An autoepistemic logical view of knowledge base

Y.J. Jiang
Computer Laboratory, Cambridge University,
Cambridge CB2, 3QG, UK

Abstract

Autoepistemic logic (AE) is a non-monotonic logic for modelling beliefs of agents who reflect on their own beliefs. In this paper, we will take such a logical view of knowledge base by treating its contents as its beliefs about a world and its integrity constraints as its beliefs about its contents. We will show how such a view can help us to represent and reason about incomplete knowledge, self-knowledge and negative information. We will also show that an AE logical closure of a knowledge base will neither suffer the inconsistency problem nor the logic-impurity problem that often persist in the standard nonmonotonic closures of a knowledge base. In particular, we will show that an AE logic view of integrity constraints provides a finer way of defining integrity constraints than existing definitions. For the logic to be effective, we introduce a stratified AE proof theory for evaluating queries and maintaining integrity constraints. It is shown that the AE logical view of a stratified knowledge base will yield a unique AE closure of the knowledge base.

Keywords: AI in Database, Deductive database, Incomplete knowledge, Non-monotonic logic, Autoepistemic logic, Closed World Assumption, Complete Database, Integrity Constraints. Nonstandard Logic, Possible worlds semantics, Modal logic.

1 Introduction

A knowledge base (KB) is generally an incomplete description of the world (or domain of application). An effective knowledge base management system therefore should make the proper plan on the basis of what currently the KB knows and does not know, rather than to wait for the state of complete world which could never be obtained. This suggests the need for a formalism to represent and reason about incomplete knowledge and self-knowledge.

The incompleteness problem is not new in database models. A number of solutions have been proposed. In Codd's relational model [Codd 82], the unknown number in "Mary's tele-no is not known" is indicated by a null value. This approach however is not sufficient enough to represent disjunctive incomplete knowledge such as "Mary's tele-no is 1 or 2". This problem is solved in Lipski's model by associating a set of elements for Mary's tele-no. If the set is empty, then it is not known; else if the set contains one element, then the value is known; otherwise, the value is only known to be a member of the set.

These solutions are nevertheless very restrictive. For example, they cannot represent accumulative objects, eg. "the person who likes Mary also likes Sue"; neither can they represent disjunctive information of the kind in which there is uncertainty regarding which object is being characterized, eg. "Mary's tele-no is 1 or John's teleno is 2".

Although many of these problems can be solved in a deductive logical view of database systems [Reiter 84, Galliari et al 84], they still cannot represent and reason about what a KB knows and does not know. Often, negative information cannot be represented. This is because positive information that is not present in the KB, is usually assumed asymmetrically to be negative in order to form a complete closure of a partial KB. As a result, negative connectives becomes non-logical, usually in the disguise of Clark's Negation as Failure [Clark 78].

These problems have been partially addressed by Levesque [84]. In his approach, a first order (FOL) knowledge base is characterized by an autoepistemic logic closure. In Levesque's view, a KB is simply a set of facts about the world; while queries and integrity constraints are treated as its beliefs about the contents of the KB. However Levesque's approach is limited to knowledge bases that contain only FOL formulae. Thus in Levesque's approach, we can make queries and integrity constraints about what a KB knows and does not know; but we cannot represent in the KB itself what it knows and does not know. For example,
given Tele-no(Mary,100) in a KB, we can ask if the KB knows about Mary's tele-no; on the other hand, we cannot represent in a KB the knowledge "Mary's tele-no is known" (i.e. $\exists x \text{Know}(\text{Tele-no}(\text{Mary},x))$) without giving the actual number.

Our position in this paper is to extend Levesque's approach by additionally viewing KB as a set of beliefs rather than facts about the world. In this way, a KB can express beliefs about its beliefs and non-beliefs of the world. For example, a KB can represent its belief "if it is not known that Mary works, then Mary is a housewife". We will formalize our position in an autoepistemic predicate logic. In particular in contrast with Levesque's approach which uses the full power of first order proof mechanization for his query evaluation, we will attempt to develop an effective epistemic proof procedure for query processing and integrity checking.

The paper is organized as follows. Section 2 introduces an AE predicate logic that allows quantifying-in variables. It also defines the semantics of such a logic view of KB. Section 3 and 4 compares and contrasts such a semantical view of KB with some existing semantics for queries and integrity constraints. Section 5 then develops an effective proof procedure for a stratified AE predicate logic which has the property of unique AE extension of a KB.

2 Autoepistemic predicate logic

Autoepistemic (AE) logic was defined by Moore [85] as a formal account of an agent reasoning about her own beliefs. In the autoepistemic view of a KB, the KB is identified as such an agent. It can thus express knowledge such as "It is not known that John has a telephone" and "It is known that John lives either in the city where he works or in the city where his wife lives". For convenience reason, we will use KB to denote the knowledge base both as an agent and as the initial set of AE formulae in the knowledge base.

AE logic was originally only defined for propositional case. Otherwise, the logic is restricted to closed epistemic formulæ in the sense that no quantifying-in variables (i.e. variables quantified outside the scope of their B-operators) are allowed. Although AE logic is often claimed for its ability to reason about incompleteness, however many types of incomplete knowledge in KB applications are quantifying-in types of AE formulæ. For example, the beliefs "All the managers are known" and "There are definitely some girls who are not known" can be represented in a KB respectively as follows.

\[
\forall x (\text{Manager}(x) \rightarrow B(\text{Manager}(x))).
\]

\[
\exists x (\text{Girl}(x) \land \neg B(\text{Girl}(x))).
\]

In this section, we will therefore attempt to develop an AE logic that allows quantifying-in variables.

2.1 Semantics of AE logic of KB

The language of AE predicate logic is that of standard first order logic, augmented by the modal operator B. We will call a set of AE formulæ an AE theory.

The standard semantics of an AE propositional logic is usually formulated in a recursive fashion (e.g. [Konolige 87]). This makes it difficult to exhibit AE extensions. It also fails to establish a connection between the monotonic logic of belief [Halpern & Moses 85] and AE logic. To avoid these problems without changing AE logic, we reformatulate the standard semantics of AE predicate logic (e.g. [Konolige 89]) in a possible worlds semantics.

We define the following Kripke-style possible worlds model structure $M=(W, D, S5, w_0)$ for AE predicate logic.

**Definition 1**  
1. $W$ is a non-empty set of possible worlds;
2. $D$ is the constant domain of all possible worlds.
3. $S5$ denotes the equivalence accessibility relation for all worlds in $W$; i.e. $W$ forms a complete S5 structure.
4. $w_0$ is the actual world and every world in $W$ is accessible from $w_0$.
5. A world is a function which assigns to each n-place function symbol a function from $D^n$ to $D$, and which assigns to each n-place predicate symbol, a set of tuples of $D^n$; and which assigns to each non-domain constant an element in $D$. 

Given the above semantic structure $M$, we can define the following standard satisfiability (or truth assignment) relation $|=_{M}$ recursively between a world and a belief as follows for AE predicate logic.

1. $w|=_{M} F(t_{1},..,t_{n})$ iff $\langle \text{Val}(t_{1},w),..,\text{Val}(t_{n},w)\rangle \in w(F(p))$ where $\text{Val}(t,w) = w(t)$ if $t$ is a non-domain constant, $\text{Val}(t,w) = t$ if $t$ is a domain element, $\text{Val}(f(t_{1},..,t_{n}),w) = w(f)(\langle \text{Val}(t_{1},w),..,\text{Val}(t_{n},w)\rangle)$

2. $w|=_{M} B\phi$ iff for all $w' \in W$, $w'|=_{M} \phi$

3. $w|=_{M} \neg \phi$ iff $w|=_{M} \phi$

4. $w|=_{M} \phi \lor \psi$ iff $w|=_{M} \phi$ or $w|=_{M} \psi$

5. $w|=_{M} \forall x \phi(x)$ iff for all $d \in D$, $w|=_{M} \phi(d/x)$ where $\phi(d/x)$ is obtained by replacing all $x$ in the formula $\phi$ by $d$.

Definition 2 An AE theory is satisfied/true in a world iff every formula in the theory is true in the world.

Definition 3 An AE theory is satisfied/true in a set of worlds $W$ iff the theory is true in all the worlds of $W$.

The semantical structure $M=(W,D,S^5,\omega_0)$ defined so far thus corresponds to a weak S5 structure $M'=(W+\omega_0, D, K45)$ for a monotonic modal logic of belief. This is because $W$ is a complete S5 structure, and all worlds in $W$ are accessible from $\omega_0$, hence the union of $W$ and $\omega_0$ forms a weak S5 structure. The question now is how the nonmonotonic part of AE logic comes into the possible worlds semantics. This is achieved by defining possible worlds interpretations of AE theory at a meta-level that would reflect the nonmonotonic part or reflective nature of the AE logic.

Definition 4 A Kripke structure $M=(W,D,S^5,\omega_0)$ is a possible worlds interpretation of an AE theory $T$ iff $T$ consists of all the formulae that are satisfied by $W$ in $M$.

Definition 5 A possible world model of an AE theory $T$ is a possible worlds interpretation of $T$ such that $T$ is true in $\omega_0$.

Definition 6 An AE theory is AE-satisfiable iff it has a possible world model.

It can be shown that the possible world models of an AE predicate theory are the possible world interpretations of the theory where $\omega_0$ has the same satisfiability relation as one of the possible worlds in $W$ in the Kripke structure $M=(W,D,S^5,\omega_0)$.

Theorem 1 If $M=(W,D,S^5,\omega_0)$ is a possible world interpretation of an AE theory $T$, then $M$ will be a possible world model of the theory iff $\omega_0 \in W$.

Proof: The if-case is obvious. If $\omega_0 \in W$, naturally all formulae in $T$ will be true in $\omega_0$. Hence $M$ is a possible world model of $T$. To prove the only-if-case, we suppose that $\omega_0$ differs from all the worlds in $W$. Then we can construct a disjunction of atoms in such a way that each atom is selected from each world in $W$ which differs from $\omega_0$. This disjunction is true in $W$ but false in $\omega_0$. Since $T$ contains this disjunction if $M$ is its possible world interpretation, thus not all formulae in $T$ will be true in $\omega_0$. Hence $M$ cannot be a possible world model of $T$.

Our semantics can be compared with Levesque's (also reformulated in Reiter [88]) AE semantics of a FOL KB. In his formalization, an autoepistemic model of an autoepistemic theory of a FOL KB consists of a S5 world structure $W$ which is formed of all the models of the sentences of the KB. The satisfiability relation between a possible world and an AE sentence is defined in the same way as our semantics. The following theorem establishes the relationship of Levesque’s semantics and our semantics.

Theorem 2 In the case of a FOL KB, Levesque’s semantics $M=(W,D,S^5)$ of FOL KB is equivalent to our semantics $M'=(W,D,S^5,\omega_0)$ of the KB in which $\omega_0$ is an element of $W$. 
Proof: Since all the FOL consequence of KB is true in all the possible worlds of W(or models of FOL KB) in a Levesque’s possible world model, for these consequences to be true in our corresponding possible world model, \( w_0 \) must be a model of FOL KB, ie. an element of W.

Now a traditional AE logical approach usually defines an AE extension T of a theory KB self-referentially in a non-constructive fixpoint as follows:

\[
T = \{ \phi \mid (KB + \{Bp \mid p \in T\} + \{\neg Bp \mid \neg(p \in T)\}) \models \phi \}.
\]

We can avoid this self-referential problem by defining AE extension in a possible world semantics.

Definition 7 T is an AE extension of a theory KB iff there exists a complete S5 structure \((W,D,S5)\) such that, for every possible world interpretation \(M=(W,D,S5,w_0)\) of T that has KB true in W, KB is true in \(w_0\) iff \(M\) is a possible world model of \(T\) (ie. \(w_0 \in W\)).

This definition can be shown to be equivalent to the traditional definition [Jiang 89].

Theorem 3 The possible world definition is equivalent to the self-referential definition.

Correspondingly we can also define completeness and soundness in the possible worlds semantics that equivalent [Jiang 89] match their counterparts (ie. stableness and groundness) in the self-referential semantics.

Definition 8 An AE theory T is semantically complete with respect to a set of premises KB iff there exists a possible world interpretation \(M=(W,D,S5,w_0)\) of T in which KB is true in W.

Definition 9 An AE theory T is semantically sound with respect to a set of premises KB iff there exists a complete S5 structure of worlds \((W,D,S5)\) such that for all \(w_0\) in all AE Kripke structures of the form \(M=(W,D,S5,w_0)\) such that KB and T are true in W, KB and T are true in \(w_0\) iff \(w_0 \in W\) of M.

It can be easily seen that if T consists of all formulae that are true in all the worlds of W in the soundness definition, then T is also semantically complete. We thus have the following obvious theorem:

Theorem 4 T is an AE extension of a set of premises KB iff T is semantically complete and sound with respect to KB.

Another obvious theorem is:

Theorem 5 An AE theory T that contains KB is AE-satisfiable if T is an AE extension of a KB; The converse however is not true (eg. let KB={Bp}, then T={p,Bp,...} is AE-satisfiable and contains KB, but it is not an AE extension of KB).

3 AE closure, Closed World Assumption and Complete database

Given the above semantical characterization, queries to a knowledge base KB simply become formulae to be checked to be members of an AE extension of the KB and integrity constraints simply become members of such an AE extension. In this section, we will compare and contrast the approach of using an AE extension as a closure of KB with two notable closure functions in database semantic: Reiter’s Closed World Assumption [Reiter 78] and Clark’s Complete database [Clark 78]. Both functions are also intended to form a total closure on a partially complete database. The intuitive effect is that facts about the world that are not known to be positively true, are assumed to be false.

Theorem 6 An AE extension of a KB is syntax independent.

Our first observation is that unlike Completion of database which is syntax-dependent, an AE closure is not. For example, the completion of the database \(\{p\rightarrow q, p\rightarrow p\}\) is quite different from the completion of the database \(\{p\rightarrow q\}\) although both databases are FOL semantically equivalent. In contrast, both sets of sentences would have the same AE extension.

The second observation is that unlike CWA or Comp, an AE closure will not affect the consistency property of a FOL KB. Formally, we have the following theorems.
Theorem 7 Assume KB is a FOL knowledge base, if CWA(KB) is FOL-satisfiable, then KB is FOL satisfiable; the converse, however, is not true.

Theorem 8 Assume KB is a FOL knowledge base, if Comp(KB) is FOL-satisfiable, then KB is FOL satisfiable; the converse, however, is not true.

Theorem 9 Assume KB is a FOL knowledge base, KB is FOL-satisfiable iff there exists an AE extension of KB (which implies that the extension is AE-satisfiable).

Proof:
1. If KB is satisfiable, then construct a model structure $M=(W,D,S,S_0,tv_0)$ such that $W$ is formed of all the models of KB and $w_0$ is a world in $W$. Let $T$ be all the formulae that are true in $W$. $T$ is an AE extension of KB.

2. If $T$ is an AE extension of KB, then in any possible world model $M=(W,D,S,S_0,w_0)$ of $T$, KB will be satisfied in $W$ (and $w_0$) as well. Now assuming KB is not FOL-satisfiable, then $W$ must be empty. But according to the definition of an AE extension, $W$ must not be empty, ie. a contradiction.

Although Lifschitz's [88] circumscriptive view of a KB also preserves the consistency property of a FOL KB, however it is not powerful enough to represent disjunctive incomplete knowledge and ignorance. For example, they can neither model $B(p \lor q)$ nor $\neg B(p \lor q)$.

Following Theorem 7, Theorem 8 and Theorem 9, we immediately have the following relationships between CWA, Comp and AE.

Lemma 1 Assume KB is a FOL knowledge base, then if CWA(KB) is FOL-satisfiable, then KB has an AE extension, hence AE-satisfiable; the converse, however, is not true.

Lemma 2 Assume KB is a FOL knowledge base, then if Comp(KB) is FOL-satisfiable, then KB has an AE extension, hence AE-satisfiable; the converse, however, is not true.

However it should be noted that an AE closure of a non-FOL KB may not be AE-satisfiable. Or more precisely, such a KB may not have an AE extension at all. For example, given $Ba \lor Bb$, the AE closure will become AE-unsatisfiable or empty. In addition, unlike FOL KB, there can be multiple AE extensions of a non-FOL KB in which a formula is true in one extension and false in another. These problems will be addressed in Section 5 where we introduce a restricted AE logic to ensure that an AE closure of a KB is always unique and non-empty.

However for satisfiable FOL KB, we can have the uniqueness property. For this reason, we will use AE(KB) to indicate the unique closure of such a KB [Jiang 89].

Theorem 10 There exists a unique AE extension for the AE closure of a satisfiable FOL KB.

Now it is well-known that CWA and Comp closures of a KB are both asymmetric assumptions about the incomplete information in the KB in the sense that only positive ground atoms are considered for candidates of negative conclusions. However often in KB applications, we also assume unspecified information to be true. For example, we generally regard that every one lives somewhere unless otherwise stated. The point is that we have a no-win situation for any asymmetric treatment. On the other hand, an AE closure would provide a symmetric treatment of assumption. If $\phi$ is not known in an AE extension of a KB, then $\neg B\phi$ is known in the extension. Here $\phi$ can be either positive or negative atoms. In fact, it can be any formulae.

Another advantage of the AE treatment is that it can also represent negative information explicitly. This is not possible for the asymmetric treatment of assumption in CWA and Complete database. If negation is introduced, it usually appears in the form of "negation as failure" which makes the negation non-logical. In particular, it is not clear a "Yes" answer to a negative atomic query means "the query is proven to be negative" or "the query is not proven to be positive". In contrast, all the connectives are logical in our AE closure of KB.

It is well known that CWA is generally incompatible with Comp. This can be demonstrated by the following FOL database $\{p \Leftarrow q, q \Leftarrow \neg p, q \Leftarrow q\}$. Although both CWA and Comp of the database is consistent, the union of the two closure is not. It appears that CWA and Comp closures may be compatible to AE closure since AE may bring out the possible incomparability outside FOL into epistemic level. This however in fact is not possible as shown below.
Theorem 11 Assume KB is a FOL database. Even if CWA(KB) is FOL satisfiable, CWA(KB) + AE(KB) may still not be AE-satisfiable.

Proof: This is because AE closure is symmetric. So if both p and ~p are not followed from a FOL database KB, CWA(KB) would have ~p true (when p is an atom); while AE(KB) would have ~B~p (and ~Bp true) which is contradicting to the CWA closure.

Theorem 12 Assume a FOL database KB. Even if Comp(KB) is FOL satisfiable, Comp(KB) + AE(KB) may still not be AE satisfiable.

Proof: Similar to Theorem 11.

4 AE closure for integrity constraints

Introducing epistemic notions into knowledge base can also open a new era of research in Integrity Constraints (IC) [Reiter 88, Jiang 88]. Integrity constraints (IC) are meant to characterize the acceptable states of a Knowledge Base (KB) through incremental changes. In this section, we will argue, contrary to the prevailing view, that integrity constraints are autoepistemic in nature.

Traditionally, integrity constraints are usually formalized under the following two definitions.

Definition 10 Satisfiability (eg. [Sadri & Kowalski 87])

KB satisfies IC iff Closure(KB) + IC is satisfiable.

Entailment (eg. [Lloyd & Topor 85])

KB satisfies IC iff Closure(KB) |= IC

The Closure concept is crucial here because it reflects the indexical nature of of integrity constraints satisfaction (This point is overlooked in Reiter's criticism of the two definitions). It generates a kind of complete extension of a KB based on the current context of KB.

Our formalization of integrity constraint is to identify the closure concept in the above definitions with an AE extension of a KB in our AE logic. In this case however, the standard concept of an integrity rule of the form p→q must be represented in AE logic as Bp→Bq (or equivalently B(p→q). If there is any NAF notions of the form not(p), then it will also be translated into ~Bp in AE logic. We denote these translations as IC AE. The translations are necessary because the original formulae mention nothing explicitly about the current context although it is what it is intended. We may wonder why such a translation is not needed for CWA and Complete closures functions that are normally chosen for the two definitions. This is because these closures are based on FOL reflections of KB. As a result, the conclusions at epistemic level (eg. not provable(p)) are reduced to FOL level (eg. ~p).

However it should be noted that the IC AE is only defined for IC rules of standard concept in FOL formalization of a KB. In the context of an AE formalization of KB, we do not need such a translation function. IC simply denotes a set of AE formulae. Thus we are not only able to represent AE formulae of the form Bp→Bq (which replaces the standard FOL IC notion) where p,q are in FOL, but also of any form of AE predicate logic. In particular, we can represent quantifying-in IC rules that have no analogs in standard concept of FOL IC rules. For example, the standard notion of IC can say something like “for x to be a person, he must live somewhere”; while the AE notion of IC can additionally say something like “for x known to be a person, the KB must currently know where he lives”.

We can now compare and contrast our AE formalization of integrity constraints with alternative formalizations. For this purpose, we first show that consistent CWA of a FOL database has the nice categorical property of a unique model [Reiter 87].

Theorem 13 FOL satisfiable CWA of a FOL KB is categorical.

Lemma 3 FOL satisfiable CWA of a FOL KB entails either p or ¬p for any FOL formula p.

Due to the complexity of AE logic, Reiter [88] introduces a way of combining AE closure with CWA in such a way that the satisfaction of AE integrity constraints become the satisfaction of FOL integrity constraints. This is shown by the following theorem:
Theorem 14  Assume KB is FOL, \( AE(CWA(KB)) \models IC^{AE} \) iff \( CWA(KB) \models IC \)
where \( KB \) and \( IC \) are sets of FOL sentences.

Proof: The reason for this to happen is that CWA(KB) is catagorical in the sense that it has at most one model. Thus in the case that there is no model, the theorem is trivally true. In the case that there is only one model, then this model must be the only possible world in an AE model of AE((CWA(KB)). Thus, we always have \( \forall x p(x) \leftrightarrow Bp(x) \) for any FOL formula \( p \) with free variables \( x \).

However Reiter's approach is too coarse-grain to distinguish the definite terms from indefinite terms. For example, given \( p(a) \lor p(b) \), in our AE formalization of integrity constraints we would not have \( \exists z Bp(z) \) true although \( \exists z p(z) \) is true in FOL. However, they become indistinguishable when the closure of a knowledge base is CWA followed by an AE closure.

It is also interesting to note that under the AE closure for a consistent FOL KB, the two IC definitions mentioned earlier become equivalent for standard IC rules. This is however not true for other closures. The main reason for this is that AE closure preserves consistency and yields a unique AE extension for FOL KBs.

Theorem 15  Let KB be a set of FOL formulae and IC a set of FOL formulae of standard notion, then the entailment definition of IC and the satisfiability definition of IC is equivalent iff CWA(KB) is consistent.

Proof: This is because consistent CWA(KB) is catagorical.

Theorem 16  Let KB be a set of FOL formulae and IC a set of FOL formulae of standard notion, then the entailment definition of IC and the satisfiability definition of IC will not be equivalent even if Comp(KB) is consistent.

Proof: This is because Comp(KB) is not catagorical.

Theorem 17  Let KB be a set of FOL formulae and IC a set of FOL formulae of standard notion, then the entailment definition of IC and the satisfiability definition of IC is equivalent under AE closure where IC becomes IC \( \models IC^{AE} \) iff KB is consistent.

Proof: 
1. Assume the two definitions are equivalent. Then if the KB is inconsistent, then clearly the satisfiability approach would invalid \( IC^{AE} \); while the the entailment approach would entail or validify anything. Hence the two definitions become inequivalent; ie. contradicting the assumption.

2. Assume KB is consistent. If the two definitions are not equivalent, then there must exist an IC such that \( AE(KB) \models IC^{AE} \) while \( AE(KB) + IC^{AE} \) is AE-unsatisfiable or \( \neg AE(KB) \models IC^{AE} \) while \( AE(KB) + IC^{AE} \) is AE-satisfiable. Since AE(KB) has a unique AE extension for consistent FOL KB according to Theorem 10, both cases would not be possible either.

Because of the lack of a unique AE extensions for a non-FOL KB, the two definitions will not be equivalent for the AE closure of a non-FOL KB. However as shown in the next section, a restricted kind of AE logic can ensure that there is a unique AE extension for a non-FOL KB. In this case, the two definitions would again become equivalent.

5  Stratified AE logic

Since the semantics of AE logic consists of a complete S5 structure together with an actual world, there appears to be a connection between a S5 modal logic and AE extensions. In this section, we thus attempt to establish such an connection proof theoretically.

We first note that the semantical structure of AE logic is a weak S5, thus the monotonic component of the proof theoretic connection is a proof theory for K45 epistemic logic.
Definition 11 Epistemic K45 proof theory consists of the following set of axioms and rules of inference:

the axioms are:

\[ \text{FOL} - \text{-axioms} \]
\[ B(\phi \rightarrow \psi) \& B\phi \rightarrow B\psi \]
\[ B\phi \rightarrow BB\phi \]
\[ \neg B\phi \rightarrow B\neg B\phi \]

and the rules of inference are:

\[ \phi, \phi \rightarrow \psi \Rightarrow \psi \]
\[ \phi \Rightarrow B\phi \]

Because of the non-monotonic nature of the AE logic, the deployment of the above monotonic K45 mechanization to the AE logic need to be augmented by indexical introspections. This however presents several difficulties. To start with, the negative introspection is itself not even semi-decidable [Niemela 88]. Furthermore, these introspections are part of an AE extension itself, thus involving a circular way of making assumptions followed by checking on soundness. For example, from \( Bp \rightarrow p \), an AE extension could assume \( p \) to be true and be justified through soundness.

We believe that the only way to solve the self-referential problem is to introduce some kind of stratification. This is essentially what Konolige’s Hierarchical AE (HAE) Predicate Logic [Konolige 88] attempted to do although it does not appear to be so obvious. By allowing B formulae to introspect only formulae at a lower level theory in a hierarchical structure, HAE essentially removes the self-referential part of an AE proof theory. However to achieve this, HAE requires an unnaturally indexed set of B operators rather than a single B operator. Our solution is to introduce an explicitly stratified AE predicate logic. First, we introduce two rules characterizing introspections. Instead of having a general negation as failure type of introspection, we adopt an Epistemic Negation As Failure (ENAF) rule of inference:

Epistemic NAF:

If \( \neg (A \vdash_{AE} p) \), then \( A \vdash_{AE} \neg Bp \)

where \( p \) is ordinary and AE is an autoepistemic proof system that includes ENAF, K45 and PI (defined later).

The ENAF rule is in fact similar to Clark’s NAF [Clark 78] except that it is made at an epistemic level and can be applied to non-grounded formulae as well. However like Clark’s NAF, ENAF is not-semidecidable for general logic due to the semi-decidability of FOL. In addition, the ENAF can still result an inconsistent AE closure at epistemic level. For example, given \( Bp \), the ENAF would produce \( \neg Bp \).

On the other hand, from an autoepistemic point of view, this kind of inconsistency would result an empty AE extension. Our solution is to restrict the AE logic to avoid an empty extension. This is a reasonable restriction for a rational agent such as KB. Since given some beliefs of an agent, the possibility of an empty AE extension appears to be intuitively contradictory. After all, the agent did have some beliefs before his reflection. We will show how to achieve the restriction later.

Next, to characterize the positive introspection, we introduce another rule of inference:

PI: If \( A \vdash_{AE} p \), then \( A \vdash_{AE} Bp \)

where \( p \) is ordinary and AE = K45 + ENAF + PI.

This is in fact the kind of rule of inference used in McDermott’s modal non-monotonic logics [82]. As pointed out by Moore [85], instead of treating the inference as part of a AE theory, ie. \( Bp \in T \) if \( p \in T \), the mere use of inference as a rule would result no AE extension of \( Bp \rightarrow p \) that contains \( p \). To avoid this incompleteness problem, we will restrict the AE logic to one that does not define \( p \) recursively through \( Bp \) as shown later.

The AE logic also suffers the problem of ‘theoremhood’ in the sense that there could be multiple AE extensions. Multiple extensions may be possible for human beings. They however do not seem to be
reasonable in many practical Knowledge Base (KB) applications. We cannot afford to have $p$ and $\neg\Phi$ at the same time from a KB due to its multiple extensions. Nor can we afford to find all AE extensions (which is possible for AE propositional logic) as indicated in [Gelfond 88] because this could be infinite for AE predicate logic.

To avoid (not solve) all these problems, we define a stratified AE predicate logic. We extend Gelfond's [87] definition for propositional logic to predicate logic. The intuition is that we want to avoid to define recursively a predicate through its introspection (positive or negative). We will also indicate how stratification would restrict certain kinds of formulae that cause problems in our AE proof system.

**Definition 12**  
An AE theory $T$ is stratified if

1. $T$ consists of clauses of the form $S \rightarrow V$. $V$ is a disjunction of FOL atoms and $S$ is a conjunction of \( B \) literals and FOL atoms. Each \( B \)-literal is not allowed to be nested. Hence the introspection rules only need to introspect ordinary FOL formulae.
2. There is a partition $T=T_0+\ldots+T_n$ such that
3. $T_0$ is ordinary (possibly empty).
4. clauses with empty conclusions do not belong to $T_k$ where $k>0$.

Thus the case $\Phi \rightarrow \neg\Phi$; ie. $\neg\Phi \rightarrow \neg\Phi$ that would cause inconsistency of ENAF would not occur.

5. If a predicate $p$ belongs to the conclusion of a clause in $T_k$, then (positive/negative) literals with predicate $p$ do not belong to $T_0,\ldots,T_{k-1}$ and where literals $B\Phi$ and $\neg B\Phi$ where $f$ contains $p$, do not belong to $T_0,\ldots,T_k$.

Thus the case $B\Phi \rightarrow p$ and $\neg B\Phi \rightarrow \neg p$ would not occur.

Now Gelfond [87] has shown that every consistent stratified AE propositional theory has a unique AE extension. In [Jiang 89], we have shown that Herbrand theorem "a set of universal sentences is unsatisfiable iff a finite subset of its ground instances are" remains to be valid for epistemic logic with the introduction of level of intension. Thus with unification and Gelfond's propositional result, we can have the following theorem:

**Theorem 18**  
Every consistent set of stratified AE predicate formulae has a unique AE extension.

If we use $K45+ENAF+PI$ to characterize the proof mechanization of our stratified AE predicate logic, we first can show the following soundness theorem.

**Theorem 19**  
If $KB$ is a consistently stratified AE theory, then there exists a unique $AE(KB)$ such that

If $KB \vdash K45+ENAF+PI \phi$, then $AE(KB) \models \phi$.

Since the stratified AE logic still allows ordinary FOL formulae in the AE logic, we thus cannot obtain full completeness result for our proof mechanization. Given the presence of function symbols in FOL, the ENAF and PI rules can only be semi-decidable. A consequence is that the stratified AE predicate logic itself is not even semi-decidable. Hence it is not complete either. However if we limit ourselves to function-free knowledge base which is generally the case for deductive database applications, we can obtain the following completeness result:

**Theorem 20**  
If $KB$ is a consistently stratified and function-free AE theory, then there exists a unique $AE(KB)$ such that

If $AE(KB) \models \phi$, then $KB \vdash K45+ENAF+PI \phi$.

We illustrate our proof theory on the following stratified AE KB (assuming the unique name assumption).

Query: Student(Tom)?
Knowledge base:

1. teacher(John). "John is a teacher"
2. teacher(father(John)). "Father of John is a teacher"
3. \(-B(teacher(x)) \rightarrow student(x)\). "If x is not currently known to be a teacher, then x is a student"

From 3, we have:

4. \(-B(teacher(Tom)) \rightarrow student(Tom)\)

Apply ENAF rule to the initial knowledge base, yield:

5. \(-B(teacher(Tom))\)

Apply modus ponens between 4 and 5, yield student(Tom).

Note if we add \(father(John) = Tom\), then 5 cannot be obtained.

Despite the catagorical property of stratified AE predicate logic which helps resolve the constructive proof theoretical problem of AE logic, it is still too restrictive for KB applications. For examples, the two quantifying-in formulae mentioned in Section 2 would fall outside the scope of our stratification. On-going research is to extend the boundary of AE logic and at the same time maintain the constructive nature of AE logic.

Acknowledgements

The author would like to thank Professor R. Reiter, Professor J. Lloyd for their helpful discussions. This research is supported by a British Telecom Fellowship.

References

I. Niemela *Autoepistemic Predicate logic* ECAI 88, pp.595-599.