# Autoepistemic Logic of First Order and Its Expressive Power *

Li Yan Yuan
Department of Computing Science
The University of Alberta
Edmonton, CANADA T6G 2H1
yuan@cs.ualberta.ca

**Abstract**

We study the expressive power of first order autoepistemic logic [1] . We argue that full introspection of rational agents should be carried out by minimizing positive introspection and maximizing negative introspection. Based on full introspection, we propose the generalized stable semantics that characterizes autoepistemic reasoning processes of rational agents, and show that the breadth of the semantics covers all theories in autoepistemic logic of first order, Moore's AE logic, and Reiter's default logic.

Our study demonstrates that autoepistemic logic of first order is a very powerful framework for nonmonotonic reasoning, logic programming, deductive databases, and knowledge representation.

**Key Words:** nonmonotonic reasoning, autoepistemic logic, default logic, logic programming, knowledge representation

## 1 Introduction

An idea rational agent has to decide which set of propositions to believe according to her knowledge. Moore's AE logic is a powerful framework for this kind of introspective reasoning which means that an agent is capable of reasoning not only about the world, but also of reasoning about its own knowledge and beliefs about the world [16]. Introspective reasoning

*This paper is available as Technical Report TR93-17 at the Department of Computing Science, University of Alberta, Canada.

[1] By first order autoepistemic logic, we mean a first order logic without modal operators, not a modal logic with quantifying-in

by rational agents is characterized by the concept of a *stable expansion*. By a stable expansion of an AE theory $E$ we mean a set $T$ of formulas such that

$$T = \{\phi | E \cup \{\mathcal{K}\psi | \psi \in T\} \cup \{\neg\mathcal{K}\psi | \psi \notin T\} \models_S \phi\} \tag{1}$$

where $S$ is the propositional calculus in the modal language [16]. The terms $\{\mathcal{K}\phi | \phi \in T\}$ and $\{\neg\mathcal{K}\phi | \phi \notin T\}$ express the positive and negative introspection of an agent respectively.

However, Moore's logic has some serious limitations. First, since the same approach is applied to both positive and negative introspection, Moore's logic has difficulty in accounting precisely for the mechanism used in Reiter's default logic [11, 15]. For example, $E = \{\mathcal{K}p \supset p\}$ has a stable expansion containing $p$, which seems to be quite counterintuitive. In order to get rid of such ungrounded expansions, various notions of grounded stable expansions have been proposed [11, 23]. However, as indicated by Schwarz [22], due to inappropriate positive introspection, those notions of grounded expansions are logically unsatisfactory in that an ungrounded expansion may become grounded by introducing some trivial denotations [2] without really modifying the system. Since such difficulties are caused by inappropriate positive introspection, we believe positive and negative introspection must be dealt with differently in autoepistemic reasoning processes.

Another problem associated with Moore's logic as well as Reiter's default logic is the so called *semantical partial-ness*, that is, the breadth of the semantical definitions does not cover all logic theories. It is a well known fact that many AE sets have no stable expansions and default theories may not always have extensions. Similar problems exist with the stable semantics for logic programming proposed by Gelfond and Lifschitz [8]. The well-founded semantics for logic programming, proposed by Van Gelder, Ross, and Schlipf [5], does cover all logic programs, but it may not be able to characterize the intended meaning of some programs since the negative introspection is not fully realized [2, 3].

Many attempts have been made to establish full semantics [3] for various formalisms in that every reasonable theory should have a consistent set of theorems under the given semantics [1, 2, 3, 9, 15, 18, 24, 25]. Among all proposals, the regular model semantics for logic programs, first proposed by You and Yuan [24, 26] and then redefined as the preferential semantics by Dung [3], and the stable class semantics for logic programs, AE logic, and default logic, proposed by Baral and Subrahmanian [1, 2], have attracted considerable attentions. The regular model semantics provides a satisfactory semantics for logic programs, but can not be directly applied to more general nonmonotonic frameworks because of the absence of positive introspection in logic programming. The stable class semantics is defined for various formalisms, however we have found out that sometimes, it may lead to counterintuitive conclusions, as demonstrated in Example 5.4. Note that Baral and Subrahmanian have suggested that different semantics may be obtained by selecting other stable classes instead of the choices they made in [1, 2], and following this suggestion, we have found out that the above problem can be avoided if non-normal stable classes are eliminated.

[2] By a denotation of an atom, say $\mathcal{K}p$, we mean another atom $q$ such that $q \equiv \mathcal{K}p$.

[3] By full semantics we mean semantics that defines a consistent set of theorems for any reasonable theory in the concerned logic.

Perlis and Lifschitz have initiated study of first order autoepistemic logic [12, 17]. First order autoepistemic logic is a first order logic with a set of belief predicates $\mathcal{L}P$ standing for "believing in $P$", and therefore, combines the advantages of Moore's logic with those of first order logic. Przymusinski, Yuan and You have extended the well-founded semantics of logic programming into the context of autoepistemic logic of first order [18, 25]. However, similar to its counterpart for logic programming, the well-founded semantics for autoepistemic logic does not characterize full introspection. Besides, it can not be extended to other nonmonotonic formalisms due to its limitation on belief-levels.

In this paper, we argue that the autoepistemic reasoning process of rational agents should be characterized by distinguishing positive and negative introspection. The basic idea is simple but very effective: **positive introspection should be minimized while negative introspection should be maximized**. Based on these two principles, we define the generalized stable semantics for autoepistemic logic of first order. Since the generalized stable semantics is based on full introspection it defines a consistent set of theorems for any reasonable autoepistemic theories.

We have shown that both Moore's logic and Reiter's logic can be represented by autoepistemic logic of first order. We also demonstrate that in the context of logic programs, the generalized stable semantics coincides with the regular model semantics, though the later does not endorse the principle of minimizing positive introspection.

Our results demonstrate that autoepistemic logic of first order is an attractive nonmonotonic reasoning formalism. It combines explicit belief representation of Moore's logic with the formal reasoning approach of first order logic.

## 2 Autoepistemic Logic of First Order

In this section, we present the basic notations of first order autoepistemic logic. Following other researchers, we restrict all of our discussions to the language of *propositional logic*, unless otherwise indicated.

By a *propositional autoepistemic language* we mean a propositional language $\mathcal{P}$ with two distinct sets of predicates $P$ and $\mathcal{L}P$, where $P$ is the set of *objective predicates* and $\mathcal{L}P$ the set of *belief predicates* such that for each belief predicate $\mathcal{L}p$ in $\mathcal{L}P$, there exists an objective predicate $p$ in $P$ and the intended meaning of $\mathcal{L}p$ is "believing in $p$". An *autoepistemic formula* is a well formed formula in an autoepistemic language and an *autoepistemic theory* is a set of autoepistemic formulas. A theory may also be viewed as the conjunction of all formulas in the theory. (Note that (1) an objective predicate may not always have a belief counterpart but each belief predicate must have an objective counterpart; and (2) $\mathcal{L}p$ is a predicate with a special name to indicate its intended meaning and should not be viewed as applying some *modal operator* $\mathcal{L}$ to $p$.)

Let $A(P, \mathcal{L}P)$ be a logic theory. A model of $A$ is a set of atoms such that all formulas in $A$ are true in the set, according to the usual propositional evaluation. $A$ is consistent if $A$ has at least one model. By $A \models \phi$ we mean $\phi$ is a logical consequence of $A$. A $P$-interpretation (resp. $\mathcal{L}P$-interpretation) of $A$ is a set of atoms whose predicates are in $P$ (resp. $\mathcal{L}P$). For convenience, we abuse notation in the following way: an $\mathcal{L}P$-interpretation is also called a *belief* when no confusion arises.

# 3 Introspection and Beliefs

In this section, we characterize full introspection in terms of generalized stable beliefs.

We are convinced that the intended meaning of an autoepistemic theory should be determined by appropriate beliefs while appropriateness of beliefs should be determined based on introspective reasoning.

Let $A(P, \mathcal{L}P)$ be an autoepistemic theory and $\mathbf{m}_{\mathcal{L}}$ be a belief representing truth-assignments of belief predicates. We use $N(\mathbf{m}_{\mathcal{L}}) = \{\neg\mathcal{L}a | \mathcal{L}a \notin \mathbf{m}_{\mathcal{L}}\}$ and $P(\mathbf{m}_{\mathcal{L}})$ to represent the negative and positive introspection under $\mathbf{m}_{\mathcal{L}}$ [4] respectively.

First, we discuss some necessary restrictions upon introspection.

By the nature of introspection, $P(\mathbf{m}_{\mathcal{L}})$ and $N(\mathbf{m}_{\mathcal{L}})$ should satisfy the following two equations.

$$P(\mathbf{m}_{\mathcal{L}}) = \{\mathcal{L}a \mid A \cup P(\mathbf{m}_{\mathcal{L}}) \cup N(\mathbf{m}_{\mathcal{L}}) \models a\} \tag{2}$$

$$N(\mathbf{m}_{\mathcal{L}}) \supseteq \{\neg\mathcal{L}a \mid A \cup P(\mathbf{m}_{\mathcal{L}}) \cup N(\mathbf{m}_{\mathcal{L}}) \models \neg a\}$$

In an idea situation, a rational agent shall believe anything is either true or false. Thus, we say the introspection under $\mathbf{m}_{\mathcal{L}}$ is *rigorous* if

$$P(\mathbf{m}_{\mathcal{L}}) = \mathbf{m}_{\mathcal{L}}$$

The rigorous introspection assigns a truth value to each and every belief predicate.

However, we have to consider many reasonable applications in which the rigorous introspection can not be achieved. According to the negation as failure rule, $\neg\mathcal{L}a$ can be assumed only if $a$ cannot be possibly true, even if all $\mathcal{L}a$ not in $P(\mathbf{m}_{\mathcal{L}})$ are false. Consequently, we say the introspection under $\mathbf{m}_{\mathcal{L}}$ is *negatively sound* if

$$N(\mathbf{m}_{\mathcal{L}}) = \{\neg\mathcal{L}a \mid A \cup P(\mathbf{m}_{\mathcal{L}}) \cup \{\neg\mathcal{L}a | \mathcal{L}a \notin P(\mathbf{m}_{\mathcal{L}})\} \not\models a\} \tag{3}$$

Naturally, the positive and negative introspection should not overlap. Hence, the introspection under $\mathbf{m}_{\mathcal{L}}$ is said to be *normal* if

$$P(\mathbf{m}_{\mathcal{L}}) \cap N(\mathbf{m}_{\mathcal{L}}) = \emptyset$$

Obvious, rigorous introspection is both negatively sound and normal, but not vice versa. We believe that the normality and negative soundness are necessary restrictions on introspection.

[4] Using $N(\mathbf{m}_{\mathcal{L}})$ to denote the negative introspection is merely for convenience. The results will be the same if we use $\mathbf{m}_{\mathcal{L}}$ for the positive introspection, and redefine $N(\mathbf{m}_{\mathcal{L}})$ accordingly.

### 3.1 minimizing positive introspection

The *groundedness* defined below is to characterize minimized positive introspection.

Let $A$ be an autoepistemic theory and $N$ be a set of negative belief literals. Then $A \uplus N$ is used to denote an autoepistemic theory obtained from $A$ by replacing each **positive occurrence** $\mathcal{L}a$ in $A$ with `false` if $\neg\mathcal{L}a \in N$.

**Example 3.1** Let $A = \{a \vee \mathcal{L}a; b \vee \neg\mathcal{L}a \vee \mathcal{L}c\}$ and $N = \{\neg\mathcal{L}a\}$.
Then $A \uplus N = \{a; b \vee \neg\mathcal{L}a \vee \mathcal{L}c\}$. □

**Definition 3.1** Let $A(P, \mathcal{L}P)$ be an autoepistemic theory and $\mathbf{m}_\mathcal{L}$ a belief. We say an atom $a$ is *grounded* with respect to $\mathbf{m}_\mathcal{L}$ if either

1. $A \uplus N(\mathbf{m}_\mathcal{L}) \models a$, or recursively

2. there exist grounded atoms $a_1, \ldots, a_n$ with respect to $\mathbf{m}_\mathcal{L}$ such that

$$(A \uplus N(\mathbf{m}_\mathcal{L})) \cup \{\mathcal{L}a_1 \cdots \wedge \mathcal{L}a_n\} \models a.$$

Let $N$ denote the necessitation rule $\vdash a \rightarrow \vdash \mathcal{L}a$, and $\models_N$ propositional calculus augmented with $N$ [10]. Then $a$ is grounded with respect to $\mathbf{m}_\mathcal{L}$ if and only if $A \uplus N(\mathbf{m}_\mathcal{L}) \models_N a$. □

Note that (1) if $A$ contains no negative occurrences of belief atoms then $A \uplus N(\mathbf{m}_\mathcal{L}) \models_N a$ if and only if $A \cup N(\mathbf{m}_\mathcal{L}) \models a$; and (2) if $P_A(\mathbf{m}_\mathcal{L}) \subseteq \mathbf{m}_\mathcal{L}$ then $A \uplus N(\mathbf{m}_\mathcal{L}) \models_N a$ if and only if $A \cup N(\mathbf{m}_\mathcal{L}) \models_N a$.

Now we define a very important transformation, based on the concept of the grounded-ness, which is a generalization of the Gelfond-Lifschitz transformation defined in [8].

**Definition 3.2** Let $A(P, \mathcal{L}P)$ be an autoepistemic theory. Then the transformation $P_A$, which maps sets of beliefs to sets of beliefs, is defined as follows.

$$P_A(\mathbf{m}_\mathcal{L}) = \{\mathcal{L}a \mid A \uplus N(\mathbf{m}_\mathcal{L}) \models_N a\}$$ □

$P_A(\mathbf{m}_\mathcal{L})$, which satisfies Equation (2), represents the minimized positive introspection under $\mathbf{m}_\mathcal{L}$. Therefore, the introspection under $\mathbf{m}_\mathcal{L}$ is *positively minimized* if $P(\mathbf{m}_\mathcal{L}) = P_A(\mathbf{m}_\mathcal{L})$.

**Theorem 3.1** $P_A$ is anti-monotonic, that is,

$$P_A(\mathbf{m}_{\mathcal{L}_1}) \subseteq P_A(\mathbf{m}_{\mathcal{L}_2}) \text{ if } \mathbf{m}_{\mathcal{L}_2} \subseteq \mathbf{m}_{\mathcal{L}_1}$$

*Proof:* This follows from a simple fact that $A \uplus N(\mathbf{m}_{\mathcal{L}_2}) \models A \uplus N(\mathbf{m}_{\mathcal{L}_1})$ if $\mathbf{m}_{\mathcal{L}_2} \subseteq \mathbf{m}_{\mathcal{L}_1}$. □

## 3.2 maximizing negative introspection

First, we present the following definition.

**Definition 3.3** Let $A$ be an autoepistemic theory and $\mathbf{m}_{\mathcal{L}}$ a belief. Then $\mathbf{m}_{\mathcal{L}}$ is called

1. a *fixpoint* of $A$ if $\mathbf{m}_{\mathcal{L}} = P_A(\mathbf{m}_{\mathcal{L}})$;
2. an *alternating fixpoint* of $A$ if $\mathbf{m}_{\mathcal{L}} = {P_A}^2(\mathbf{m}_{\mathcal{L}}) = P_A(P_A(\mathbf{m}_{\mathcal{L}}))$. [5]
3. a *normal alternating fixpoint* if (1) $\mathbf{m}_{\mathcal{L}} = P_A(P_A(\mathbf{m}_{\mathcal{L}}))$ and (2) $P_A(\mathbf{m}_{\mathcal{L}}) \subseteq \mathbf{m}_{\mathcal{L}}$.

Let $F_{\mathbf{m}_{\mathcal{L}}} = P_A(\mathbf{m}_{\mathcal{L}}) \cup N(\mathbf{m}_{\mathcal{L}})$ denote the introspection under $\mathbf{m}_{\mathcal{L}}$. Then it is straightforward to show that

1. $F_{\mathbf{m}_{\mathcal{L}}}$ is rigorous and positively minimized if and only if $\mathbf{m}_{\mathcal{L}}$ is a fixpoint of $A$;
2. $F_{\mathbf{m}_{\mathcal{L}}}$ is negatively sound and positively minimized if and only if $\mathbf{m}_{\mathcal{L}}$ is an alternating fixpoint of $A$; and
3. $F_{\mathbf{m}_{\mathcal{L}}}$ is negatively sound, normal, and positively minimized if and only if $\mathbf{m}_{\mathcal{L}}$ is a normal alternating fixpoint of $A$.

Note the introspection under an alternating fixpoint may not be normal.

Furthermore, the negative introspection should be maximized otherwise little objective conclusion may be derived. Therefore, we define the following.

**Definition 3.4** Let $A(P, \mathcal{L}P)$ be an autoepistemic theory and $\mathbf{m}_{\mathcal{L}}$ a belief. Then $\mathbf{m}_{\mathcal{L}}$ is said to be a *generalized stable belief* of $A$ if

1. $\mathbf{m}_{\mathcal{L}}$ is a normal alternating fixpoint of $A$, and
2. there exists no normal alternating fixpoint $\mathbf{n}_{\mathcal{L}}$ of $A$ such that $\mathbf{n}_{\mathcal{L}} \subset \mathbf{m}_{\mathcal{L}}$. □

The introspection under a generalized stable belief is positively minimized and negatively maximized. Therefore, full introspective reasoning by rational agents should be represented by generalized stable beliefs.

Let $\mathbf{m}_{\mathcal{L}_1}, \ldots, \mathbf{m}_{\mathcal{L}_n}$ be the set of all generalized stable beliefs of $A$. Then *the generalized stable semantics* of $A$ is defined by

$$A(P, \mathcal{L}P) \wedge (F_{\mathbf{m}_{\mathcal{L}1}} \vee \cdots \vee F_{\mathbf{m}_{\mathcal{L}n}})$$

The generalized stable semantics of an autoepistemic theory is defined using the so called *skeptical approach*, that is, the semantics is characterized by the theory and the disjunction of the introspections under all the generalized stable beliefs.

[5] The alternating fixpoint was first proposed by Van Gelder for representing the well-founded semantics for logic programs [4].

# 4 Analysis and Justification

Since introspection is usually represented by expansions, we first discuss relationships between various expansions and corresponding beliefs, and then characterize the generalized stable semantics in terms of expansions with full introspection. Finally we show that the generalized stable semantics defines a consistent set of theorems for any epistemically coherent theory.

Recall that an autoepistemic theory $T$ is a *stable expansion* of $A$ if $T$ satisfies the following introspective equation:

$$T = \{\phi \mid A \cup \{\mathcal{L}a|a \in T\} \cup \{\neg\mathcal{L}a|a \notin T\} \models \phi\} \tag{4}$$

Let $\mathbf{m}_\mathcal{L}(T) = \{\mathcal{L}a \mid \mathcal{L}a \in T\}$. Since $P_A(\mathbf{m}_\mathcal{L})$ represents the set of all atoms that are grounded with respect to $\mathbf{m}_\mathcal{L}$, a stable expansion $T$ is *grounded* if $\mathbf{m}_\mathcal{L}(T) \subseteq P_A(\mathbf{m}_\mathcal{L}(T))$. That is, $T$ is a grounded stable expansion of $A$ if $T$ satisfies

$$T = \{\phi \mid A \cup \{\neg\mathcal{L}a \mid a \notin T\} \models_N \phi\} \tag{5}$$

The following example shows that a minimal stable expansion (in terms of set inclusions) is not necessarily grounded. We use $Cn(T)$ to represent the set of all logical consequences of $T$.

**Example 4.1** Consider $A = \{a \subset \mathcal{L}a, b \subset a, a \subset \neg\mathcal{L}b\}$. Then $A$ has a unique stable expansion $T = Cn(A \cup \{\mathcal{L}a, \mathcal{L}b\})$. However, $T$ is not grounded since $P_A(\{\mathcal{L}a, \mathcal{L}b\}) = \emptyset$. It follows that $A$ has no grounded stable expansions. □

It is interesting to observe that the relationship between the ground stable expansion and the minimal stable expansion is analogous with that between the stable model and minimal model of logic programs. Note that a stable model is always a minimal model but not vice versa [8, 13, 14].

The stable expansion and the grounded stable expansion can also be defined in terms of appropriate beliefs. A belief $\mathbf{m}_\mathcal{L}$ is *stable* if $\mathbf{m}_\mathcal{L} = \{\mathcal{L}a \mid A \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L}) \models a\}$. Then $T$ is a stable expansion of $A$ if and only if there exists a stable belief $\mathbf{m}_\mathcal{L}$ such that $T = Cn(A \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L}))$. Furthermore,

**Theorem 4.1** Let $A$ be an autoepistemic theory. Then $T$ is a grounded stable expansion of $A$ if and only if there exists a fixpoint $\mathbf{m}_\mathcal{L}$ of $P_A$ such that

$$T = Cn(A \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L}))$$

*Proof:* It follows from the fact that for each belief $\mathbf{m}_\mathcal{L} = P_A(\mathbf{m}_\mathcal{L})$, $A \uplus N(\mathbf{m}_\mathcal{L}) \models_N a$ if and only if $A \cup N(\mathbf{m}_\mathcal{L}) \models_N a$ if and only if $A \cup N(\mathbf{m}_\mathcal{L}) \cup \mathbf{m}_\mathcal{L} \models a$. □

From the above discussion, we can see that the stable expansion is based on rigorous introspection while the grounded stable expansion is based on rigorous and positively minimized introspection.

Since $P_A$ is not monotonic, many reasonable autoepistemic theories may not always have grounded stable expansions. For example, $A = \{a \subset \neg\mathcal{L}a\}$ has no stable expansions at all. However, since $P_A$ is anti-monotonic, every autoepistemic theory has at least one alternating fixpoint. By generalizing Van Gelder's alternating fixpoint theory, we define *a well-founded belief* as a normal alternating fixpoint of $A$, and the well-founded expansion of $A$ as a theory $T$ that satisfies the following introspective equation.

$$T = \{\phi \mid A \cup N_T \models_N \phi\} \tag{6}$$

where $N_T = \{\neg\mathcal{L}p \mid A \cup \{\mathcal{L}a \mid a \in T\} \cup \{\neg\mathcal{L}a | a \notin T\} \not\models p\}$.
Therefore, $T$ is a well-founded expansion of $A$ if and only if there exists a well-founded belief $\mathbf{m}_\mathcal{L}$ such that $T = Cn(A \cup P_A(\mathbf{m}_\mathcal{L}) \cup N(\mathbf{m}_\mathcal{L}))$. The well-founded semantics provides a solution to the no-grounded-stable-expansion problem. However, the well-founded semantics does not preserve some desirable capacities of the grounded stable semantics. Consider the following example.

**Example 4.2** Consider $A = \{a \subset \neg\mathcal{L}b;\ \ b \subset \neg\mathcal{L}a;\ \ c \subset a;\ \ c \subset b\}$. Then $A$ has two fixpoints, viz. $\mathbf{m}_{\mathcal{L}_1} = \{\mathcal{L}a\}$, $\mathbf{m}_{\mathcal{L}_2} = \{\mathcal{L}b\}$, and three normal alternating fixpoints, viz. $\mathbf{m}_{\mathcal{L}_1}$, $\mathbf{m}_{\mathcal{L}_2}$, $\mathbf{m}_{\mathcal{L}_3} = \{\mathcal{L}a, \mathcal{L}b\}$.

Our intuition tells us that $c$ should be true in the intended meaning of $A$, which is exactly the value of $c$ assigned by the grounded stable (expansion) semantics. But $c$ is unknown in the well-founded (expansion) semantics of $A$. The only well-founded expansion of $A$ that does not contain $c$ is $Cn(A \cup P_A(\mathbf{m}_{\mathcal{L}_3}) \cup N(\mathbf{m}_{\mathcal{L}_3}))$. However, readers may find out that

$$Cn(A \cup P_A(\mathbf{m}_{\mathcal{L}_3}) \cup N(\mathbf{m}_{\mathcal{L}_3})) \subset Cn(A \cup P_A(\mathbf{m}_{\mathcal{L}_1}) \cup N(\mathbf{m}_{\mathcal{L}_1})).$$

□

This example motivates us to define the generalized stable expansion.

**Definition 4.1** $T$ is said to be *a generalized stable expansion* of $A$ if

1. $T$ is a well-founded expansion of $A$, and
2. there exists no well-founded expansion $T'$ of $A$ such that $T \subset T'$. □

Assume $T_1$ and $T_2$ are two well-founded expansions such that $T_1 \subset T_2$. Then $\{\neg\mathcal{L}a \mid \neg\mathcal{L}a \in T_1\} \subset \{\neg\mathcal{L}a \mid \neg\mathcal{L}a \in T_2\}$. Therefore, a generalized stable expansion of $A$ represents

> **the full introspection that is both positively minimized and negatively maximized.**

The following theorem shows that the generalized stable expansion is characterized by the generalized stable beliefs. The proof of the theorem is straightforward.

**Theorem 4.2** $T$ is a generalized stable expansion of an autoepistemic theory $A$ if and only if there exists a generalized stable belief $\mathbf{m}_{\mathcal{L}}$ such that

$$T = Cn(A \cup P_A(\mathbf{m}_{\mathcal{L}}) \cup N(\mathbf{m}_{\mathcal{L}}))$$

□

It is easy to check that a grounded stable expansion is a generalized stable expansion, and of course, a generalized stable expansion is a well-founded expansion, but not the vise versa. Among all semantics, the generalized stable semantics is the only one that captures the intuitive meaning of theories in the above examples.

**Remark:** It is easy to show that among all well-founded expansions of an autoepistemic theory, there exists a least well-founded expansion that is a subset of any well-founded expansions. The least well-founded expansion may be used to extend the well-founded semantics of logic programs into the context of autoepistemic theories [18, 25]. Since the generalized stable semantics is characterized by all maximal well-founded expansions, it may also be called the *maximal well-founded semantics.*

Since ${P^2}_A$ is monotonic, every autoepistemic theory has at least one normal alternating fixpoint, as shown below.

**Theorem 4.3** Any autoepistemic theory $A$ has at least one generalized stable belief.

*Proof:* It is sufficient to show that $A$ has at least one normal alternating fixpoint. Since ${P_A}^2$ is monotonic, its least and greatest fixpoints, denoted by lfp$({P_A}^2)$ and gfp$({P_A}^2)$ respectively, do exist, and ${P_A}^2(gfp({P_A}^2)) = lfp({P_A}^2)$. Let $\mathbf{m}_{\mathcal{L}} = gfp({P_A}^2)$, then $P_A(\mathbf{m}_{\mathcal{L}}) = lpf({P_A}^2) \subseteq \mathbf{m}_{\mathcal{L}}$. It follows that $\mathbf{m}_{\mathcal{L}}$ is a normal alternating fixpoint of $A$. □

The existence of generalized stable beliefs does not imply that the generalized stable semantics is always consistent.

**Example 4.3** Consider $A = \{a; \neg\mathcal{L}a\}$. Since $A \models a$, $A$ has a unique alternating fixpoint which is also its unique fixpoint, viz. $\mathbf{m}_{\mathcal{L}} = \{\mathcal{L}a\}$. Therefore $A$ has a unique generalized stable expansion that contains all sentences. □

An autoepistemic theory $A$ is said to be *epistemically coherent* if $A$ is *consistent* under the necessitation rule, that is, for any $a$ such that $A \models_N a$, $A \not\models_N \neg\mathcal{L}a$. $A$ in the above example is not epistemically coherent. If a theory is not epistemic ally coherent then its objective and belief predicates are not compatible in any circumstances, and therefore its behavior may not be very normal. The following theorem shows that the generalized stable semantics of a theory is consistent if and only if it is epistemic ally coherent.

**Theorem 4.4** A well-founded expansion of $A$ is consistent if and only if $A$ is epistemically coherent.

*Proof:* We show the if part for the only if part is trivial.

Assume $\mathbf{m}_{\mathcal{L}}$ is a normal alternating fixpoint of an autoepistemic theory $A$ and $A$ is epistemically coherent.

Assume $A \cup P_A(\mathbf{m}_{\mathcal{L}}) \cup N(\mathbf{m}_{\mathcal{L}})$ is inconsistent then $A \cup P_A(\mathbf{m}_{\mathcal{L}}) \cup N(\mathbf{m}_{\mathcal{L}}) \models a$ for any atom $a$. Thus, $P_A(\mathbf{m}_{\mathcal{L}}) = \mathcal{L}P$ and $N(\mathbf{m}_{\mathcal{L}}) = \emptyset$, for $P_A(\mathbf{m}_{\mathcal{L}}) \subseteq \mathbf{m}_{\mathcal{L}}$, which implies that $A \cup N(\mathbf{m}_{\mathcal{L}})$ is not epistemically coherent. This contradicts to the assumption that $A$ is epistemically coherent since $N(\mathbf{m}_{\mathcal{L}}) = \emptyset$. □

# 5 Comparisons with Various Formalisms

In this section, we compare the semantics of autoepistemic logic with various nonmonotonic formalisms and demonstrate the expressive power of autoepistemic logic by representing Moor's autoepistemic logic ( AE logic for short) and Reiter's default logic as autoepistemic theories.

In order to compare intended meanings of various theories we define the concept of *obj-equivalence.*

**Definition 5.1** Let $A_1$ and $A_2$ be two logic theories, and let $P$ be the set of all objective predicates appearing in both $A_1$ and $A_2$. Let $A_{1obj}$ and $A_{2obj}$ be the sets of all objective formulas, whose predicates are from $P$, that are logical consequences of $A_1$ and $A_2$ respectively. We say $A_1$ and $A_2$ are *obj-equivalent*, denoted as $A_1 \equiv_{obj} A_2$, if $A_{1obj} = A_{2obj}$. □

By the definition, $A_1$ and $A_2$ are *obj*-equivalent if and only if they entail the same set of objective formulas over the common objective predicates $P$.

## 5.1 Moore's AE-logic

Moore's AE-logic [16] is defined on a propositional language augmented by a special modal operator $\mathcal{K}$. The semantics of an AE-logic theory is characterized by the AE-stable expansions [11, 16].

Now we present a translation to represent each AE set as an autoepistemic theory.

**Definition 5.2** Let $E$ be an AE set. Then $AT_E(P, \mathcal{L}P)$, called *the autoepistemic translation* of $E$, is an autoepistemic theory obtained from $E$ using the following procedure.

1. $AT_E := E$.
2. Replace each modal operator $\mathcal{K}\phi$ in $AT_E$ with $\mathcal{L}\phi$ if $\phi$ is an atom.
3. For each modal operator $\mathcal{K}\phi$ in $AT_E$ such that $\phi$ is not an atom do
   (a) replacing $\mathcal{K}\phi$ with $\mathcal{L}p_\phi$, and
   (b) adding $p_\phi \equiv \phi$ into $AT_E$,

   where $p_\phi$ and $\mathcal{L}p_\phi$ are a new objective predicate and its belief counterpart.
4. Repeat (3) until $AT_E$ contains no modal operators. □

**Example 5.1** Let $E = \{p \subset q; \;\; q \subset \neg\mathcal{K}(\neg p)\}$. $AT_E = \{p \subset q; \;\; q \subset \neg\mathcal{L}\bar{p}; \;\; \neg p \equiv \bar{p}\}$. $\square$

**Theorem 5.1** Let $E$ be an AE set and $AT_E(P, \mathcal{L}P)$ be the autoepistemic translation of $E$. Then $\Gamma$ is an AE stable expansion of $E$ if and only if there exists a stable expansion $T$ of $AT_E$ such that $\Gamma \equiv_{obj} T$.

*Proof:* By [11], $\Gamma$ is an AE stable expansion of $E$ if and only if $\Gamma$ satisfies Equation (1), i.e.,

$$\Gamma = \{\phi | E \cup \{\mathcal{K}\phi \mid \phi \in \Gamma\} \cup \{\neg\mathcal{K}\phi \mid \phi \notin \Gamma\} \models_S \phi\}.$$

Let $P(\Gamma) = \{\mathcal{K}\phi \mid \phi \in \Gamma \;\text{ and }\; \mathcal{K}\phi \;\text{ appears in }\; E\}$, and
$N(\Gamma) = \{\neg\mathcal{K}\phi \mid \phi \notin \Gamma \;\text{ and }\; \mathcal{K}\phi \;\text{ appears in }\; E\}$.
Since for any $\mathcal{K}\phi$ that does not appear in $E$, $\mathcal{K}\phi$ can not be used to derive any objective formulas, an objective formula $\phi$ is contained in an AE stable expansion $\Gamma$ of $E$ if and only if

$$E \cup P(\Gamma) \cup N(\Gamma) \models_S \phi$$

($\Rightarrow$) Assume $\Gamma$ is an AE stable expansion of $E$ and $AT_E$ be the autoepistemic translation of $E$. For each $\mathcal{K}\phi$ in $E$, let $\mathcal{L}p_\phi$ be the corresponding belief atom in $AT_E$. Construct $\mathbf{m}_\mathcal{L} = \{\mathcal{L}p_\phi \mid \mathcal{K}\phi \in P(\Gamma)\}$, and $N(\mathbf{m}_\mathcal{L}) = \{\neg\mathcal{L}p_\phi \mid \mathcal{L}p_\phi \notin \mathbf{m}_\mathcal{L}\}$. Then for each objective formula $\phi$ whose predicates appear in $E$, $\phi$ is contained in $\Gamma_{obj}$ if and only if $AT_E \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L}) \models \phi$. Let $T = Cn(AT_E \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L})$. Then $\Gamma \equiv_{obj} T$.

Since for each $\mathcal{L}p_\phi$, $\mathcal{L}p_\phi \in \mathbf{m}_\mathcal{L}$ if and only if $E \cup P(\Gamma) \cup N(\Gamma) \models_S \phi$ if and only if $AT_E \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L}) \models p_\phi$, $\mathbf{m}_\mathcal{L}$ is a stable belief of $AT_E$, and therefore, $T$ is a stable expansion of $AT_E$.
($\Leftarrow$) Assume $T$ is a stable expansion of $AT_E$ and $\mathbf{m}_\mathcal{L} = \{\mathcal{L}a \mid a \in T\}$. Then $T = Cn(AT_E \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L}))$. Let $\Gamma'$ be the set of all objective formulas that are logical consequences of $AT_E \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L})$ and whose predicates are in $E$, and $\Gamma = \{\phi | E \cup \{\mathcal{K}\phi \mid \phi \in \Gamma'\} \cup \{\neg\mathcal{K}\phi \mid \phi \notin \Gamma'\} \models \phi\}$.

Since $T$ is a stable expansion of $A$, $\mathcal{L}p_\phi \in \mathbf{m}_\mathcal{L}$ if and only if $AT_E \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L}) \models p_\phi$, and therefore, if and only if $\phi \in \Gamma'$ for $p_\phi \equiv \phi$ is in $AT_E$. Therefore $\Gamma = \{\phi | E \cup \{\mathcal{K}\phi | \phi \in \Gamma\} \cup \{\neg\mathcal{K}\phi \mid \phi \notin \Gamma\} \models_S \phi\}$. It follows that $\Gamma$ is an E stable expansion of $E$ and $\Gamma \equiv_{obj} Cn(AT_E \cup \mathbf{m}_\mathcal{L} \cup N(\mathbf{m}_\mathcal{L}))$. $\square$

The above theorem establishes a one-to-one correspondence between the set of AE stable expansions of $E$ and the set of stable expansions of $AT_E$.

## 5.2 Reiter's default logic

A default theory $\Delta$ is a pair $< D, W >$, where $W$ is a set of propositional sentences and $D$ is a set of defaults, each of which has the form

$$\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\omega}$$

The *extension semantics* of a default theory is characterized by its extensions [21].

Now we translate a given default theory into an autoepistemic theory. Our method is based on the translation given by Konolige [11].

**Definition 5.3** Let $\Delta =< D, W >$ be a default theory. Then $DT_\Delta(P, \mathcal{L}P)$, called *the default translation* of $\Delta$, is an autoepistemic theory defined as follows

1. $W \subseteq DT_\Delta$
2. for each default $\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\omega} \in D$,
   (a) $\omega \subset \mathcal{L}p_\alpha \wedge \neg\mathcal{L}p_{\bar{\beta}_1} \wedge \cdots \wedge \neg\mathcal{L}p_{\bar{\beta}_m}$ is in $DT_\Delta$,
   (b) $p_\alpha \equiv \alpha$ is in $DT_\Delta$,
   (c) $p_{\bar{\beta}_i} \equiv \neg\beta_i$ is in $DT_\Delta$, for $i = 1, \ldots, m$,

   where $p_\alpha, p_{\bar{\beta}_i}$ and $\mathcal{L}p_\alpha, \mathcal{L}p_{\bar{\beta}_i}$ are newly introduced predicates and their belief counterparts.
3. Nothing else is in $DT_\Delta$. □

**Example 5.2** Assume $\Delta_1 =< D_1, W_1 >$, where
$D_1 = \{\frac{:Ma}{a}; \ \frac{:Mb}{b}; \ \frac{:Mc}{c}\}$, and $W_1 = \{\neg a \subset b; \ \neg c \subset b\}$.
Then $DT_{\Delta_1} = \{a \subset \neg\mathcal{L}\bar{a}; \ b \subset \neg\mathcal{L}\bar{b}; \ c \subset \neg\mathcal{L}\bar{c}; \ \bar{a} \equiv \neg a; \ \bar{b} \equiv \neg b; \ \bar{c} \equiv \neg c; \ \neg a \subset b; \ \neg c \subset b\}$.
$\Delta_2 =< D_2, W_2 >$, where $D_2 = \{\frac{a:}{a}, \frac{:\neg a}{a}\}$ and $W_2 = \phi$.
Then $DT_{\Delta_2} = \{a \subset \mathcal{L}a, a \subset \neg\mathcal{L}\bar{a}, \bar{a} \equiv a\}$. □

The following theorem demonstrates the equivalence between the extension of a default theory and the fixpoint of the corresponding autoepistemic theory.

**Theorem 5.2** Let $\Delta =< D, W >$ be a default theory, and $DT_\Delta(P, \mathcal{L}P)$ be the default translation of $\Delta$. Then $E$ is an extension of $\Delta$ if and only if there exists a fixpoint $\mathbf{m}_\mathcal{L}$ of $DT_\Delta$ such that

$$E \equiv_{obj} (DT_\Delta(P, \mathcal{L}P) \cup F_{\mathbf{m}_\mathcal{L}})$$

*Proof:* Without losing generality, assume each default is in the form of $\frac{\alpha : M\beta_1, \ldots, M\beta_n}{\omega}$, where $\alpha, \beta_i$'s and $\omega$ are distinct atoms, and therefore is translated into a clause $\omega \subset \mathcal{L}p_\alpha, \neg\mathcal{L}p_{\bar{\beta}_1}, \ldots, \neg\mathcal{L}p_{\bar{\beta}_n}$. (Otherwise we can always use renaming technique to achieve this.)

Define a transformation $R$ from sets of objective formulas to sets of objective formula as follows.

Let $E$ be a set of objective formula.

1. $R^0 = Cn(W)$.
2. $R^{n+1} = Cn(R^n \cup \{\omega | \frac{\alpha : M\beta_1, \ldots, M\beta_n}{\omega} \in D \text{ and } \alpha \in R^n \text{ and } \neg\beta_i \notin E\})$.

Then $R(E)$ is defined as the least fixpoint of $R^i$, i.e., $R \uparrow \omega$.

Assume $DT_\Delta(P, \mathcal{L}P)$ is the default translation of $\Delta$ and $\mathbf{m}_{\mathcal{L}E} = \{\mathcal{L}p | \mathcal{L}p \in \mathcal{L}P \text{ and } p \in E\}$. Then for any objective formula $\phi$, we are going to show that $\phi \in R(E)$ if and only if

$$DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \cup \{\mathcal{L}a | DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \models_N a\} \models \phi$$

by induction on $R^i$. Assume $\phi \in R(E)$. (We show the only if part and the if part is similar and therefore omitted.)
**basis** $\phi \in R^0$. Since $W \subseteq DT_\Delta$, $DT_\Delta \models \phi$.
**hypothesis** Assume for each $\phi \in R^n$, $DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \cup \{\mathcal{L}a | DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \models_N a\} \models \phi$.
**induction** Then consider $\phi \in R^{n+1}$. We need to show that for each default $\frac{\alpha:M\beta_1,\ldots,\beta_n}{\omega}$ such that $\alpha \in R^n$ and $\neg\beta^i \notin E$, $DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \cup \{\mathcal{L}a | DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \models_N a\} \models \omega$.

By the inductive hypothesis, $DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \cup \{\mathcal{L}a | DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \models_N a\} \models \alpha$, and therefore, $\mathcal{L}\alpha$ is contained in $\{\mathcal{L}a | DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \models_N a\}$. Since $\neg\beta^i \notin E$, $\mathcal{L}p_{\bar{\beta}^i} \notin \mathbf{m}_{\mathcal{L}E}$ and therefore, $\neg\mathcal{L}p_{\bar{\beta}^i} \in N(\mathbf{m}_{\mathcal{L}E})$. It follows that $DT_\Delta \cup N(\mathbf{m}_{\mathcal{L}E}) \cup \{\mathcal{L}a | DT_\Delta \cup F_{\mathbf{m}_{\mathcal{L}E}} \models_N a\} \models \omega$, since $\omega \subset \mathcal{L}a, \neg\mathcal{L}p_{\bar{\beta}^1}, \ldots, \neg\mathcal{L}p_{\bar{\beta}^n}$ is contained in $DT_\Delta$.

Therefore, $E = R(E)$ if and only if $\mathbf{m}_{\mathcal{L}E} = P_{DT_\Delta}(\mathbf{m}_{\mathcal{L}E})$, i.e., if and only if $\mathbf{m}_{\mathcal{L}E}$ is a fixpoint of $DT_\Delta$. However, by [21], $E = R(E)$ if and only if $E$ is an extension of $\Delta$. It follows that $E$ is an extension of $\Delta$ if and only if $\mathbf{m}_{\mathcal{L}E}$ is a fixpoint of $DT_\Delta$ and $E \equiv_{obj} (DT_\Delta \cup F_{\mathbf{m}_\mathcal{L}})$.
□

**Example 5.3** Consider $\Delta_1$ and $\Delta_2$ in Examples 5.2. $DT_{\Delta_1}$ has two fixpoints, viz. $\mathbf{m}_{\mathcal{L}_1} = \{\mathcal{L}\bar{a}, \mathcal{L}\bar{c}\}$ and $\mathbf{m}_{\mathcal{L}_2} = \{\mathcal{L}\bar{b}\}$. Therefore, $\Delta_1$ has two extensions, one contains $W \cup \{b\}$ and the other contains $W \cup \{a, c\}$.

$DT_{\Delta_2}$ has no fixpoints, and therefore, $\Delta_2$ has no extensions. □

The following corollary follows from Theorem 4.4 and 5.2.

**Corollary 5.3** An extension of a default theory $\Delta$ is consistent if and only if the default translation of $\Delta$ is epistemically coherent. □

## 5.3 the stable class semantics

Following the direction of Gelfond [6], a logic program can be represented as an autoepistemic theory of the form

$$a \subset b_1 \wedge \cdots \wedge b_n \wedge \neg\mathcal{L}c_1 \wedge \cdots \wedge \neg\mathcal{L}c_m.$$

Van Gelder has shown that the stable and the well-founded semantics are characterized by the set of all fixpoints and the set of all alternating fixpoints of the GL transformation $T_\Pi$ respectively [4]. Since the GL-transformation is just a special case of our transformation, it is easy to show that our stable and the well-founded semantics for autoepistemic theories coincide with the stable and the well-founded semantics in the content of logic programs [4].

However, as demonstrated earlier, both semantics suffer from various problems [2, 3, 9]. Baral and Subrahmanian have realized the problems associated with the stable and the well-founded semantics and proposed the stable class semantics to resolve the problems. The basic idea behind the stable class semantics is that the transformation of $P_A$ may not always have fixpoints, but there exists a class of points, called the *stable class*, such that the transformation cycles around the class. Then the stable class semantics is defined as the set of minimal strict stable classes, according to a preference relation defined in [2]. However the following example demonstrates that the stable class semantics may lead to counterintuitive conclusions.

**Example 5.4** Let $A = \{a \subset \neg\mathcal{L}a;\ \ b \subset \neg\mathcal{L}b;\ \ c \subset a, \neg\mathcal{L}a;\ \ c \subset b, \neg\mathcal{L}b\}$. Then $A$ has four alternating fixpoints, viz. $\mathbf{m}_{\mathcal{L}_1} = \{\mathcal{L}a, \mathcal{L}b\}$, $\mathbf{m}_{\mathcal{L}_2} = \emptyset$, $\mathbf{m}_{\mathcal{L}_3} = \{\mathcal{L}a\}$, and $\mathbf{m}_{\mathcal{L}_4} = \{\mathcal{L}b\}$. According to Baral and Subrahmanian, these four alternating fixpoints form two stable classes, i.e., $C_1 = \{\mathbf{m}_{\mathcal{L}_1}, \mathbf{m}_{\mathcal{L}_2}\}$ and $C_2 = \{\mathbf{m}_{\mathcal{L}_3}, \mathbf{m}_{\mathcal{L}_4}\}$. The preference relation adapted in the stable class semantics, based on the largeness, prefers $C_2$ to $C_1$, and the consequence is that $c$ is true in the stable class semantics of this logic program.

Since the premises of clauses that derive $c$ can not be satisfied in any circumstances, $c$ should not be true in any reasonable semantics.

Since $\mathbf{m}_{\mathcal{L}_2}$, $\mathbf{m}_{\mathcal{L}_3}$ and $\mathbf{m}_{\mathcal{L}_4}$ are non-normal alternating fixpoints, the generalized stable semantics is characterized by $\mathbf{m}_{\mathcal{L}_1}$ and therefore avoids the counter-intuitive conclusion. □

The stable class semantics took an important step forward to resolving the problems associated with the stable and the well-founded semantics. However, the preference relation used by Baral and Subrahmanian is based on the largeness only and ignores the normality which is the source of the counterintuitive conclusions of the stable class semantics.

## 5.4 the regular model semantics

The regular model semantics, proposed by You and Yuan in [24, 26] and the preferential semantics, proposed by Dung [3], provide a satisfactory semantics for logic programs. We are going to show that the regular model semantics is just a special case of the generalized stable semantics.

Let $\Pi$ be a logic program consisting of clauses of the form

$$a \subset b_1 \wedge \ldots \wedge b_n \wedge \neg c_1 \wedge \ldots \wedge \neg c_m$$

A *three-valued interpretation* of $\Pi$ is defined as a tuple $I =< T, F >$, where $T$ and $F$ are two disjoint sets of atoms such that atoms in $T$ and $F$ are considered `true` and `false` in $I$ respectively. A three-valued interpretation $M =< T, F >$ is said to be a *three-valued model* of $\Pi$ just in case that for each clause

$$a \subset b_1 \wedge \ldots \wedge b_n \wedge \neg c_1 \wedge \ldots \wedge \neg c_m$$

in $\Pi$, (1) $a \in T$ if $b_i$s are in $T$ and $c_j$s are in $F$, and (2) there exists either $b_i \in F$ or $c_j \in T$ if $a \in F$. An atom $a$ in $T$ is said to be *founded* with respect to a three-valued model $M =< T, F >$ if

1. there exists a clause
$$a \subset \neg c_1 \wedge \ldots \wedge \neg c_m$$
such that $c_j \in F$ for $1 \leq j \leq m$, or recursively,

2. there exists a clause
$$a \subset b_1 \wedge \ldots \wedge b_n \wedge \neg c_1 \wedge \ldots \wedge \neg c_m$$
such that $b_i$ are founded and $c_j \in F$ for $1 \leq j \leq m$.

A *justified model* of $\Pi$ is a three-valued model $M = < T, F >$ such that every atom in $T$ is founded. A *regular model* of $\Pi$ is then defined as a justified model $M = < T, F >$ of $\Pi$ such that there exists no other justified model $M' = < T', F' >$ of $\Pi$ such that $T \subseteq T'$ and $F \subseteq F$.

The following theorem shows that the regular model semantics of logic programs coincides with the generalized stable semantics of the corresponding autoepistemic logic theories.

**Theorem 5.4** Let $\Pi$ be a logic program and $A(P, \mathcal{L}P)$ be the corresponding autoepistemic theory. Then $\mathbf{m}_\mathcal{L}$ is a generalized stable belief of $A$ if and only if $M = < T, F >$ is a regular model of $\Pi$, where $F = \{a \mid \neg\mathcal{L}a \in N(\mathbf{m}_\mathcal{L})\}$ and $T = \{a \mid \mathcal{L}a \in P_A(\mathbf{m}_\mathcal{L})\}$.

*Proof:* ($\Rightarrow$). Let $\mathbf{m}_\mathcal{L}$ be a generalizes stable belief of $A$, $T = \{a \mid \mathcal{L}a \in P_A(\mathbf{m}_\mathcal{L})\}$ and $F = \{a \mid \mathcal{L}a \not\in \mathbf{m}_\mathcal{L}\}$. Assume $M = < T, F >$ is a not a three-valued model of $\Pi$. Since $\mathbf{m}_\mathcal{L}$ is normal and $T = \{a \mid \mathcal{L}a \in P_A(\mathbf{m}_\mathcal{L})\}$, there exist an atom $a \in F$ and a clause

$$a \subset b_1 \wedge \ldots \wedge b_n \wedge \neg c_1 \wedge \ldots \wedge \neg c_m$$

in $\Pi$ such that $b_i$s are not in $F$ and $c_j$s are not in $T$. Since $\mathbf{m}_\mathcal{L}$ is an alternating fixpoint of $A$, we have $neg\mathcal{L}c_j$ are in $N(P_A(\mathbf{m}_\mathcal{L}))$ and $\mathcal{L}b_i$ are in $P_A(P_A(\mathbf{m}_\mathcal{L}))$. Thus $A \uplus N(P_A(\mathbf{m}_\mathcal{L})) \models_N a$, which contradicts that $a \in F$. Therefore, $M$ is a three-valued model of $\Pi$.

It follows that $M$ is a regular model of $\Pi$ since $T = \{a \mid \mathcal{L}a \in P_A(\mathbf{m}_\mathcal{L})\}$ and $\mathbf{m}_\mathcal{L}$ is a minimal, normal alternating fixpoint.

($\Leftarrow$) It is similar to the if-part and thus omitted.

$\Box$

It is not difficult to show that the preferential semantics defined in [3] coincides with the regular model semantics.

Recently attentions have been focused on semantics of general logic programs with negation, i.e., disjunctive logic programs and deductive databases with both negation as failure and the classical negation [7, 19, 20]. The generalized stable semantics for autoepistemic theories defines a natural semantics for such general logic programs.

# 6 Conclusions

We conclude the paper by the following observations:

1. The stable expansion of Moore's logic is based on rigorous introspection.

2. The extension of Reiter's default logic is based on rigorous and positively minimized introspection.

3. The stable class semantics is based on negatively maximized introspection that may not be normal.

4. The well-founded semantics for logic programs is based on negatively sound introspection.

5. The regular model semantics for logic programs is based on negatively maximized introspection.

**Acknowledgements** The author would like to thank reviewers for providing valuable suggestions and constructive criticism to improve the presentation of the paper.

## References

[1] C.R. Baral and V.S. Subrahmanian. Dualities between alternative semantics for logic programming and nonmonotonic reasoning. In *Proceedings of the First International Workshop on Logic Programming and Mon-monotonic Reasoning,* the full version to appear in *Journal of Automated Reasoning*, 1991.

[2] C.R. Baral and V.S. Subrahmanian. Stable and extension class theory for logic programs and default logics. *Journal of Automated Reasoning*, 8: 345 – 366, 1992.

[3] P.M. Dung. Negations as hypotheses: an abductive foundation for logic programming. In *Proceedings of the 8th Intern. Conference on Logic Programming*, pages 3–17, 1991.

[4] A. Van Gelder. The alternating fixpoints of logic programs with negation. In *Proceedings of the 8th ACM PODS*, pages 1 – 10, 1989.

[5] A. Van Gelder, K. Ross, and J.S. Schlipf. The well-founded semantics for general logic programs. *JACM*, 38:620 – 650, 1991.

[6] M. Gelfond. On stratified autoepistemic theories. In *Proceedings of AAAI-87*, 1987.

[7] M. Gelfond and V. Lifschitz. Logical programs with classical negation. In *Proceedings of the 7th International Conference on Logic Programming*, pages 579–597, 1990.

[8] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *Proc. of the 5th Intl. Conference and Symposium on Logic Programming*, pages 1070–1080, 1988.

[9] Y. Hu and L. Yuan. Extended well founded semantics for logic programs with negations. In *Proceedings of the 8th Intl. Conference on Logic Programming*, pages 412 – 425, 1991.

[10] G.E. Hughes and M.J. Cresswell. *Introduction to Modal Logic*. Routledge, 1968.

[11] K. Konolige. On the relationship between default and autoepistemic logic. *Artificial Intelligence*, 35:343 – 382, 1988.

[12] V. Lifschitz. Between circumscription and autoepistemic logic. *KR'89*, 235–244, 1989.

[13] W. Marek and V.S. Subrahmanian. The relationship between logic program semantics and non-monotonic reasoning. In *Proceedings of the 6th International Conference and Symposium on Logic Programming*, pages 600 – 617, 1989.

[14] W. Marek and V.S. Subrahmanian. The relationship between stable, supported, default and auto-epistemic semantics for general logic programs. *Theoretical Computer Science*, 1989.

[15] W. Marek and M. Truszczunski. Relating autoepistemic and default logic. In *Proc. of KR'89*, pages 276 – 288, 1989.

[16] R.C. Moore. Semantic considerations on non-monotonic logic. *AI*, 25:75–94, 1985.

[17] D. Perlis. Autocircumscription. *Artificial Intelligence*, 36:223 – 236, 1988.

[18] T.C. Przymusinski. Autoepistemic logics of closed world beliefs and logic programming. In *Proc. of the Workshop on Nonmonotonic Reasoning and Logic Programming*, pages 3–20, 1991.

[19] T.C. Przymusinski. Extended stable semantics for normal and disjunctive logic programs. In *Proceedings of the 7th International Conference and Symposium on Logic Programming*, pages 459–477, 1990.

[20] T.C. Przymusinski. Semantics of disjunctive logic programs and deductive databases. In *Proceedings of the 2nd DOOD conference*, 1991.

[21] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[22] G. Schwarz. Bounding introspection in nonmonotonic reasoning. *KR'92*, 581 – 590, 1992.

[23] M. Tiomkin and M. Kaminski. Nonmonotonic default modal logics. In *Proceeding of TARK 1990*, pages 73 – 84, 1990.

[24] J.-H. You and L. Y. Yuan. Three-valued formalization of logic programming: is it needed? In *Proceedings of the 9th ACM Symposium on Principles of Database Systems*, pages 172–182, 1990.

[25] L. Yuan and J. You. Autoepistemic circumscription and logic programs. *Journal of Automated Reasoning*, 10: 143 – 160, 1993.

[26] L.Y. Yuan and J.H You. Justification rules and justified model semantics. In *Proc. of Pacific Rim International Conference on AI*, pages 822–827, 1990.