, called the *kernel* of the extension.

We use $Cn(X)$ to mean the first-order consequences of a first-order set of sentences X. A sentence $\phi \in \mathcal{L}$ is called an AE-consequence of the theory A if it is contained in all extensions of A.

3 Equality

The equality predicate is treated in the standard first-order way. Let the interpretation w consist of $\langle U, v, R \rangle$, where U is a universe of individuals, v is a mapping from terms to individuals and predicate names to relations, and R is a set of relations. Then:

$$(10) \quad w \models t_1 = t_2 \quad \text{iff} \quad v(t_1) = v(t_2).$$

In this semantics, the following axioms and schemata all hold:

$$(11) \quad \begin{aligned} &\forall x.\, x = x \\ &\forall xy.\, x = y \supset y = x \\ &\forall xyz.\, x = y \wedge y = z \supset x = z \\ &\forall xy.\, x = y \supset [\phi(x) \equiv \phi(y)] \end{aligned}$$

Here $\phi(x)$ is any first-order formula with the free variable x.

In AE logic, since extensions are closed under first-order consequence these expressions are also in any extension. In addition, because extensions are closed under instantiation, all substitution instances of these expressions also hold.

The equality schema does not have much impact on modal atoms, because they do not contain free variables. For example, $\forall xy.\, x = y \supset [LP(x) \equiv LP(y)]$ is not an expression of $\mathcal{L}$, and so is not part of the schema. But the substitution instances $a = b \supset [LP(a) \equiv LP(b)]$ are in $\mathcal{L}$, and in general they are not AE consequences of a theory.

EXAMPLE 3.1 Let $\mathrm{Cn}\{Pa, Qb\}$ be the kernel of an AE extension T. T contains LPa, $\neg LPb$, and $\neg L(a = b)$. Let w be a first-order interpretation in which Pa and Qb hold, and for which $a = b$. Since w respects T and $w \not\models a = b \supset [LPa \equiv LPb]$, $(a = b \supset [LPa \equiv LPb])$ is not in T.

This example highlights the interaction between names and equality in AE logic. Take P to be the property of being rich, a to be a descriptive term for the mayor, and b for the police chief. An agent has proof that the mayor is rich (LPa) and no evidence that the police chief is $(\neg LPb)$. These are statements about the *intension* of the terms a and b, that is, the agent believes that whatever individual goes under the description of "mayor," whoever he is, is rich; and the police chief is not. It says nothing about whether the agent knows who is denoted by these terms: for all he knows, the mayor and the police chief could be the same individual (assume that this is a small town on a tight budget). So if the agent does not know that the two names refer to the same individual, he cannot conclude that knowing one is rich implies that he knows or does not know the other is rich. Contrast this with $(a = b \supset [Pa \equiv Pb])$, which

is true in all extensions. Here we are simply claiming that if a and b are the same individual, he must be rich or poor; the names a and b are interpreted extensionally in this context, that is, they simply denote the same individual.

The equality substitution schema is true in a modalized form, given by the next proposition.

PROPOSITION 3.1 *In any AE extension, all instances of $L(a = b) \supset [L\phi(a) \equiv L\phi(b)]$ are present, where $\phi(x)$ is a nonmodal formula.*

> *Proof.* Either $L(a = b)$ or $\neg L(a = b)$ is in every extension. If the latter, then the schema instance is also. If the former, then $a = b$ is in the extension, and since it is closed under first-order consequence, so is $\phi(a) \equiv \phi(b)$. Since extensions are closed under $S5$-consequence (see [8], Proposition 3.2), $L\phi(a) \equiv L\phi(b)$ is in the extension.

From this it is possible to derive certain nonbeliefs about equality, e.g., $(LPa \wedge \neg LPb) \supset \neg L(a = b)$.

We can exploit the self-referential nature of AE logic to draw new conclusions about equality by explicitly asserting some defaults. For example, we may want to infer that two descriptions for which we have differing beliefs actually refer to different individuals. Then we would add the schema:

(12) $L\phi(a) \wedge \neg L\phi(b) \supset a \neq b$.

In the example above, this would lead to the default conclusion that the mayor and the police chief are different, because we know one of them is rich, but not the other.

The unique names assumption is that every term refers to a different individual. This property arises naturally in the context of deductive databases [24]. We can use equality schemata to state a flexible unique names assumption. For any two distinct terms a and b, assert the schema

(13) $\neg L(a = b) \supset a \neq b$.

Any names that can consistently refer to different individuals are forced to do so. If there are no positive assertions of equality predicates in the premises A, then any extension of A will have the unique names property.

Domain closure is the assumption that the only individuals that exist are those referred to by the terms in a theory. This assumption is also important in the theory of deductive databases [24]. AE logic, even with equality, is not strong enough to assert domain closure as a default when there are infinitely many terms.

A stronger concept than domain closure is domain circumscription: only those individuals necessary to satisfy a give theory T are assumed to exist [18]. Unlike domain closure, domain circumscription tries to make the universe of individuals as small as possible. In AE logic, we could try:

10

(14) $\neg L(a \neq b) \supset a = b$,

but this just says that as many names as possible refer to the same individuals. There can also be unnamed individuals in the domain, and no ground equality statements will restrict the number of such individuals.

4 Quantifying-in

The language $\mathcal{L}$ does not allow the presence of free variables in a modal context (called familiarly "quantifying-in," e.g., $\exists x.LPx$). In this section we extend the logic to deal with such expressions, and look at the consequences for statements of equality and identity.

Let $q\mathcal{L}$ be defined as the sentences of $\mathcal{L}_0$, together with the formation rule

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

As usual, the sentences of $q\mathcal{L}$ are the formulas with no free variables.

It is not obvious how to extend the semantics of the logic to deal with quantified-in expressions. Recall that the belief set Γ is a set of sentences that form the beliefs of an agent. To interpret $L\phi$, we simply ask whether the expression ϕ is in Γ; if so, every interpretation w that respects Γ must satisfy $L\phi$. But with the quantified-in language, we must also be able to interpret $L\phi(x)$, where $\phi(x)$ 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 referring to individuals in the domain. We will give three basic methods for doing this, and show how these methods interact with different assumptions about the nature of terms in the language.

The basis for all of these schemes is a set of terms of $q\mathcal{L}$ that are singled out as appropriate "names" for individuals. We call this set $\mathcal{N}$. $\mathcal{N}$ can be any subset of the terms of $q\mathcal{L}$, including all terms, in which case the language is said to have a *full name set*. To distinguish elements of $\mathcal{N}$ from terms not in $\mathcal{N}$, we call them *names*.

Obviously, there can be individuals in a domain that do not have a name, even if the language has a full name set. If we restrict the semantics of $q\mathcal{L}$ to just those domains in which every individual has a name, then we make the assumption of *parameter interpretations*. Herbrand interpretations are one type of parameter interpretation, in which every term denotes itself. Parameter interpretations are more general than Herbrand interpretations, since every Herbrand term is unequal to all others. In parameter interpretations, two terms can refer to the same individual.

4.1 Any-Name Semantics

The simplest scheme for reference to domain elements is to use the denotation map v already present in the first-order interpretation w, and to let any name $a \in \mathcal{N}$ such that $v(a) = x$ suffice to pick out the individual x in modal contexts. We extend the rule for interpretations respecting a belief set (7) to the case of formulas with free variables.

Any-name semantics:

(16)
w respects Γ if for all formulas $L\phi(x)$ and substitutions x/k,

$$w, x/k \models L\phi(x) \quad \text{iff} \quad \text{for some name } t \in \mathcal{N} \text{ such} \\ \text{that } v(t) = k, \ \phi(t) \in \Gamma.$$

Here w is an interpretation and x/k means that the variable x is to refer to the individual k; so we say that $\phi(x)$ is believed if k, the individual x refers to, has a name t such $\phi(t)$ is believed. The normal truth-recursion rules for quantifiers hold, e.g.,

(17)
$$w \models \forall x.\phi(x) \quad \text{iff} \quad \text{for all individuals } k \in U, \\ w, x/k \models \phi(x).$$

The definition of extension for $q\mathcal{L}$ does not need any changes, but because the concept of respecting a belief set is expanded, the resulting extensions are not the same for the two languages. One of the most useful results in AE logic over $\mathcal{L}$ is that any set of nonmodal sentences has a unique AE extension. This result carries over to the quantified-in language.

PROPOSITION 4.1 *If A is a set of nonmodal sentences from $q\mathcal{L}$, it has exactly one AE extension T. T_0 is the first-order closure of A.*

Proof. We define the sets $S(n)$ in the following iterative fashion:

$$S(0) = \{\phi \in \mathcal{L}_0 \mid A \models \phi\} \\ S(n) = \{\phi \in \mathcal{L}_n \mid A \models_{S(n-1)} \phi\}$$

Let T_n be the set of sentences of T from $\mathcal{L}_n$, and let S be the infinite union of all $S(i)$. We can show that if T is an AE extension of A, $T_n = S(n)$; there is thus at most one AE extension T, with $T_0 = S(0)$, the first-order closure of A. We prove existence by showing that S is always an AE extension of A. See the proof of Proposition 2.1 [8, pp. 371–372] for the details, since it applies here without change.

In addition, extensions in $q\mathcal{L}$ are stable sets [27].

PROPOSITION 4.2 *Any extension T satisfies the following properties*

1. *T is closed under first-order consequence.*

2. *If $\phi \in T$, then $L\phi \in T$.*

3. *If $\phi \notin T$, then $\neg L\phi \in T$.*

14

Proof. The first item follows from the fact that extensions are defined using consequence under first-order valuations. Note that $L\phi(x)$, with free variable x, is treated from a first-order viewpoint as a unary predicate, dissimilar to any other predicate, including $L\phi(a)$ for any term a.

The second two items follow from the first-order valuations respecting the belief set. In fact, only the nonquantified form of respect (7) is needed here, so any semantic account of quantifying-in will yield extensions that are stable sets.

Finally, if the sentences of A do not have any nested modal operators, than the kernel of the extension satisfies the following reduced fixed point equation.

PROPOSITION 4.3   *If A is a set of sentences from $\mathcal{L}_1$, then*

$$S = \{\phi \in \mathcal{L}_0 \mid A \models_S \phi\}$$

if and only if S is the kernel of an extension of A.

Proof. We define the sets $S(n)$ in the following iterative fashion:

$$
\begin{aligned}
S(1) &= S \quad \text{as defined above} \\
S(n) &= \{\phi \in \mathcal{L}_n \mid A \models_{S(n-1)} \phi\}
\end{aligned}
$$

The proof proceeds along the same lines as that for Proposition 4.1. The only change is that we start the recursive definition with $S(1)$ instead of $S(0)$.

We can use these results to explore some of the facts about quantification and equality in modal contexts. A given individual x may have none or many names in a model, a circumstance which leads to some interesting behavior in extensions.

EXAMPLE 4.1   Let $A = \{Pa\}$, and $\mathcal{N} = \{a, b\}$. By Proposition 4.1 there is a single extension T of A, with $T_0 = \text{Cn}(Pa)$. Therefore, we know that LPa and $\neg LPb$ are in T. By the valuation rule for modal atoms (16), $\exists x.LPx$ will be true in any interpretation respecting T_0, 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 model of A respecting T_0, and hence in T.

Another sentence contained in T is $\forall x. x = a \supset LPx$. To see why this is so, let x be an individual with $x = v(a)$. Since LPa is in T, LPx is true in all interpretations which respect T, and so the whole sentence is true in all such interpretations. On the other hand, consider a similar sentence $\forall x.x = b \supset \neg LPx$. It might be suspected that this sentence is a member of T, but it is

not. For although x is the denotation of b, it may also be the denotation of a in some first-order interpretation, and for this interpretation, LPx will be true. In the other direction, the sentence $\forall x.\, LPx \supset x = a$ is in T, since in every model respecting T, only the individual named by a will make LPx true. On the other hand, $\forall x.\, Px \supset x = a$ is not in T: there is a model w of A for which Px is true of some individual $x \neq a$.

As in $\mathcal{L}$, in $q\mathcal{L}$ knowledge of properties of individuals hinges on having a name for that individual. The expression LPx when x is a quantified-in variable says that the agent believes Pc to be true for some intensional concept c whose denotation is x. Using the above example, and letting a stand for the mayor, b for the chief of police, and P for the property of being rich, we can interpret the formal results as follows. The agent has a belief that the individual described as "the mayor" is rich, and so affirms that, if x is that individual, LPx is a belief, because under at least one of the names of x (namely, a), LPa is true. On the other hand, the agent cannot say that he disbelieves that the individual y who is described as "the police chief" is rich because this individual may have another name for which LPy is true, that is, y may also be the mayor.

These observations hold for arbitrary extensions containing LPa and $\neg Pb$, and we can make this precise by introducing the meta-logical sentence $\phi \mapsto \psi$. This sentence is true just in case whenever ϕ is in some extension (over $q\mathcal{L}$), so is ψ. Similarly, $\phi \not\mapsto \psi$ means that there exists an extension containing ϕ that does not contain ψ. Note that $A \mapsto \phi$ is stronger than saying that ϕ is an AE-consequence of A, since A does not have to be the premise set of an extension, only included within it.

PROPOSITION 4.4   *Let* $a, b \in \mathcal{N}$, *and let* $\phi(x)$ *be any first-order formula with free variable* x. *The following statements are true of any-name semantics.*

$$L\phi(a) \;\mapsto\; \exists x. L\phi(x)$$

$$L\phi(a) \quad \mapsto\; \forall x. x = a \supset L\phi(x)$$
$$\neg L\phi(a) \;\not\mapsto\; \forall x. x = a \supset \neg L\phi(x)$$
$$\forall x.\, x = a \supset L\phi(x) \quad \mapsto\; L\phi(a)$$
$$\forall x.\, x = a \supset \neg L\phi(x) \;\mapsto\; \neg L\phi(a)$$
$$\mapsto\; \forall xy.\, x = y \supset (L\phi(x) \equiv L\phi(y))$$
$$\not\mapsto\; a = b \supset (L\phi(a) \equiv L\phi(b))$$

Proof. The first three statements can be proven from the example by a simple generalization. For the fourth, consider an interpretation in which $\forall x.\, x = a \supset L\phi(x)$ is true. For some individual k such that $v(a) = k$, there is a name b for k such that $L\phi(b)$ is in the extension. Since extensions

16

are stable sets, $\phi(b)$ is in the extension, and since $v(a) = v(b)$, so is $\phi(a)$, and hence $L\phi(a)$. For the fifth statement, in any interpretation there is some individual k for which $v(a) = k$, and since $\neg L\phi(x)$ is true for this individual, $\phi(a)$ cannot be in the extension, and so $\neg L\phi(a)$ is.

The sixth statement follows directly from the rule of interpretation for modal atoms with free variables, since the variables refer to the same individuals. Finally, the example furnishes a counterexample to the inclusion of the last sentence in all extensions.

Note that the universal statement of substitution of equal individuals in a modal context is in every extension, but its substitution instances over names are not. The rule of instantiation of universal quantifiers, which holds for first-order logic, is not valid in quantified AE logic with any-name semantics. This can be understood by considering the subexpression $L\phi(x)$. If this formula is asserted positively in some sentence, and we substitute a name a, we are not sure that a is the name that led to the evaluation of $L\phi(x)$ as true, since there could be many such names. On the other hand, if $L\phi(x)$ occurs negatively, any name can be substituted, since any such name should make $L\phi(x)$ false. The moral: substitution of a particular name for x into positive modal contexts is not a valid operation within extensions, given the any-name semantics. Substitution is valid for negative modal contexts, since $\neg L\phi(x)$ is a strong statement in any-name semantics: there is no name a such that $L\phi(a)$ is believed.

Finally, we note that the Barcan formula $\forall x L\phi(x) \supset L\forall x \phi(x)$ is true in every extension, while the converse $L\forall x \phi(x) \supset \forall x L\phi(x)$ may be false. The reason for the latter is that even though every individual x has the property ϕ, some individuals may not be given a name in an interpretation, and so $L\phi(x)$ will be false. By the properties of stable sets, we can reduce the subexpression $L\forall x\phi(x)$ to $\forall x\phi(x)$ in these sentences.

PROPOSITION 4.5 *For any-name semantics,*

$$\forall x L\phi(x) \mapsto \forall x \phi(x)$$
$$\forall x \phi(x) \not\mapsto \forall x L\phi(x).$$

If only parameter models are considered, then both the Barcan formula and its converse are in every extension.

4.2   All-Name Semantics

An alternative scheme for reference is to assume that $L\phi(x)$ is true in case the expressions formed by substitution of *all* names for an individual x in the denotation

map are part of the belief set. The rule for interpretations respecting a belief set (16) is modified:

All-name semantics:

(18) w respects Γ if for all formulas $L\phi(x)$ and substitutions x/k,

$$w, x/k \models L\phi(x) \quad \text{iff} \quad \text{for every name } t \in \mathcal{N} \text{ such} $$
$$\text{that } v(t) = k, \, \phi(t) \in \Gamma.$$

Proposition 4.1 remains true, and all extensions are still stable sets. However, some of the results for equality and substitution are reversed from any-name semantics.

PROPOSITION 4.6   *Let $a, b \in \mathcal{N}$, with $\phi(x)$ a nonmodal formula. The following statements are true of all-name semantics.*

$$L\phi(a) \quad \nvdash \exists x. L\phi(x)$$

$$L\phi(a) \quad \nvdash \forall x.\, x = a \supset L\phi(x)$$
$$\neg L\phi(a) \quad \vdash \forall x.\, x = a \supset \neg L\phi(x)$$
$$\forall x.\, x = a \supset L\phi(x) \quad \vdash L\phi(a)$$
$$\forall x.\, x = a \supset \neg L\phi(x) \quad \vdash \neg L\phi(a)$$
$$\vdash \forall xy.\, x = y \supset (L\phi(x) \equiv L\phi(y))$$
$$\nvdash a = b \supset (L\phi(a) \equiv L\phi(b))$$

Proof.   The truth of these statements can be verified using the same techniques as for Proposition 4.4.

Again, the rule of instantiation is not valid in quantified AE logic, but with all-name semantics the reasons are opposite from any-name semantics. Here the positive statement $LP(x)$ is strong, since it says that for any name a, $L\phi(a)$ must be true. So substitution of a particular name for x into positive modal contexts is valid within extensions for all-name semantics. Substitution is not valid for negative modal contexts, since $\neg L\phi(x)$ just means there is some name for x that the agent does not connect with property P, and for any particular name a, $L\phi(a)$ may be a belief.

The results on the Barcan formula (Proposition 4.5) are the same for all-name semantics.

4.3   Unique-Name Semantics

A specialization of the any-name and all-name semantics is to let each member of $\mathcal{N}$ stand for a unique individual. Unique names of this sort are called "standard names," and for any two standard names we have:

(19) $n_i \neq n_j$ iff $i \neq j$.

The rule for interpretations respecting a belief set is:

Unique-name semantics:

(20) w respects Γ if every name in $\mathcal{N}$ refers to a unique individual in U, and for all formulas $L\phi(x)$ and substitutions x/k,
$w, x/k \models L\phi(x)$ iff for any name $t \in \mathcal{N}$ such that $v(t) = k$, $\phi(t) \in \Gamma$.

Unique-name semantics merges any-name and all-name semantics, since there is exactly one name for each individual. Proposition 4.1 remains true, and all extensions are still stable sets. These are the results for equality and substitution.

PROPOSITION 4.7 *Let $n, n_i, n_j \in \mathcal{N}$, with $\phi(x)$ a nonmodal formula. The following statements are true of unique-name semantics.*

$$LPn \; \mapsto \exists x. L\phi(x)$$

$$LPn \quad \mapsto \forall x. x = n \supset L\phi(x)$$
$$\neg LPn \; \mapsto \forall x. x = n \supset \neg L\phi(x)$$
$$\forall x. x = n \supset L\phi(x) \quad \mapsto LPn$$
$$\forall x. x = n \supset \neg L\phi(x) \; \mapsto \neg LPn$$
$$\mapsto \forall xy. x = y \supset (L\phi(x) \equiv L\phi(y))$$
$$\mapsto n_i = n_j \supset (L\phi(n_i) \equiv L\phi(n_j))$$

Proof. The truth of these statements can be verified using the same techniques as for Proposition 4.4.

The main change is that the rule of instantiation is valid for unique-name semantics, as long as only standard names are substituted into the context of modal atoms.

The results on the Barcan formula (Proposition 4.5) are the same for unique-name semantics.

4.4 A Note on the Modal Predicate Calculus

We have chosen a semantic approach to analyzing quantifying-in because it clearly shows the relation between assumptions about naming individuals and self-beliefs. It is known from the original definition of AE logic [21] that a proof-theoretic version of the fixed-point equation (9) exists; this fixed-point can be expressed in terms of deduction in the modal calculus $K45$, or weak $S5$.

PROPOSITION ([8], 3.6) *A set T is an AE extension of $A \subseteq \mathcal{L}$ if and only if it satisfies the equation*

$$T = \{\phi \in \mathcal{L} \mid A \cup LT_0 \cup \neg L\overline{T}_0 \vdash_{K45} \phi\} \,,$$

where $LT_0 = \{L\phi \mid \phi \in T \cap \mathcal{L}_0\}$ and $\neg L\overline{T}_0 = \{\neg L\phi \mid \phi \notin T \text{ and } \phi \in \mathcal{L}_0\}$. This definition can be extended to $q\mathcal{L}$ by using the first-order version of the calculus $K45$. However, the intuitions we have developed in this section are violated by the proof-theoreitc extensions of some simple AE theories. For example, let $A = \{\exists x.LPx\}$. This theory should have no extension, since there is no individual $c \in \mathcal{N}$ such that Pc is derivable from A. Yet there is a solution to the fixed-point equation. Let T_0 be Cn(); then the $K45$-consequences of $A \cup LT_0 \cup \neg L\overline{T}_0$ are T_0, giving a fixed-point. T_0 does not contain $\exists x.Px$, and so T contains $\neg L\exists x.Px$. The agent does not believe that there is any individual with property P, yet he believes that he believes it.

5 Knowing Who

In an ordinary epistemic logic, the expression $\exists x.K\phi(x)$ is contrasted with $K\exists x.\phi(x)$. The latter is meant to express that an agent knows that something has the property ϕ, e.g, the agent knows that somebody is rich, without necessarily being able to pick out that individual. On the other hand, the quantified-in version expresses the fact that the agent has a particular individual in mind who is rich; we, the observers, represent this fact without saying who the individual is. The formal technique was originally suggested by Hintikka [6], and in the AI literature there are many examples: Moore's theory of knowledge and action [20], and Levesque's approach to self-knowledge in knowledge bases [11] are two well-known ones. Levesque's work is especially appropriate here: he points out that the expressions $\exists x.L\phi(x)$ and $L\exists x.\phi(x)$ in a knowledge base differentiate the type of information contained in it. In the first case, the knowledge base has information about a particular individual with property ϕ; in the second, it just has the information that some individual has property ϕ.

The contents of the knowledge base should reflect the different kinds of information present. Since Levesque takes knowledge bases to be described by collections of sentences, similar to the belief sets of AE logic, knowing who a particular individual is depends on having a certain type of name for the individual, often called a *standard name*. The knowledge base knows that an individual has the property ϕ just in case the expression $\phi(a)$ is in the KB, where a is the standard name for the individual. Suppose, for example, that we take an individual's proper name to be a standard name (this is only a good assumption for a small group of people, which is why social security numbers are useful). Then *Rich(John Doe)* in the knowledge base counts as knowing of the individual John Doe that he is rich, while *Rich(richest-man-in-the-world)* just means that the knowledge base knows someone is rich, without being able to identify that person.

Using standard names is one way in which an agent can identify individuals, but it is certainly not the only one. We have argued (in [7]) that an appropriate notion of knowing who an individual is is that the agent have a description sufficient to pick out the individual for a particular task in a particular context. For example, a robot trying to go through a doorway in front of it needs only to have an expression of the form $Closed(d1) \wedge Door(d1) \wedge Dist(d1, 3ft)$ in its knowledge base in order to take an appropriate action relative to the door. Having a standard name for the door would require too much: that the agent be able to distinguish this door from any other known door, for example.

Just what constitutes an adequate description is still a matter of research, both in the study of epistemology, where it has a long tradition, and in the AI and computer science community. Recent work of interest in the latter are the proposals of Lespérance [10] on self-naming and Grove and Halpern [5] on multi-agent naming and

reference. Grove and Halpern stress the importance of descriptive names (as opposed to standard names) as a way of identifying particular agents in context, and go on to give a theory of such descriptions, making names first-class objects in his modal language, quantifying over them and introducing predicates for describing different types of names.

We will not subscribe to a particular theory of description here, since we want to be more abstract and remain compatible with different theories. But general characteristics of a description theory can guide the choice of an appropriate semantics for $q\mathcal{L}$.

1. We make the assumption that it is the syntactic form of the name in an expression such as $P(a)$ that is important in determining whether it is an appropriate description or not. For example, 322-3646 is an appropriate name if the agent is trying to dial a phone number, while *phone-num(Bill)* is not, since the agent may not know the digits of Bill's phone number. The appropriateness of a name is task and context dependent.

2. Different appropriate names may refer to the same individual, without an agent knowing whether those individuals are the same or not. This is especially true if the names are used in different contexts. An agent might use the name "mayor" when talking to the phone operator and trying to get in touch with the mayor; he might use the term "police chief" if trying to get in touch with that individual. These are appropriate titular names, and the operator knows how to connect the agent with the right person; yet the agent might not know whether the mayor and the police chief are the same individual. As we have seen, standard names do not satisfy this condition, since the refer to unique individuals, and the agent knows they are unique.

3. There may be individuals for which the agent has no appropriate identifying names, that is, the names do not cover the set of individuals an agent can have beliefs about. Also, there may be terms for individuals that do not count as identifying names (e.g., Skolem constants).

How do the proposed semantics for quantified-in AE logic meet the above criteria? The first condition is the hardest one, since it really demands a theory of appropriate identifiers for individuals. Since we do not want to complicate $q\mathcal{L}$ by introducing names as objects of the domain and subscribing to a particular theory of description, the best that can be done is to use the facilities of the language specification to select a set of terms $\mathcal{N}$ that approximate the appropriate identifiers for a class of applications.

All of the semantics use the names in $\mathcal{N}$ as identifying names for individuals. If we do not make the parameter interpretation assumption (i.e., that all individuals are

referred to by names in $\mathcal{N}$), then the models of $q\mathcal{L}$ may have unnamed individuals. Also, if $\mathcal{N}$ is not coextant with the set of all terms, then some terms for individuals do not count as identifying those individuals, thus satisfying the third condition. There may be some cases in which we have a restricted domain and can name all individuals with $\mathcal{N}$, but in general we will have to deal with the problem of "open" domains.

The any-name and all-name semantics fulfill second condition, while the unique-name semantics does not. The unique-name semantics is essentially the same as using standard names. In special cases standard names are appropriate, e.g., it is possible to tell, for any two telephone numbers, whether those numbers are the same or not (actually, even this case is not strictly correct: there can be multiple numbers that ring at the same location, as in large customer service operations).

The most appropriate semantics, from a representational point of view, appears to be any-name semantics. The results of the rest of this paper are for this semantics.

6 MIN Theories

One of the uses of quantifying-in is to specify minimization over known instances of a predicate. We now examine a particular class of AE theories that can serve this purpose; they are called MIN theories. Any such theory has the form

$$(21) \quad A \bigcup_i \{\forall x(\neg LP_i x \supset \neg P_i x)\} \,,$$

where A is a set of first-order sentences and the P_i are a sequence of unary predicates.[2] We write $M(W; P_i)$ to indicate the MIN theory of W over the predicates P_i. The idea behind MIN theories is to select AE valuations in which every individual not known to have the property P_i does not have this property, i.e., to minimize the extension of each P_i.

Before describing the properties of MIN theories, we need the following result on a reduced fixed point form for their kernels. Define $\text{Atoms}(\mathcal{N}; P_i)$ to be the set of all ground atoms of the form $P_i(a)$, with $a \in \mathcal{N}$.

PROPOSITION 6.1 *S is the kernel of an extension of* $M(A; P_i)$ *if and only if*

$$S = \{\phi \in \mathcal{L}_0 \mid M(A; P_i) \models_{S \cap \text{Atoms}(\mathcal{N}; P_i)} \phi\}$$

Proof. By Proposition 4.3, S is a kernel of a MIN theory iff

$$S = \{\phi \in \mathcal{L}_0 \mid M(A; P_i) \models_S \phi\} \,.$$

Since the only modal atoms in $M(A; P_i)$ are of the form $LP_i x$, the only sentences of S relevant to the truth of these atoms are of the form $P_i(a)$ for $a \in \mathcal{N}$.

A set $S \cap \text{Atoms}(\mathcal{N}; P_i)$ satisfying the equation above is called an *atom base* of the theory $M(A; P)$.

EXAMPLE 6.1 Let $\mathcal{N} = \{a_1, a_2, \cdots\}$, and $A = S = \{Pa_1\}$. Every interpretation w respecting S which satisfies $M(A; P)$ makes Px true for $x = v(a_1)$, and false for every other x. Thus, if we define T by

$$T = \{\phi \in \mathcal{L}_0 \mid M(A; P) \models_S \phi\} \,,$$

---

[2] All results of this section can be readily extended to predicates of arity greater than one.

it is clear that $T = \mathrm{Cn}(\forall x. Px \equiv x = a_1)$. Since $S = T \cap \mathrm{Atoms}(\mathcal{N}; P)$, T is the kernel of an extension of $\mathrm{M}(A; P)$. There are no other kernels, since no other atoms Pa_i can be deduced from $\mathrm{M}(A, P)$.

Let $A = \{Pa_1 \vee Pa_2\}$, and as before let $S = \{Pa_1\}$. Again we can show that every valuation w respecting S and satisfying $\mathrm{M}(A; P)$ satisfies $T = \mathrm{Cn}(\forall x. Px \equiv x = a_1)$, and so T is the kernel of an AE extension of $\mathrm{M}(A; P)$. In this case the kernel is not unique; there is another one $\mathrm{Cn}(\forall x. Px \equiv x = a_2)$. The sentence $(\forall x. Px \equiv x = a_1) \vee (\forall x. Px \equiv x = a_2)$ is an AE-consequence of $\mathrm{M}(A; P)$.

Let $A = \{\exists x Px\}$, and let $S = \{Pa_1\}$. Again we can show that every valuation w respecting S and satisfying $\mathrm{M}(A; P)$ satisfies $T = \{\forall x. Px \equiv x = a_1\}$, and so T is the kernel of an extension. But the choice of the constant a_1 was arbitrary, and we can use any other constant in defining S and T. Hence there are an infinite number of extensions of $\mathrm{M}(A; P)$; the sentence $\exists! x. Px$ is an AE-consequence of $\mathrm{M}(A; P)$.

These examples are very suggestive of a correspondence between MIN theories and the minimal models of A. But we must define what we mean by minimal models, and there are several choices. First we assume that the extensions of all predicates other than the P_i can vary across compared models. In the next section we consider the case of *fixed* predicates. Second, we must choose whether to allow the universe and denotation function of comparable models to be different. As we will see in the section on equality reasoning (7.3), the minimization of MIN theories is somewhere between having a fixed and varying denotation function. Also, the extensions of $\mathrm{M}(A; P)$ characterize minimal models that have a *parameter cover* of the predicate P: in every such model, the individuals with the property P are all named in $\mathcal{N}$. This is what the expression $\forall x.\, Px \supset LPx$ means: if any individual has the property P, it is believed to have it, i.e., it has a name a and Pa is in the belief set.

PROPOSITION 6.2 *Let V be an atom base for $\mathrm{M}(A; P)$.[3] Every model w of $\mathrm{M}(A; P)$ respecting V has a parameter cover of P, and for every $a \in \mathcal{N}$,*

$$w \models Pa \quad \textit{iff} \quad Pa \in V.$$

Proof. w has a parameter cover for P because if $w, x/k \models Px$ for some individual k, then by $\forall x.\, Px \supset LPx$ and w respecting V there must be a predicate Pa of V such that $a \in \mathcal{N}$ and $v(a) = k$. Similarly, if $w \models Pa$ then Pa must be in V. Now suppose that $Pa \in V$ for some $a \in \mathcal{N}$. This means LPa is in the extension, and so is Pa since it is a stable set; hence $w \models Pa$.

---

[3]Note: we will often use a single minimized predicate P in propositions in the rest of this paper for clarity; the addition of multiple predicates causes no significant changes in the proofs.

The set of ground atoms Pa such that $a \in \mathcal{N}$ and $w \models Pa$ is called the *P-cover* of w. The P-cover is analogous to a Herbrand model for A, except that it is only for the positive part of the predicate P, and the parameters can refer to the same individual. In the same way that minimal Herbrand models are taken as a semantics for logic programs, we consider models of A that have a subset-minimal P-cover. These minimal P-covers determine the extensions of $\mathrm{M}(A; P)$.

PROPOSITION 6.3 *Let A be a set of first-order sentences. A set V is an atom base of $\mathrm{M}(A; P)$ if and only if it is a minimal P-cover for A.*

Proof. Let V be a minimal P-cover for A, and define S by

$$S = \{\phi \in \mathcal{L}_0 \mid \mathrm{M}(A; P) \models_V \phi\}.$$

We know that $S \cap \mathrm{Atoms}(\mathcal{N}; P) \subseteq V$, since by definition there is a model of A that makes V true, but no other ground P atoms. To show that S contains V, suppose to the contrary that for some $Pa \in V$, Pa is not in S. Then there is a model w that makes $V - \{Pa\}$ true, but no other P atoms, since $\forall x.Px \supset LPx$ forces all P atoms not in V to be false. But we assumed V was a minimal P-cover of A, which w contradicts.

For the converse, let S be an extension of $\mathrm{M}(A; P)$ with atom base $V = S \cap \mathrm{Atoms}(\mathcal{N}; P)$. V is a P-cover of A, since by Proposition 6.2 there is a model w of A with a parameter cover of P, such that V is the set of P atoms true in w. To show that V is a minimal P-cover, assume that there is some model w' of A with a parameter cover of P whose P-cover is a proper subset of V. Because A is first-order, we can choose w' to respect V. Then we have:

$$w' \models_V \mathrm{M}(A; P),$$

and hence $\{\phi \mid \mathrm{M}(A; P) \models_V \phi\}$ does not contain all of V, contradicting the assumption that V is an atom base of $\mathrm{M}(A; P)$.

COROLLARY 6.4 *The AE-consequences of $\mathrm{M}(A; P)$ are those sentences true in the minimal P-cover models of A.*

Because extensions of MIN theories depend on the presence of minimal P-covers, first-order theories without such covers will have no extensions. This is analogous to the case of circumscription, in which theories with no minimal models fail to have a consistent circumscription. The following example (from [3]) illustrates this correspondence.

EXAMPLE 6.2 Let A be the set

$$\exists x. \, Nx \wedge \forall y.[Ny \supset x \neq sy]$$

$$\forall x. \, Nx \supset Nsx$$

$$\forall xy. \, sx = sy \supset x = y \, .$$

Let $\mathcal{N} = \{0, s0, ss0, sss0, \cdots\}$. Any model of A that is a parameter cover for N has an N-cover of the form $\{s^i0, s^{i+1}0, s^{i+2}0, \cdots\}$. There are no minimal N-covers, and hence no extensions of A. Similarly, the circumscription of A with respect to N is inconsistent, because there are no N-minimal models of A. On the other hand, consider the skolemization of A (call it A'):

$$N0 \wedge \forall y.[Ny \supset 0 \neq sy]$$

$$\forall x. \, Nx \supset Nsx$$

$$\forall xy. \, sx = sy \supset x = y \, .$$

This theory does have one minimal N-cover, namely $\{N0, Ns0, Nss0, \cdots\}$, so there is one extension, all of whose models are isomorphic to the natural numbers, with 0 as interpreted as zero and s as successor. Similarly, the circumscription of A' with respect to N has just these models.

7 Representational Issues

We can use the results of the previous section to state defaults about knowing all the members of a class, about knowing all the members of a domain, and about equality among individuals.

7.1 Comprehensive Class Knowledge

Suppose we would like to state that the only members of a class are the ones we believe to be in the class, that is, our knowledge of the class is comprehensive. For example, we might want to make the assumption that if we don't know of any block on top of block B, then B is clear.[4] Using $\mathcal{L}$, we could add the default

$$On(a, B) \supset LOn(a, B)$$

for every term a in the language. But although this will enable us to conclude $\neg On(a, B)$ for every term, it will not sanction the generalization $\forall x. \neg On(x, B)$.[5]

In $q\mathcal{L}$, we can use quantifying-in to refer to every individual in the domain inside the belief operator:

(22) $\quad \forall x. On(x, B) \supset LOn(x, B),$

which is just the theory $M(A; \lambda(x)On(x, B))$ for empty A. From Corollary 6.4, we know that the sentences that are AE-consequences of $M(A; \lambda(x)On(x, B))$ are just those in the minimal $On(*, B)$-covered models of A. In every such model, there is no individual on B, and so $\forall x. \neg On(x, B)$ is an AE-consequence of the theory.

In general, if only a finite set of atoms $Pa_1, Pa_2, \cdots Pa_n$ are provable in a theory A, then an AE-consequence of $M(A; P)$ will be $\forall x. Px \supset (x = a_1 \lor x = a_2 \lor \cdots \lor x = a_n)$. For a theory A that implies an infinite set of ground atoms, the AE-consequences will depend on the structure of A. For example, if A asserts that all members of $\mathcal{N}$ have the property P, then $\neg \exists x. \neg Px$ will be an AE-consequence of $M(A; P)$.

It is also possible for an agent to express incomplete knowledge of class membership in $q\mathcal{L}$, with the statement

(23) $\quad \exists x. \phi(x) \land \neg L\phi(x).$

Here there is an individual who is in the class ϕ but who is not known to be by the agent.

---

[4]This example is taken from [18]; the inadequacy of default logic (and thus AE logic over $\mathcal{L}$) for this problem has been discussed in [15, 16, 22].

[5]That is, unless we also assume parameter interpretations, which in general we do not want to do, since it requires giving a name to every individual in the domain.

Generally an agent may not know either of (22) or (23). For example, consider the simple premise set $\{Pa\}$. There is a single extension whose kernel is $\text{Cn}(Pa)$, and using the any-name semantic equations we find

(24)
$$\neg L[\exists x.\,\phi(x) \wedge \neg L\phi(x)]$$
$$\neg L[\forall x.\,\phi(x) \supset L\phi(x)]\,.$$

are both in the extension.

7.2 Domain Closure and Domain Circumscription

The domain closure assumption is similar to that of parameter interpretations: every element of the domain is represented by a term, but the terms need not refer to unique individuals. In first-order logic, the domain closure assumption can be expressed for finite sets of terms by the sentence

(25) $\quad \forall x.\, x = a_1 \vee x = a_2 \cdots \vee x = a_n\,.$

There is no first-order sentence to express domain closure for a countably infinite set of terms. We can do this in $q\mathcal{L}$, however.

PROPOSITION 7.1   *If A is a set of nonmodal sentences, $A \cup \{\forall x.L(x = x)\}$ has a single extension whose kernel is the set of all first-order sentences true in models of A whose domains are covered by $\mathcal{N}$.*

> *Proof.* Let W be the set of models of A with the domain cover assumption, and let $S = \{\phi \in \mathcal{L}_0 \mid W \models \phi\}$. We want to show that
>
> $$S = \{\phi \in \mathcal{L}_0 \mid A \cup \{\forall x.L(x = x)\} \models_S \phi\}\,.$$
>
> The only effect $\forall x.L(x = x)$ has is to restrict the models of A to those with a domain cover. Since S is precisely the set of sentences true in such models, the equation holds.

If the set $\mathcal{N}$ is finite, then $\forall x.L(x = x)$ is equivalent to (25). If $\mathcal{N}$ is infinite, then all models respecting $\forall x.L(x = x)$ have domains that are countably infinite or less. The effect of this depends on the first-order axioms A. For example, if $A = \{Pa_1, Pa_2, \cdots\}$ for every name $a_i \in \mathcal{N}$, then the extension of A and $\forall x.L(x = x)$ contains $\forall x.Px$, which is not a first-order consequence of A.

This last example exhibits a phenomenon first noticed by Levesque [12], that the kernel of an extension in $q\mathcal{L}$ does not necessarily completely determine the extension. This is unlike the case for $\mathcal{L}$, in which different extensions must have different kernels ([21]). To see how this is so, consider the case above in which A asserts Pa for

30

every name $a \in \mathcal{N}$. Let us suppose also that all individuals (named or not) have the property P, so that $\forall x.Px$ is also in A. By Proposition 4.1, A has a unique extension with $\text{Cn}(A)$ as its kernel. This extension contains $s = \neg L[\forall x.\, Px \supset LPx]$, since if there are any unnamed individuals, they have the property P without the agent knowing it (see the discussion leading up to Equation (24)). Now consider the theory $M(A; P)$. By Corollary 6.4, there is a single extension of $M(A; P)$, since there is exactly one minimal set of P-atoms over $\mathcal{N}$. The first-order AE-consequences of A' are exactly the same as those of A, namely $\text{Cn}(A)$. So we have two AE theories with the same kernels, but differing in their modal sentences: in one, the agent has a name for all P-individuals, and in the other, he does not know if he does. This difference is a reflection of the increased expressive power of $q\mathcal{L}$ relative to first-order logic, since no first-order axioms can characterize countable models, but MIN theories can.

We can also express a kind of domain circumscription in $q\mathcal{L}$, by minimizing the inequality in the models of a theory. If A is a set of nonmodal sentences, this is expressed by $M(A; \lambda(x, y)x \neq y)$. According to Corollary 6.4, the AE-consequences of this theory are just the sentences true in all the minimal $\neq$-cover models of A. In such models, since any two distinct individuals are unequal, every individual must have a name, that is, they must be parameter models.

EXAMPLE 7.1   Let $A = \{a \neq b\}$. In every parameter model of A minimal in inequality statements, each element must be equal to a or to b. Hence $\forall x.\, x = a \lor x = b$ is an AE-consequence of $M(A; \lambda(x, y)x \neq y)$.

Let $A = \{a \neq b \lor (a \neq c \land a \neq d \land c \neq d)\}$. Now there are two minimal sets of inequality statements, and thus two extensions, one restricted to domains of cardinality two, and one to domains of cardinality three.

7.3   Reasoning about Equality

It has been noted (see [3]) that in its original formulation, circumscription of a predicate does not allow any conclusions about equality to be made that are not present in the original theory. This is because only interpretations with the same denotation function and domain are compared. As a consequence, reasoning about defaults sometimes turn out to be counterintuitive. The following example uses the abnormality predicate proposed by McCarthy [19].

EXAMPLE 7.2   Consider a simple abnormality theory, with $A = \{\forall x.Px \land \neg ab(x) \supset Qx,\ Pa,\ \neg Qb\}$. We would expect Qa to be a consequence of minimizing P (while allowing Q to vary), but it is not. The reason is that there are abminimal models of A in which b and a refer to the same individual, and $\neg Qa$ is true.

31

In more recent versions of circumscription [13], terms are also allowed to vary their denotation across interpretations being compared. In this case, the example above has Qa as one of its consequences. However, there are also cases where allowing terms to vary leads to undesirable results.

EXAMPLE 7.3   Consider a simple abnormality theory with $A = \{\forall x.Px \land \neg ab(x) \supset Qx, ab(a), ab(b)\}$. In this case, $a = b$ is a consequence of minimizing P while allowing Q and a, b to vary. 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 comprehensive class axioms have a similar effect to minimizing the extent of a predicate, but they allow new conclusions about equality to occur. Since the minimization is partially based on the presence of sentences in the belief set, they do not necessarily force identification of terms.

EXAMPLE 7.4   Redoing Example 7.2, let $A = \{\forall x. Px \land \neg ab(x) \supset Qx, Pa, \neg Qb\}$, and assume $a, b \in \mathcal{N}$. There is one extension of M$(A; P)$, with kernel Cn$(A, \forall x. Px \supset x = a, a \neq b)$. The conclusion that a and b are different individuals is not derivable from A alone.
For the abnormality theory $A = \{\forall x.Px \land \neg ab(x) \supset Qx, ab(a), ab(b)\}$, there is again a single extension of M$(A; P)$ with kernel Cn$(A, \forall x. ab(x) \supset (x = a \lor x = b))$. There is no conclusion about the equality or inequality of a and b.

The comprehensive class axiom is 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.

Another type of default about assumption about equality is the unique names assumption (important in the theory of deductive databases [25] and logic programming [1]). In its most basic form, it asserts that all terms denote unique individuals, i.e., it is the assumption of Herbrand models, without necessarily having a term for every individual.[6] The unique name assumption is usually axiomatized by including all sentences of the form

---

[6]Clark [1] gives a more complex notion of unique names, in which all constants and functions have their free interpretation. We have not tried to formalize this interpretation in $q\mathcal{L}$.

(26) $a_i \neq a_j$ for $i \neq j$.

We can formalize a default interpretation of the unique names assumption by minimizing equality. For any first-order theory A, the AE theory $M(A, =)$ has as its consequences all sentences true in the minimal $=$-cover models of A.

EXAMPLE 7.5 Let $A = \{a_1 = a_2\}$. In every model of A minimal in equality statements, each name of $\mathcal{N} - \{a, b\}$ must be unequal to every name in $\mathcal{N}$. Hence $a_i \neq a_j$ is an AE-consequence of $M(A; =)$ whenever $i \neq j$ and i and j are not both 1 or 2.

Recently Rathmann and Winslett [23] have proposed a different type of circumscription, using the concept of homomorphism to compare models. They use this technique to form a type of equality circumscription, in which distinct terms are assumed to designate distinct individuals whenever possible. They are able to show that the unique name axioms are present in the equality circumscription of a theory T whenever T contains no positive instances of the equality predicate.

8 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

$$A \quad \bigcup_i \{\forall x(\neg L P_i x \supset \neg P_i x)\}$$
$$(27) \qquad \bigcup_i \{\forall x(\neg L Q_j x \supset \neg Q_j x)\}$$
$$\bigcup_i \{\forall x(\neg L \neg Q_j x \supset Q_j x)\} \;,$$

where the P_i and Q_j are sequences of predicates. We write $M(A; P_i; Q_j)$ to indicate the MIN theory of A over the predicates P_i with Q_j fixed.

There is a reduced fixed point equation for MIN$^=$ theories that is similar to Proposition 6.1. Let Lits$(A; P_i; Q_j)$ stand for the set of atoms $P_i a$ and literals $Q_j a$, $\neg Q_j a$ for all $a \in \mathcal{N}$. These are the literals important in defining the fixed point of $M(A; P_i; Q_j)$.

PROPOSITION 8.1 *S is the kernel of an extension of* $M(A; P_i; Q_j)$ *if and only if*

$$S = \{\phi \in \mathcal{L}_0 \mid M(A; P_i; Q_j) \models_{S \cap \text{Lits}(\mathcal{N}; P_i; Q_j)} \phi\}$$

Proof. By Proposition 4.3, S is a kernel of a MIN$^=$ theory iff

$$S = \{\phi \in \mathcal{L}_0 \mid M(A; P_i; Q_j) \models_S \phi\} \;.$$

Since the only modal atoms in $M(A; P_i; Q_j)$ are of the form $L P_i x$, $L Q_j x$, and $L \neg Q_j x$, the only sentences of S relevant to the truth of these atoms are of the form $P_i a$, $Q_j a$, and $\neg Q_j a$ for $a \in \mathcal{N}$.

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

EXAMPLE 8.1 Let $A = \{\neg Pa \supset Qa\}$. The MIN theory $M(A; P; Q)$ has two classes of extensions: one class contains $\{\neg Pa, Qa\}$, while the other contains $\{Pa, \neg Qa\}$. 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 $\mathcal{N}$ of names in $\mathcal{L}$.

One consequence of the any-name semantics is that fixing any predicate causes every element of the domain to be named by $\mathcal{N}$, since any element in either the positive or negative extent of a predicate must have a name. Further, every extension is saturated in the fixed predicates, that is, either Qa or $\neg Qa$ is in the extension for every name $a \in \mathcal{N}$.

35

PROPOSITION 8.2 *Let S be the kernel of an extension of $M(A; P_i; Q_j)$. Any interpretation w of $M(A; P_i; Q_j)$ respecting S is a parameter model over $\mathcal{N}$. S is saturated with respect to the predicates Q_j.*

Proof. Suppose w is not a parameter model. Then there is some element e of the domain such that e is not denoted by any name. Thus both $\neg LQ_j x$ and $\neg L\neg Q_j x$ are true for $x = e$, and this leads to the contradiction $Q_j x \wedge \neg Q_j x$ for $x = e$.

To show S is saturated, assume it is not for some Q_j. Then there is some $c \in \mathcal{N}$ such that $Q_j c$ and $\neg Q_j c$ are not in S. Then there is some individual x for which $\neg LQ_j x$ and $\neg L\neg Q_j x$ hold, which leads to a contradiction.

9 MIN= Theories

The equality predicate can be fixed, just as any other predicate. We single out the class of MIN theories $M(A; P; =)$ as interesting because they can be related to minimal models of P; we call these MIN= theories. Our main result is that the extensions of a MIN= theory of A are just the sentences true in all countable P-minimal models of A. The proof is somewhat lengthy, and we defer it to the end of this section, first pointing out the significance and limitations of the result.

THEOREM 9.1 *Let A be a set of first-order sentences not containing a countably infinite number of names from $\mathcal{N}$. The first-order sentences S true in every extension of $M(A; P; =)$ are exactly those true in the countable P-minimal models of A.*

Remarks. This result shows that MIN= theories are closely related to minimization over countable model structures. Schlipf [26] gives some useful results for this type of minimization. First, the question of whether an arbitrary finite first-order A has a countably infinite minimal model is Σ_2^1, and the question of whether a formula holds in all countable minimal models is Π_2^1. These complexity classes are much worse than the undecidability of first-order logic; there can be no complete proof theory for $q\mathcal{L}$.

A second result from Schlipf is that entailment over minimal countable structures is not equivalent to entailment over all structures (unlike ordinary first-order entailment, which is reducible to entailment in countable structures). He gives an example (Example 3, p. 93) of a first-order theory that has minimal uncountable models, but no minimal countable ones. Most of the theories we are concerned with in AI will not have this complex a structure, however, and we can usually be satisfied with entailment in minimal countable structures.[7]

Given this caveat about countable structures, we can explore the relationship between predicate circumscription and autoepistemic logic with any-name semantics. The semantics of the second-order circumscription schema $Circum(A; P_i; Z_j)$ is a second-order formula characterizing the P_i-minimal models of A, allowing all predicates Z_j to vary (see [14]). Thus, assuming Z_j are all the predicates of A except P_i, the first-order consequences of circumscription over countable models are the AE-consequences of a corresponding MIN= theory $M(A; P; =)$; that is, given that $\mathcal{N}$ contains an infinite set of names not used by A.

From the results of [2], 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 $Circum(A; P_i; Z_j)$ with $Q \notin Z_j$ is equivalent

---

[7]An interesting subclass are the universal theories; it is an open question whether minimal entailment for these theories reducible to minimal entailment over countable models.

to Circum$(A \wedge Q' = \neg Q; P_i, Q, Q'; Z_j)$, where Q' is a new predicate constant. The corresponding construct for MIN= theories is to fix Q using M$(A; P_i; =, Q_j)$ (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 over countable models are given by the corresponding MIN= theory.

We now turn to the proof of the theorem. First we develop the result that a first-order sentence is true in the P-minimal parameter models of A just in case it is true in every extension of M$(A; P; =)$. Then, we show that parameter models are a sufficient semantics for minimal entailment over all countable models.

For any parameter model w of A (over $\mathcal{N}$), call a P-*diagram* of A the set of ground P atoms and ground equality literals over $\mathcal{N}$ that are true in w. A P-diagram D is minimal if there is no other diagram with the same equality literals whose P atoms are a subset of D's. We first show that any minimal P-diagram of A picks out a unique extension of M$(A; P; =)$, for which w is a model; and conversely, every extension is formed by using a minimal P-diagram of A as its base.

LEMMA 9.2 *Let A be a set of first-order sentences, and D a P-minimal diagram of A. If S satisfies the equation*

$$S = \{\phi \in \mathcal{L}_0 \mid M(A; P; =) \models_D \phi\}$$

then it is the kernel of an extension of M$(A; P; =)$.

Proof. We will show that the restriction of S to ground atomic P and equality literals is exactly the set D, and so by Proposition 8.1 S is the kernel of an extension of M$(A; P; =)$. Note that D is saturated with respect to equality literals: for all terms a and b, one of $a = b$ or $a \neq b$ is in D. From the fixing of equality in M$(A; P; =)$, all of these are also contained in S. We know that $S \cap \text{Atoms}(\mathcal{N}; P) \subseteq D$, since by definition there is model of $A \cup D$ that makes all Pc not in D false. To show that S contains $D \cap \text{Atoms}(\mathcal{N}; P)$, suppose to the contrary there is some Pc in D that is not in S. Then there is a model of M$(A; P; =)$ respecting D that makes $D - \{Pc\}$ true, but no other P atoms, since $Ax.\, Px \supset LPx$ forces all P atoms not in D to be false. But we assumed D was a minimal P-diagram, and w contradicts this. $\blacksquare$

LEMMA 9.3 *Let A be a set of first-order sentences. If S is the kernel of an extension of* M$(A; P; =)$, *then every model satisfying* M$(A; P; =)$ *and respecting S is a P-minimal parameter model of A.*

Proof. By Proposition 8.1 S satisfies the equation

38

$$S = \{\phi \in \mathcal{L}_0 \mid \mathrm{M}(A; P; =) \models_D \phi\},$$

where D is $S \cap \mathrm{Lits}(\mathcal{N}; P; =)$, and D is saturated with respect to equality literals by Proposition 8.2. Define E as the set of equality literals in D. Let w be a model of $\mathrm{M}(A; P; =)$ respecting D. w is a parameter model satisfying E by virtue of the fixed point equation for extensions; by Proposition 6.2 it also satisfies the P atoms of D, and no others. Thus D is a P-diagram of A. Further w must be minimal over parameter models satisfying $A \cup E$. Assume to the contrary that there is a model w' of $A \cup E$ whose P atoms are a subset of w's; we can choose w' to respect D. Then w' satisfies $\mathrm{M}(A; P; =)$, but some atom Pc of D is not a consequence of w', and so not in S, a contradiction. Thus D is a minimal P-diagram of A.

PROPOSITION 9.4 *Let A be a set of first-order sentences. A first-order sentence is true in all the P-minimal parameter models of A if and only if it is an AE-consequence of $\mathrm{M}(A; P; =)$.*

Proof. Let $\phi \in \mathcal{L}_0$ be true in all P-minimal parameter models of A. Suppose S is the kernel of some extension of $\mathrm{M}(A; P; =)$, and $\phi \notin S$. By Lemma 9.3, every model of $\mathrm{M}(A; P; =)$ respecting S is a P-minimal parameter model of A, and so S must contain ϕ by the fixed-point equation, a contradiction.

Conversely, suppose ϕ is true in all extensions of $\mathrm{M}(A; P; =)$. Let w be an arbitrary P-minimal parameter model, and assume ϕ is false in it. w defines a minimal P-diagram D of A, which in turn defines an extension of $\mathrm{M}(A; P; =)$ by Lemma 9.2, and w is a model of its kernel. Since ϕ is in the extension, it must be true in w, a contradiction.

We have now shown that minimal entailment over parameter models is the same as AE-consequence in MIN$^=$ theories. Recall that, 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 A has a model if and only if it has a parameter model. Note that this statement is not true in general if A contains all members of $\mathcal{N}$; for example the set $\{\exists x \neg Px, Pa_1, Pa_2, \cdots\}$, in which Pa_i is asserted for every name $a_i \in \mathcal{N}$ has a model but no parameter model. Thus we must stipulate that a countably infinite set of names in $\mathcal{N}$ do not appear in A. For minimal entailment over *countable* models, parameter models are sufficient.

PROPOSITION 9.5 *Let ϕ be a sentence of $\mathcal{L}_0$ and A a set of sentences of $\mathcal{L}_0$, such that some infinite subset of names in $\mathcal{N}$ do not appear in either ϕ or A. Then ϕ is true in all P-minimal countable models of A (with the same domain and denotation function) if and only if it is true in all P-minimal parameter models.*

39

Proof. Suppose $A \cup \{\phi\}$ has a P-minimal countable model w. Let the constants $C = \{c_1, c_2, \cdots\}$ be those members of $\mathcal{N}$ not mentioned by ϕ or A. Let k_i be the individual referred to by c_i. Let $E = \{e_1, e_2, \cdots\}$ be those individuals with no names in w. Construct w' as follows: it has the same relations and domain as w, but the denotation function v is modified so that $v(c_{2i}) = e_i$, $v(c_{2i-1}) = k_i$ for $i > 0$. Every element has a name, and furthermore the truthvalue of A and ϕ in w' is the same as w.

To show that w' is minimal in P among parameter models, assume there is another model w'' with the same domain and denotation function that has a smaller extension of P. By reassigning $v(c_i) = k_i$, we could construct a model of A and ϕ whose extension of P would be smaller than in w, a contradiction.

In the converse direction, if $A \cup \{\phi\}$ has a minimal parameter model, it obviously has a minimal countable model.

10 Related Work

This scheme for extending the semantics of AE logic to the quantified-in case is similar to that proposed for the epistemic operators in [7]. The problems of knowing who and quantifying-in are discussed there in some detail. Some of the results that are discussed here were first given in [9], especially the connection to circumscription.

Levesque [11] was the first to consider using a self-referential modal language to formalize a knowledge base's information about its own contents. Although Moore's autoepistemic logic [21] was developed independently, it is clear there are many ideas in common between the two.

In [12] Levesque considers the problem of quantifying-in for AE logic, using techniques previously developed in [11]. His approach differs from standard $q\mathcal{L}$ semantics in that it is based on the denotation of terms rather than their intension. He considers a revision of AE logic in which a belief sets are replaced by sets of interpretations, which he calls *assignments*. He then recasts the fixed point semantics of AE logic in terms of truth-recursion equations on maximal sets of assignments, in a manner similar to the possible-world semantics of standard epistemic logics. Let W be a set of assignments, and w any member of W. Then:

(28) $W, w \models L\phi$ iff for all $w' \in W$, $W, w' \models \phi$

If ϕ is nonmodal, then $W, w \models \phi$ is the same as $w \models \phi$, that is, W is used only in the interpretation of modal atoms, just like the belief set in (7).

Because the semantics is based on assignments rather than belief sets, there is a natural way to understand quantifying-in: $W, w \models \exists x. L\phi(x)$ just in case there is some individual in w such that $L\phi(x)$ is true in all assignments of W. This is similar to the analysis of quantifying-in given for possible-world semantics of epistemic logics, in which the same individual is picked out in each possible world.

To axiomatize this semantics, Levesque chooses a language in which the only terms are a countable set of standard names, so that every individual in an assignment is denoted by a unique standard name. This simplifies the analysis of equality, but it also bypasses the representational issues we raised in Sections 5 and 7; for example, there is no way to state a default version of the unique names axioms. It is an interesting conjecture, and one that we have not pursued, that our version of AE logic with the assumption of parameter models and unique names is equivalent to Levesque's.

Levesque does provide a sound proof theory for AE consequence that has complexity Π_1^0, that is, the same as the consistency problem for first-order logic. He has not been able to show that the proof theory is complete. Given the results of Schlipf on the complexity of AE consequence in $q\mathcal{L}$, it is a reasonable conjecture that it is not.

41

Closely related to free variables in modal contexts are the open defaults of default logic. Lifschitz [16] gives a semantics for open defaults that is similar to Levesque's use of sets of assignments and standard names, except that he allows standard names over arbitrary universes, including uncountable ones. He considers extensions of open default theories for a fixed universe U. An F-consequence of an open default theory is a sentence true in all such extensions. He is able to show a result similar to ours, that the default logic analogue to $M(A; P; =)$ theories characterizes the P-minimal models of A. His result is more general, since there is no restriction to countable models. On the other hand, because default logic only allows universal quantification in open defaults, it is not as expressive as AE logic over $q\mathcal{L}$.

11 Conclusion

Adding quantifying-in to the language of AE logic greatly increases its expressive power. We are able to express defaults about comprehensive class knowledge (minimizing a predicate), domain closure, and unique names. Because the semantics is based on the intension of names, rather than their denotation, we are also able to make default assumptions that are sensitive to the possibility that different names denote different individuals, without absolutely requiring that this be so, as in Herbrand interpretations. This kind of reasoning is problematic for circumscriptive techniques, and not adequately addressed in default or AE logic without quantifying-in.

Given the importance of understanding the connections between the proliferating number of nonmonotonic formalisms, it is encouraging to note the close correspondence between AE logic with quantifying-in and consequence in countable minimal models. Along with the results of Lifschitz [16], we get a much clearer picture of how minimization is expressed in nonmonotonic logics that depend on fixed-point constructions, and it is interesting that these logics have similar expressive capabilities. The unfortunate part of this correspondence is that the complexity of AE consequence for $q\mathcal{L}$ (and for open defaults under F-consequence) is too high to hope for any type of proof theory in the general case. A recourse would be to find subsets of $q\mathcal{L}$ whose AE-consequences are recursive or recursively enumerable; this is certainly a target for research.

Given the close connections between AE logic and negation as failure in logic programming [4, 17], the results of this paper, especially concerning the nature of reasoning about individuals and equality, might be usefully applied to the study of quantification in negation as failure. Such a study could yield insight into the answers that general logic programs return, especially if we want to ask, "are the answers returned *all* the individuals that satisfy the query?"

References

[1] K. Clark, Negation as failure, in: *Logic and Data Bases* (Plenum Press, New York, 1978) 293–322.

[2] J. de Kleer and K. Konolige, Eliminating the fixed predicates from a circumscription, *Artificial Intelligence* **39** (1989) 391–398.

[3] D. W. Etherington, R. E. Mercer, and R. Reiter, On the adequacy of predicate circumscription for closed-world reasoning, *Computational Intelligence* **1** (1) (February 1985).

[4] M. Gelfond and V. Lifschitz, The stable model semantics for logic programming, in: *Proceedings of the International Conference on Logic Programming* (1988).

[5] A. J. Grove and J. Y. Halpern, Naming and identity in a multi-agent epistemic logic, in: *Proceedings of the International Conference on Knowledge Representation and Reasoning*, Cambridge, MA (1991) 301–312.

[6] J. Hintikka, *Knowledge and Belief* (Cornell University Press, Ithaca, New York, 1962).

[7] K. Konolige, *A Deduction Model of Belief* (Pitman Research Notes in Artificial Intelligence, 1986).

[8] K. Konolige, On the relation between default and autoepistemic logic, *Artificial Intelligence* **35** (3) (1988) 343–382.

[9] K. Konolige, On the relation between autoepistemic logic and cirucmscription, in: *Proceedings of the International Joint Conference on Artificial Intelligence*, Detroit, MI (1989).

[10] Y. Lespérance, A formal account of self-knowledge and action, in: *Proceedings of the International Joint Conference on Artificial Intelligence*, Detroit, MI (1989) 868–874.

[11] H. J. Levesque, Foundations of a functional approach to knowledge representation, *Artificial Intelligence* **23** (2) (July 1984).

[12] H. J. Levesque, A study in autoepistemic logic, *Artificial Intelligence* **42** (2–3) (March 1990).

[13] V. Lifschitz, Some results on circumscription, in: *AAAI Workshop on Non-Monotonic Reasoning*, Menlo Park, California (1984).

[14] V. Lifschitz, Computing circumscription, in: *Proceedings of the International Joint Conference on Artificial Intelligence*, Los Angeles, California (1985) 121–127.

[15] V. Lifschitz, Between circumscription and autoepistemic logic, in: *Proceedings of the First International Conference on Knowledge Representation and Reasoning*, Toronto, Ontario (1989).

[16] V. Lifschitz, On open defaults, in: *Proceedings of the Symposium on Computational Logic*, Brussels (1990).

[17] W. Marek and V. S. Subrahmanian, The relationship between stable, supported, default and auto-epsitemic semantics for general logic programs, Technical Report 128-88, Department of Computer Science, University of Kentucky (1988).

[18] J. McCarthy, Circumscription — a form of nonmonotonic reasoning, *Artificial Intelligence* **13** (1-2) (1980).

[19] J. McCarthy, Applications of circumscription to formalizing commonsense knowledge, *Artificial Intelligence* **28** (1986).

[20] R. C. Moore, Reasoning about knowledge and action, PhD thesis, Massachusetts Institute of Technology, Cambridge, MA (1980).

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

[22] D. Poole, Fixed predicates in default reasoning, in: *Proceedings of the International Joint Conference on Artificial Intelligence*, Milan, Italy (1987) 905–908.

[23] P. K. Rathmann and M. Winslett, Circumscribing equality, in: *Proceedings of the International Joint Conference on Artificial Intelligence*, Detroit, MI (1989) 468–473.

[24] R. Reiter, Equality and domain closure in first-order data bases, *Journal of the ACM* **27** (1980).

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

[26] J. S. Schlipf, How uncomputable is general circumscription, in: *Proceedings of the Symposium on Logic in Computer Science*, Cambridge, MA (1986) 92–95.

[27] R. C. Stalnaker, A note on nonmonotonic modal logic, Department of Philosophy, Cornell University, (1980).