Autoepistemic Circumscription and Logic Programming†

Li Yan Yuan and *Jia-Huai You*

Department of Computing Science

University of Alberta

Edmonton, Alberta

Canada, T6G 2H1

Abstract

We propose a framework of autoepistemic reasoning in which the underlying semantics is determined by the choice of a nonmonotonic inference mechanism and by specifying a *belief constraint*. While the latter makes the approach flexible in meeting possibly different applications, the former links the resulting semantics to a nonmonotonic reasoning formalism and thus allows adoption of existing techniques. In this paper we choose *circumscription* as the underlying inference mechanism and use two different belief constraints to define two semantics, *the stable circumscriptive semantics* and *the well-founded circumscriptive semantics*, for autoepistemic theories. The former coincides with Moore's autoepistemic logic for logic programs and is arguably more desirable in handling disjunctive autoepistemic theories. The latter is a reconstruction and extension of Przymusinski's iterative method for computing the least $AEL\,(circ\,)$ expansions for logic programs. We show that for logic programs the two construction methods coincide. However, while Przymusinski's construction method is restricted to logic programs only, the well-founded circumscriptive semantics is applicable to more general autoepistemic theories.

KEY WORDS: *Autoepistemic Logic, Circumscription, Semantics of Logic Programming, Nonmonotonic Reasoning.*

The paper will appear in the Journal of Automated Reasoning.

1. Introduction

Moore's autoepistemic logic (AEL) is intended to model the beliefs of a reasoning agent reflecting upon her own beliefs. The truth of an agent's beliefs, expressed as an autoepistemic theory, is determined by a close relation between propositions that are true in the external world and formulas that the agent believes. In AEL these ideas are formalized by the notion of stable expansion [3, 8].

In this paper we take an alternative view of belief reasoning. A rational reasoning agent expresses her description of the world in an autoepistemic language. What propositions that the agent believes depends on some rational inference rules and suitable belief constraints. These inference rules and belief constraints may differ from one agent to another and are application oriented.

We elaborate this approach by choosing the nonmonotonic reasoning mechanism *circumscription* as our agent's inference engine. We define a semantics, called the *stable circumscriptive semantics*, expressed by a parallel circumscription formula and constrained by requiring that the truth of a belief be the same as its objective counterpart. This yields a semantics in which belief propositions are reflections of their objective counterparts in individual models. In many cases, these models express the same semantics as that of the corresponding stable expansions; this is the case, for example, for logic programs. When they differ, the models arguably offer a more desirable semantics. This is because, as pointed out by Przymusinski [9], the reasoning pattern in AEL is based on Reiter's closed world assumption [11], which insists on determination of the belief status of every proposition and as such may produce unintuitive expansions for disjunctive theories.

We show how the change of belief constraints may yield new nonmonotonic formalisms. We adopt a more sophisticated belief constraint and show that, in conjunction with parallel circumscription, a *well-founded* formalism of belief reasoning, with the well-founded semantics of logic programs being a special case, provides a desirable semantics for general autoepistemic theories.

The work reported here is closely related to Przymusinski's recent work on autoepistemic logics of closed beliefs [9], where a nonconstructive fixed point equation is used to define stable expansions of an autoepistemic theory in which one can also choose a nonmonotonic inference mechanism as the underlying inference engine. When circumscription is used, the resulting logic is named $AEL(circ)$. Our well-founded circumscriptive semantics can be viewed as a

reconstruction and extension of Przymusinski's method for generating least $AEL\,(circ\,)$ expansions. There are two important differences. First, in Przymusinski's parallel circumscription objective as well as belief propositions are minimized, while in ours only objective propositions are minimized.† Secondly, circumscription is applied in *every* step of the construction in Przymusinski's approach but in our approach only once at the beginning of the construction. The second difference does not change the underlying semantics, as we will show that Przymusinski's construction can be simplified so that circumscription only needs to be applied once at the beginning. Thus, the only crucial difference is on the set of propositions to be minimized. We show that for logic programs the two methods coincide. Interestingly, this conclusion is due to a new result which says that for logic programs minimal models coincide with *P*-minimal models. The reason that our method is more general is that the form of circumscription employed in our construction method commits a belief proposition to a truth value whenever it *has to*, parallel to the idea of "lazy evaluation" in functional programming, while in Przymusinski's the truth value of a belief proposition may be prematurely committed.

Lifschitz introduced a formalism based on *introspective circumscription* [5] which employs a different circumscription formula over a class of first order autoepistemic theories. The starting point in his approach is in the same line as Moore's; in particular, an agent disbelieves what she cannot derive from the given theory, Thus, for example, the theory $\{a \vee b\}$ has the same semantics in both Lifschitz's approach and Moore's AEL with both a and b disbelieved, while in our approach the semantics is defined by two minimal models.

The paper is organized as follows. Section 2 provides the background material. Autoepistemic circumscription is defined in Section 3. The stable circumscriptive semantics is presented in Section 4 and is compared with Moore's AEL. In Section 5 we define the well-founded circumscriptive semantics for general autoepistemic theories and prove its equivalence with Przymusinski's construction for logic programs. Section 6 contains final remarks.

The proofs that are not given in the main text can be found in the Appendix.

---

† Przymusinski has also adopted this form of circumscription in his later papers; see, for example, [10].

2. Preliminaries

2.1 Moore's Autoepistemic Logic

We restrict all definitions to the language of *propositional logic*.

A *propositional autoepistemic language* is a propositional language $\boldsymbol{L}$ containing a special belief symbol $\boldsymbol{L}$ such that for any objective proposition symbol ξ in $\boldsymbol{L}$ its alphabet also contains the corresponding *belief proposition* $\boldsymbol{L}\xi$, and the intended meaning of $\boldsymbol{L}\xi$ is "believe in ξ". A proposition without $\boldsymbol{L}$ is called *objective*. An *autoepistemic formula* is a well-formed formula and an *autoepistemic theory* is a set of autoepistemic formulas in an autoepistemic language.

It is important that a belief proposition be viewed as a proposition in its entirety, i.e., a *named* object, not as applying some *modal* operator $\boldsymbol{L}$ to a proposition.

Given an autoepistemic language, an interpretation of an autoepistemic theory is a truth value assignment to proposition symbols and belief propositions, and a model is an interpretation in which the formulas in the theory evaluate to true, according to the usual propositional valuation. We usually denote a model by a set of atoms whose elements are assigned to *true* and those not in it are assigned to *false*. A theory is said to be *consistent* if it possesses at least one model.

Note that the technical exploration of this paper is restricted to a propositional language where the belief symbol $\boldsymbol{L}$ appears only in front of an objective atom. These belief atoms and their negated forms are called *belief literals*.

In autoepistemic logic, given an autoepistemic theory, the set of beliefs of a rational agent is expressed by a *stable autoepistemic expansion* or a *minimal stable autoepistemic expansion* of the theory. First, we explain some notations.

Let T be an autoepistemic theory. Then T_o is the set of all objective formulas in T, $\boldsymbol{L}T$ is defined as $\{\boldsymbol{L}\xi \mid \xi$ is an objective atom in $T\}$, and $\neg\boldsymbol{L}\overline{T}$ as $\{\neg\boldsymbol{L}\xi \mid \xi$ is an objective atom not in $T\}$.

Definition 2.1. (Stalnaker [12]) A *stable set* Γ is a set of autoepistemic formulas such that

(1) Γ is closed under first-order consequences,

(2) $\boldsymbol{L}\Gamma \subseteq \Gamma$, and

(3) $\neg\boldsymbol{L}\overline{\Gamma} \subseteq \Gamma$. □

Definition 2.2. (Moore [8], Konolige [3]) Let A be an autoepistemic theory. Then T is said to be

(1) a *stable autoepistemic expansion* (or simply *stable expansion*) of A if

$$T = \{\phi \mid A \cup \boldsymbol{L}T \cup \neg\boldsymbol{L}\overline{T} \models \phi\}, \text{ and}$$

(2) a *minimal autoepistemic expansion* of A if T is a stable autoepistemic expansion of A such that there is no stable set S containing A such that $S_o \subset T_o$. □

2.2. Autoepistemic Translation of Logic Programs

A *logic program* is a set of clauses of the form:

$$A \leftarrow B_1, ..., B_n, \neg C_1, ..., \neg C_m$$

where $n, m \geq 0$ and A, B_i's, and C_j's are atoms. A nonground clause represents a set of ground clauses, instantiated by the function symbols and constants of the language. In the rest of this paper, we only consider (finite or infinite) sets of ground clauses. Following Gelfond's approach [1, 9], such a clause can be transformed into an epistemic clause:

$$A \leftarrow B_1, ..., B_n, \neg \boldsymbol{L}C_1, ..., \neg \boldsymbol{L}C_m.$$

Sets of epistemic clauses of the above form are called *autoepistemic translations* of logic programs. They will also be called *autoepistemic logic programs*, or simply logic programs when no confusion arises.

2.3. Circumscription

To capture the feature of common-sense reasoning that no objects should have a property unless it has been stated they must, McCarthy [6, 7] developed a nonmonotonic reasoning technique, called circumscription.

Circumscription is usually defined by a second order formula. Since the technical exploration of this paper is not directly related to this definition of circumscription, we will not provide it here but refer the reader to the paper by Lifschitz [4] for more details. Notationally, if $P = \{P_1, \cdots, P_n\}$ is a tuple of predicates to be minimized and $Z = \{Z_1, \cdots, Z_l\}$, disjoint from P, is a tuple of predicates that are allowed to vary when minimizing P, then the circumscription of P in a first order theory A with variable Z is denoted by $CIRC(A; P; Z)$. $CIRC(A; P)$ denotes the special case when Z is empty and $CIRC(A)$ when P is the set of all predicates.

Circumscription is based on the notion of minimal model. A model M of theory A is (P,Z)-*minimal* iff there is no model N, such that M and N differ only in how they interpret predicates from P and Z, and the extension of every predicate in P from N is a subset of its extension

in M. Again, such a model is called *P-minimal* if Z is empty and simply *minimal* if P is the set of all predicates.

It is well-known that circumscription in fact restricts models of a formula to its (*P*,*Z*)-*minimal* models.

Theorem 2.1 (McCarthy [6], Lifschitz [4]) Let A be a first order theory and ϕ a formula. Then, $CIRC(A; P; Z) \models \phi$ iff $M \models \phi$ *for every (P,Z)-minimal model M of A.* □

2.4. Przymusinski's $AEL(circ)$

Przymusinski proposed $AEL(circ)$ and an iterative construction method for logic programs [9]. When the language contains only belief literals rather than more general belief formulas, his definition of $AEL(circ)$ reduces to the following.

Definition 2.3. (Przymusinski [9]) A consistent theory T is a stable $AEL(circ)$ expansion of a theory A iff T satisfies the following fixed point equation:

$$T = A \cup \{\boldsymbol{L}\xi \mid T \models_{circ} \xi\} \cup \{\neg\boldsymbol{L}\xi \mid T \models_{circ} \neg\xi\},$$

where $T \models_{circ} \xi$ means the objective atom ξ is true in all minimal models of T, and $T \models_{circ} \neg\xi$ means the objective atom ξ is false in all minimal models of T. □

Notice that Przymusinski's circumscription is over the entire set of atomic propositions, both objective and belief.

An autoepistemic theory may have more than one stable $AEL(circ)$ expansion. Przymusinski has shown that there exists a least stable $AEL(circ)$ expansion for any logic program A, which is a subset of any stable $AEL(circ)$ expansion of A and can be constructed by an iterative construction:

$$T_0 = A$$
$$T_{n+1} = T_n \cup \{\boldsymbol{L}\xi \mid T_n \models_{circ} \xi\} \cup \{\neg\boldsymbol{L}\xi \mid T_n \models_{circ} \neg\xi\}.$$

A limitation of this construction is that it cannot be extended to theories outside autoepistemic logic programs. Here is an example.

Example 2.1. Consider $A = \{b,\ a \leftarrow \boldsymbol{L}b,\ \neg a \leftarrow \neg\boldsymbol{L}b\}$. By definition, $T = \{\boldsymbol{L}b, \boldsymbol{L}a\} \cup A$ is the unique stable $AEL(circ)$ expansion of A. However, since A is not an autoepistemic logic program, the iterative definition above could not be used to find T. By the definition, $T_0 = A$, and T_1 should contain $\boldsymbol{L}b$ and $\neg\boldsymbol{L}a$, since $T_0 \models_{circ} b$ and $T_0 \models_{circ} \neg a$. The fact that $\boldsymbol{L}b$ is in T_1 forces $T_1 \models_{circ} a$ and therefore, T_2 contains $\boldsymbol{L}a$, resulting in inconsistency. □

3. Autoepistemic Circumscription

In this section we introduce the notion of autoepistemic circumscription.

The particular form of circumscription over autoepistemic theories, which will be extensively used in this paper, is given below.

Definition 3.1. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory, where P and $\boldsymbol{L}P$ denote sets of objective atoms and belief propositions respectively. *Discriminant circumscription* of $A(P, \boldsymbol{L}P)$, denoted as $DC(A)$, is defined as

$$CIRC(A(P, \boldsymbol{L}P); P).$$ □

Discriminant circumscription is meant to compare models with a fixed interpretation for belief atoms $\boldsymbol{L}P$. The set of *P*-minimal models is thus a subset of all models. A proposition is therefore easier to commit to *false* in the set of all *P*-minimal models than in the set of all models. That is, discriminant circumscription $DC(A)$ defined above commits objective propositions as much as possible according to the satisfiability in all *P*-minimal models while leaving the status of belief propositions to be the same as in the semantics of all models.

Of course, no agent may choose her beliefs arbitrarily. One way to constrain such a choice is by specifying a desired relationship between objective propositions and belief propositions.

Definition 3.3. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory. Then *autoepistemic circumscription* of $A(P, \boldsymbol{L}P)$ is defined as

$$CIRC(A(P, \boldsymbol{L}P); P) \wedge CONT(P, \boldsymbol{L}P)$$

where $CONT(P, \boldsymbol{L}P)$ is an autoepistemic theory, in the same underlying language, that constrains the relationship between objective and belief propositions. □

4. The Stable Circumscriptive Semantics of Autoepistemic Theories

The stable circumscriptive semantics of autoepistemic theories is derived from autoepistemic circumscription by a special type of constraint that requires an agent's beliefs faithfully reflect upon her own beliefs.

Definition 4.1. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory. Then the *stable circumscriptive semantics* of $A(P, \boldsymbol{L}P)$ is defined as the model-theoretic semantics of the following formula:

$$CIRC(A(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P)$$

where $P \equiv \boldsymbol{L}P$ denotes for every objective atom ξ in P, $\xi \equiv \boldsymbol{L}\xi$. □

The definition essentially says only those *P*-minimal models of A where belief propositions agree with their objective counterparts are chosen to serve the semantics of A.

Example 4.1. Consider a simple autoepistemic theory $A = \{a \leftarrow \neg \boldsymbol{L}b\}$. Out of sixteen ways of assigning truth values to the entire set of the predicates $\{a, b, \boldsymbol{L}a, \boldsymbol{L}b\}$, only four are *{a,b}*-minimal models, which are given in the table below where 1 denotes *true* and 0 *false*:

a	b	$\boldsymbol{L}a$	$\boldsymbol{L}b$
1	0	0	0
0	0	0	1
1	0	1	0
0	0	1	1

Obviously, only the third row satisfies the constraint $P \equiv \boldsymbol{L}P$ and it thus serves as the stable circumscriptive semantics of theory A. □

4.1 Relation with Stable Expansion Semantics of *AEL*

In the rest of this section we present a number of results that relate the stable circumscriptive semantics with the stable expansion semantics of *AEL*.

In the first theorem we are to show that the stable circumscriptive semantics is more regular than Moore's *AEL* in the sense that whenever a theory has an "epistemically coherent" stable expansion there is a corresponding model by the stable circumscriptive semantics.

Definition 4.2. An autoepistemic theory A is said to be *epistemically coherent* if there exists a model M of A such that for any objective atom ξ, $M \models \xi \equiv \boldsymbol{L}\xi$. Such a model is referred to as an *autoepistemic model*. □

Note that the notion of epistemic coherence does not require an agent's beliefs perfectly match the truth status of their objective counterparts in each way the theory is interpreted. It is reasonable for an interpretation of a theory to have different truth values for an objective atom and its belief counterpart. However, the nonexistence of an interpretation in which the truth of an objective proposition agrees with its belief counterpart nevertheless exhibits certain irregularity, either in the theory itself or due to the semantics of the logic. Besides obvious cases like theory $a \wedge \neg \boldsymbol{L}a$, disjunctive theories will especially be affected. A simple example of a disjunctive theory is $A = \{a \vee b\}$, which has a unique stable expansion that contains $a \vee b$, $\neg \boldsymbol{L}a$, and $\neg \boldsymbol{L}b$, etc. This expansion thus fails to be epistemically coherent. On the other hand, the semantics of A under the stable circumscriptive semantics is expressed by the two models, $\{a, \boldsymbol{L}a\}$

and $\{b, \boldsymbol{L}b\}$, from which our reasoning agent can be said as either disbelieving a or disbelieving b, but not both at the same time. The drawback of failing epistemic coherence in belief modeling becomes more evident when a slightly different theory $A' = \{a \vee b, \boldsymbol{L}a\}$ is considered. In this case, the only stable expansion we had for theory A disappeared. But the stable circumscriptive semantics retains model $\{a, \boldsymbol{L}a\}$ and our agent can then be said as believing a.

In view of these observations, the following theorem essentially says that the impact of the stable circumscriptive semantics is really on disjunctive theories.

Theorem 4.1. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory and

$$AEC = CIRC(A(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P).$$

Then

(1) for any epistemically coherent stable expansion T of A, there exists a model **m** of AEC such that **m** is the unique autoepistemic model of T; and

(2) for any epistemically coherent minimal stable expansion T of A, there exists a minimal model **m** of AEC such that **m** is the unique autoepistemic model of T. □

Example 4.2. Consider $A_1 = \{a \vee \boldsymbol{L}b, b \vee \boldsymbol{L}a\}$. A_1 has two stable expansions Γ_1 and Γ_2, where Γ_1 contains the atoms in **m** $= \{a, \boldsymbol{L}a, \neg\boldsymbol{L}b\}$ and Γ_2 contains those in **n** $= \{b, \boldsymbol{L}b, \neg\boldsymbol{L}a\}$. Both Γ_1 and Γ_2 are epistemically coherent, and **m**' $= \{a, \boldsymbol{L}a\}$ and **n**' $= \{b, \boldsymbol{L}b\}$ are their unique autoepistemic models respectively. In fact, **m**' and **n**' are exactly the two autoepistemic models of $CIRC(A_1(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P)$.

Consider $A_2 = \{a \leftarrow \boldsymbol{L}a\}$. A_2 has two stable expansions, both of which are epistemically coherent. Furthermore, one of them has **m** $= \{a, \boldsymbol{L}a\}$ as the unique autoepistemic model, and the other has **n** $= \varnothing$ as the unique autoepistemic model. Again, **m** and **n** are the only models of $CIRC(A_2(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P)$. However, **n** is a minimal model but **m** is not. Note that **m**, which is a P-minimal model of A_2, is considered not desirable by some authors [3], since it is as if the reasoning agent believes a first and then derives a. □

In the second theorem, we are going to show a result which says that the stable circumscriptive semantics can in fact be achieved in AEL by adding a set of extra axioms to a given theory. This result provides further insight into the relationship between the stable circumscriptive semantics and that based on stable expansions.

Theorem 4.2. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory and

$$AEC = CIRC(A'(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P),$$

where $A' = A \cup \{\boldsymbol{L}\xi \leftarrow \xi \mid \xi \in P\}$. Then

(1) T is a consistent stable expansion of A' if and only if there exists a model **m** of AEC such that **m** is the unique autoepistemic model of T;

(2) T is a consistent minimal stable expansion of A' if and only if there exists a minimal model **m** of AEC such that **m** is the unique autoepistemic model of T. □

Example 4.3. Consider $A = \{a \vee b\}$ again. Then, $A' = \{a \vee b\} \cup \{\boldsymbol{L}a \leftarrow a,\ \boldsymbol{L}b \leftarrow b\}$. There are two stable expansions of A', one containing $\neg a$, $\neg\boldsymbol{L}a$, b, and $\boldsymbol{L}b$, and the other containing a, $\boldsymbol{L}a$, $\neg b$, and $\neg\boldsymbol{L}b$. The two stable expansions correspond to the two models $\{b, \boldsymbol{L}b\}$ and $\{a, \boldsymbol{L}a\}$ of AEC given in Theorem 4.2. □

The stable model semantics of logic programs was proposed by Gelfond and Lifschitz with its root in autoepistemic logic [2]. Our formulation provides an alternative definition of the stable model semantics. A proof sketch of the following theorem can be found in [13].

Theorem 4.3. (Yuan and You [13]) Let Π be a logic program and $\Pi'(P, \boldsymbol{L}P)$ be its autoepistemic translation. Then, **m** is a stable model of Π if and only if the set $\mathbf{m} \cup \{\boldsymbol{L}\zeta \mid \zeta \in \mathbf{m}\}$ is a model of

$$CIRC(\Pi'(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P). \quad \square$$

5. The Well-Founded Circumscriptive Semantics of Autoepistemic Theories

In this section we define a constraint theory, called the *well-founded constraint theory*, which will be used to define a well-founded semantics for autoepistemic theories. We then study its properties and its relation with Przymusinski's $AEL\,(circ)$ and iterative construction for logic programs.

Definition 5.1. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory and $DC(A) = CIRC(A(P, \boldsymbol{L}P); P)$. Then the *well-founded constraint theory* of A, denoted by $WFT(A)$, is iteratively constructed as follows:

$$W^0 = \varnothing.$$

$$W^{n+1} = W^n \cup \{\boldsymbol{L}\xi \mid \xi \text{ is an objective atom and } A \wedge W^n \models \xi\}$$
$$\cup \{\neg\boldsymbol{L}\xi \mid \xi \text{ is an objective atom and } DC(A) \wedge W^n \models \neg\xi\}.$$

Then $WFT(A)$ is defined as the least fixed point of sequence $\{W^i\}$. □

Remark 5.1.

1. Since the usual logic valuation $\models$ is monotonic, the sequence W^n defined in this way is nondecreasing and thus has a least fixed point $W^m = W^{m+1}$ for some ordinal m. Notice that this claim holds even for theories whose least fixed point is inconsistent.

2. Also because $\models$ is monotonic, the inductive step in Definition 5.1 can be equivalently defined as:

$$W^{n+1} = \{\boldsymbol{L}\xi \mid \xi \text{ is an objective atom and } A \wedge W^n \models \xi\}$$
$$\cup \{\neg\boldsymbol{L}\xi \mid \xi \text{ is an objective atom and } DC(A) \wedge W^n \models \neg\xi\}.$$

3. $DC(A)$ need only be computed once and the results are used in every step of the iteration.

We are now ready to define the well-founded circumscriptive semantics.

Definition 5.2. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory, and $WFT(A)$ be the well-founded constraint theory of A. Then the *well-founded circumscriptive semantics* of A is defined by the following autoepistemic circumscription formula AEC:

$$AEC = CIRC(A(P, \boldsymbol{L}P); P) \wedge WFT(A).$$

The beliefs of a rational agent based on the premise A are the theorems of AEC. □

Example 5.1. Consider $A = \{b,\ a \leftarrow \boldsymbol{L}b,\ \neg a \leftarrow \neg\boldsymbol{L}b\}$, which we have used in Subsection 2.4 to illustrate a limitation of Przymusinski's construction.

$WFT(A) = \{\boldsymbol{L}a, \boldsymbol{L}b\}$ can be easily constructed. Since $DC(A) = CIRC(A; \{a, b\}) = A \wedge (a \leftrightarrow \boldsymbol{L}b)$, we have $W^1 = \{\boldsymbol{L}b\}$ for $DC(A) \models b$. Then $W^2 = W^1 \cup \{\boldsymbol{L}a\}$ for $DC(A) \wedge W^1 \models a$, and therefore, $WFT(A) = W^2$. It then can be seen that the atoms that are true of the formula $CIRC(A(P, \boldsymbol{L}P); P) \wedge WFT(A)$ are b, $\boldsymbol{L}b$, a, and $\boldsymbol{L}a$. This coincides with Przymusinski's $AEL(circ)$. □

Another way to understand the well-founded construction is by first computing $CIRC(A(P, \boldsymbol{L}P); P)$ and then adding (conjuncting) belief literals iteratively until the fixed point is reached. The latter process is equivalent to closing the theory under the two inference rules:

$$\phi/\boldsymbol{L}\phi \quad \text{and} \quad \neg\phi/\neg\boldsymbol{L}\phi.$$

We now illustrate this seemingly intricate process for the theory in Example 5.1.

First note that any *{a,b}*-minimal model must assign *true* to b and the same truth value to a and $\boldsymbol{L}b$. We thus have four *{a,b}*-minimal models of the theory, represented by the four rows in the following table (as usual, 1 denotes *true* and 0 *false*).

a	b	$\boldsymbol{L}a$	$\boldsymbol{L}b$
0	1	0	0
0	1	1	0
1	1	0	1
1	1	1	1

Because b is true in all these models, $\boldsymbol{L}b$ is added by the well-founded construction; this corresponds to the removal of the first two rows as models. In the remaining two models, a becomes true in both and therefore $\boldsymbol{L}a$ is added; this leaves the fourth row as the only model at last. This is in contrast with Przymusinski's method where the objective as well as belief propositions are minimized in parallel. Since the first row in the above table is the only minimal model, we add $\boldsymbol{L}b$ and $\neg\boldsymbol{L}a$. This leaves the third row as the only model satisfying the original theory and the newly added belief literals, in which a is true and thus $\boldsymbol{L}a$ should be concluded, resulting in inconsistency.

5.1. Properties of the Well-Founded Circumscriptive Semantics

We now formally describe some of the properties of the well-founded circumscriptive semantics.

The following theorem shows that this semantics is more cautious than either Moore's AEL or Przymusinski's $AEL\,(circ)$.

Theorem 5.1. Let $A\,(P, \boldsymbol{L}P)$ be an autoepistemic theory and $WFT\,(A)$ be the well-founded constraint theory of A. Then

(i) $WFT\,(A)$ is contained in every stable expansion of A;

(ii) $WFT\,(A)$ is contained in every stable $AEL\,(circ)$ expansion of A. □

A question arises naturally as whether the well-founded circumscriptive semantics is equivalent to Przymusinski's $AEL\,(circ)$ in some sense. First, the semantics may not necessarily correspond to a particular $AEL\,(circ)$ expansion. This is because a stable $AEL\,(circ)$ expansion may not be minimal in the sense of Konolige. For example, the theory $\{a \leftarrow \boldsymbol{L}a\}$ has two stable $AEL\,(circ)$ expansions, one containing a and $\boldsymbol{L}a$ and the other containing $\neg\boldsymbol{L}a$. These two stable $AEL\,(circ)$ expansions correspond to two P-minimal models and therefore our well-founded construction generates no belief literals. If for this example the intersection of the two stable $AEL\,(circ)$ expansions happens to be equivalent to its well-founded circumscriptive semantics, then the next example shows this is not true in the general case.

Example 5.2.

This example is much involved. We first assume an autoepistemic theory A whose only *P*-minimal models are as follows:

	$\boldsymbol{L}a$	a	$\boldsymbol{L}b$	b	$\boldsymbol{L}c$	c
m_1	1	1	1	0	0	1
m_2	1	1	1	1	1	0
m_3	0	0	1	0	0	1
m_4	0	0	1	0	1	0

The formula representation of such a theory is of little relevance in our discussion and thus will not be given here (but it can be easily obtained, for example, as a disjunctive normal form).

Note that m_1 and m_2 are minimal models of $A \wedge \boldsymbol{L}a$ and m_3 and m_4 are minimal models of $A \wedge \neg\boldsymbol{L}a$. In these four *P*-minimal models, the status of any objective atom cannot be determined. Thus $WFT(A) = \varnothing$.

Now consider $AEL(circ)$. Any minimal model of a theory must be one of its *P*-minimal models. Clearly, m_3 and m_4 are the only minimal models of A. But unfortunately any expansion based on m_3 and m_4 will be inconsistent, due to the fact that $\boldsymbol{L}b$ is true and b is false. This eliminates the possibility of forming an $AEL(circ)$ expansion based on these two models. However, in $AEL(circ)$ (as well as in Moore's AEL), an expansion may be "ungrounded", i.e., generated in such a way that a belief atom is assumed before its objective counterpart is concluded. This is the case for theory A above, of which the unique stable $AEL(circ)$ expansion is $A \cup \{\boldsymbol{L}a\}$. Therefore, the intersection of all stable $AEL(circ)$ expansions of a theory may not be equivalent to its well-founded circumscriptive semantics. This is due to the combined effect of "eager" minimization, i.e., minimizing all propositions, and tolerance of "ungrounded" expansions. □

The well-founded model semantics and the stable model semantics of a logic program coincide if the program has a two-valued well-founded semantics. Not surprisingly, this result can be extended to autoepistemic theories, as shown below.

Theorem 5.3. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory such that any objective atom ξ is either true or false in the well-founded circumscriptive semantics. Then $CIRC(A(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P)$ and $CIRC(A(P, \boldsymbol{L}P); P) \wedge WFT(A)$ are logically equivalent.

Proof: Assume A is such a theory and let $\Gamma_1 = CIRC(A(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P)$ and $\Gamma_2 = CIRC(A(P, \boldsymbol{L}P); P) \wedge WFT(A)$. Then for each objective atom ξ, we have either $\Gamma_2 \models \xi$ or

$\Gamma_2 \models \neg\xi$, and therefore, either $\boldsymbol{L}\xi$ or $\neg\boldsymbol{L}\xi$ is in $WFT(A)$. It follows that Γ_2 implies $\xi \leftrightarrow \boldsymbol{L}\xi$ for every objective atom ξ, which in turn shows that Γ_2 implies $P \equiv \boldsymbol{L}P$. On the other hand, for any ξ such that $\Gamma_1 \models \boldsymbol{L}\xi$, we can similarly show that $WFT(A) \models \boldsymbol{L}\xi$, and hence Γ_1 implies $WFT(A)$. Therefore, Γ_1 and Γ_2 are logically equivalent. □

Just like Moore's AEL or Przymusinski's $AEL(circ)$, not every autoepistemic theory has a consistent well-founded circumscriptive semantics. For example, with the theory $A = \{\boldsymbol{L}p\}$, since $CIRC(A; \{p\})$ implies $\neg p$ but $\boldsymbol{L}p$ is contained in A, the well-founded circumscriptive semantics of A yields an inconsistent theory.

However, the following theorem shows the application range of the well-founded circumscriptive semantics is at least as large as either AEL or $AEL(circ)$.

Theorem 5.4. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory. The well-founded circumscriptive semantics of A is consistent if (a) A has a consistent stable expansion, or (b) A has a stable $AEL(circ)$ expansion.

Proof: (a) Let T be a consistent stable expansion of A, **m** be a minimal model of T, and $M_{\boldsymbol{L}P} = \{\boldsymbol{L}\xi \mid \xi \in T\} \cup \{\neg\boldsymbol{L}\xi \mid \xi \notin T\}$. Then, T is logically equivalent to $A \wedge M_{\boldsymbol{L}P}$, for $T = \{\phi \mid A \wedge M_{\boldsymbol{L}P} \models \phi\}$. Thus, **m** is a minimal model of $A \wedge M_{\boldsymbol{L}P}$, which in turn implies that **m** is a P-minimal model of A, i.e., **m** is a model of $CIRC(A; P) \wedge M_{\boldsymbol{L}P}$. By Theorem 5.1 (i), we have $M_{\boldsymbol{L}P} \models WFT(A)$. It follows that **m** is a model of $CIRC(A; P) \wedge WFT(A)$.

(b) The proof is similar to (a), using Theorem 5.1 (ii), and thus is omitted. □

5.2. Relation with Przymusinski's Construction and the Well-Founded Model Semantics

We first show that for logic programs our well-founded construction coincides with Przymusinski's construction.

Recall that in the well-founded construction circumscription is applied only once at the beginning. We now show that the semantics remains unchanged, even if circumscription is applied at every step of the construction. Likewise, in Przymusinski's construction it is not necessary to apply circumscription in every step of the construction.

Proposition 5.5. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory and Φ be a set of belief literals. Then

(i) $CIRC(A; P) \wedge \Phi \equiv CIRC(A \wedge \Phi; P)$

(ii) $CIRC(A) \wedge \Phi \equiv CIRC(A \wedge \Phi)$.

Proof: We only need to show (i) since (ii) is a special case of (i).

($\Rightarrow$) We show that a model of the left hand side is a model of the right hand side.

A model **m** of the left hand side is a *P*-minimal model of A in which Φ is also satisfied. Thus it is a model of $A \wedge \Phi$. It is also a *P*-minimal model of $A \wedge \Phi$, since if not, there would exist a model **n** of $A \wedge \Phi$ such that **n** and **m** have the same assingment to the belief atoms and **n** $\subset$ **m**. Thus, **n** is a model of A and *P*-smaller than **m**. This contradicts the fact that **m** is *P*-minimal of A.

($\Leftarrow$) The proof is similar and thus omitted. □

Now we can rewrite the definition of the well-founded circumscriptive semantics in a style similar to that of Przymusinski's.

Corollary 5.6. For an autoepistemic theory $A(P, \boldsymbol{L}P)$, the semantics defined by the following iterative construction is equivalent to the well-founded circumscriptive semantics defined in Definition 5.2:

$$W^0 = CIRC(A; P)$$
$$W^{n+1} = W^n \cup \{\boldsymbol{L}\xi \mid W^n \models_{circ(P)} \xi\} \cup \{\neg\boldsymbol{L}\xi \mid W^n \models_{circ(P)} \neg\xi\}$$

where $W^n \models_{circ(P)} \xi$ means the objective atom ξ is true in all *P*-minimal models of W^n and $W^n \models_{circ(P)} \neg\xi$ means the objective atom ξ is false in all *P*-minimal models of W^n.

Likewise, Przymusinski's construction can be equivalently defined as:

$$T_0 = A$$
$$T_{n+1} = T_n \cup \{\boldsymbol{L}\xi \mid T_n \wedge CIRC(A) \models \xi\} \cup \{\neg\boldsymbol{L}\xi \mid T_n \wedge CIRC(A) \models \neg\xi\}. \quad \square$$

Notice that the inductive bases are different between ours and Przymusinski's; the former starts from $CIRC(A; P)$ and the latter begins with the initial theory A. The model-theoretic semantics of the two are therefore different with the former being strictly stronger (there are theories for which the former contains more theorems than the latter). However, we can associate logic program semantics with the truth status of the belief literals so that this difference plays no significant role (cf. Theorem 5.8).

With Corollary 5.6, to show that our well-founded construction coincides with Przymusinski's, we only need to show

Lemma 5.7. Let $\Pi(P, \boldsymbol{L}P)$ be an autoepistemic logic program and ξ be an atom. Then $CIRC(\Pi; P) \models \xi$ iff $CIRC(\Pi) \models \xi$ and $CIRC(\Pi; P) \models \neg\xi$ iff $CIRC(\Pi) \models \neg\xi$.

Proof: The first equivalence is trivial for both are equivalent to $\Pi \models \xi$.

Since each minimal model of Π is also its P-minimal model, we have $CIRC(\Pi) \models \neg\xi$ if $CIRC(\Pi; P) \models \neg\xi$.

For the converse, assume $CIRC(\Pi) \models \neg\xi$ but $CIRC(\Pi; P) \not\models \neg\xi$, for some objective atom ξ. Then, there exist two models **m** and **n** such that

(1) **m** is a model of $CIRC(\Pi; P)$,

(2) **n** is a model of $CIRC(\Pi)$,

(3) $\xi \in \mathbf{m} - \mathbf{n}$, and

(4) $\mathbf{n} \subset \mathbf{m}$.

Let

$M_P = \{\zeta \mid \zeta$ is an objective atom in **m**$\}$

$N_P = \{\zeta \mid \zeta$ is an objective atom in **n**$\}$

$M_{\boldsymbol{L}P} = \{\boldsymbol{L}\zeta \mid \boldsymbol{L}\zeta \in \mathbf{m}\}$

$N_{\boldsymbol{L}P} = \{\boldsymbol{L}\zeta \mid \boldsymbol{L}\zeta \in \mathbf{n}\}$.

Then, we have $N_P \subset M_P$ for $\xi \in \mathbf{m} - \mathbf{n}$ and $\mathbf{n} \subset \mathbf{m}$.

Let Π_M and Π_N be the logic programs obtained, respectively, from Π by replacing each belief atom with its truth value according to $M_{\boldsymbol{L}P}$ and $N_{\boldsymbol{L}P}$. Then, M_P and N_P are models of Π_M and Π_N respectively.

Since Π is a logic program, all belief literals are negative in the bodies of the clauses. Therefore, any clause in Π with a belief atom replaced by *true* becomes a tautology.

Since $\mathbf{n} \subset \mathbf{m}$, that a belief atom $\boldsymbol{L}\zeta$ is true in $N_{\boldsymbol{L}P}$ implies that it is also true in $M_{\boldsymbol{L}P}$. Thus, we have $\Pi_N \models \Pi_M$. It follows that N_P is also a model of Π_M. Therefore, $N_P \cup M_{\boldsymbol{L}P}$ is a model of Π. However, we have $\mathbf{m} = M_P \cup M_{\boldsymbol{L}P}$ and $N_P \subset M_P$, which contradicts that **m** is a P-minimal model of Π. □

With the above lemma and Corollary 5.6, and the fact that Przymusinski's construction yields the least stable $AEL(circ)$ expansion which coincides with the well-founded model of a logic program, we have

Theorem 5.8. Let Π be a logic program and $\Pi'(P, \boldsymbol{L}P)$ be its autoepistemic translation. Let $WFT(\Pi')$ be the well-founded constraint theory of $\Pi'(P, \boldsymbol{L}P)$. Then $WFT(\Pi')$ is exactly the set of belief literals generated by Przymusinski's construction method. That is, if we let T =

$\{\xi \mid \boldsymbol{L}\xi \in WFT(\Pi')\}$ and $F = \{\xi \mid \neg\boldsymbol{L}\xi \in WFT(\Pi')\}$, then $<T, F>$ is the (3-valued) well-founded model of Π, where T is the set of atoms assigned to *true* and F the set of the atoms assigned to *false*. □

6. Final Remarks

The semantics of an autoepistemic logic is based on the notion of stable expansions, which in the past have typically been defined by nonconstructive fixed point equations. A semantics defined this way appears to bear more irregular behaviors in their inferences than a semantics that can be defined more concretely.

The proposal in this paper is an attempt towards more concrete, flexible definitions of autoepistemic logics which give rise to semantics of autoepistemic logics that behavior more regularly and may suit possibly different applications. The stable circumscriptive semantics defined in this paper has been shown to be more regular than Moore's *AEL* in that whenever a stable expansion is epistemically coherent there is a corresponding model by the stable circumscriptive semantics. The well-founded circumscriptive semantics proposed here coincides with Przymusinski's iterative construction for logic programs. However, a crucial difference is that the former is applicable to more general autoepistemic theories.

Belief formulas in this paper are restricted to belief literals where the belief symbol $\boldsymbol{L}$ may only appear in front of an atomic objective proposition. One possible approach to accommodating general autoepistemic theories is to tactically replace a belief formula $\boldsymbol{L}\phi$ by $\boldsymbol{L}\sigma \wedge \sigma \leftrightarrow \phi$, where σ is a new proposition symbol, and employ prioritized circumscription in such a way that $\boldsymbol{L}\phi$ is minimized with a lower priority than ϕ. This approach is worth future investigation.

Acknowledgement:

We would like to thank the two anonymous referees for their valuable comments that helped improve the presentation of the paper. We also benefited from a discussion with Przymusinski on the results of this paper. The work is partly supported by grants from the Natural Sciences and Engineering Research Council of Canada.

APPENDIX: Proofs of Theorems

Theorem 4.1. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory and

$$AEC = CIRC(A(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P).$$

Then

(1) for any epistemically coherent stable expansion T of A, there exists a model **m** of AEC such that **m** is the unique autoepistemic model of T; and

(2) for any epistemically coherent minimal stable expansion T of A, there exists a minimal model **m** of AEC such that **m** is the unique autoepistemic model of T.

We need two lemmas first.

Lemma A.1 Let A be an autoepistemic theory, T be an epistemically coherent stable expansion of A, and **m** be an autoepistemic model of T. Then

(1) for any objective atom ξ, ξ is in T if and only if ξ is in **m**, and

(2) **m** is the unique autoepistemic model of T.

Proof: (1) Let ξ be an objective atom. If ξ is in T then obviously ξ is in **m** for **m** is a model of T. Assume ξ is in **m** but not in T. By Definition 2.2, $\neg\boldsymbol{L}\xi$ is in T, and therefore, $\boldsymbol{L}\xi$ is not in **m**, which contradicts the fact that **m** is an autoepistemic model of T.

(2) For any autoepistemic model **n** of T, by (1),

$$\mathbf{n} = \{\xi \mid \xi \text{ is an objective atom in } T\} \cup \{\boldsymbol{L}\xi \mid \xi \text{ is an objective atom in } T\},$$

which implies T has a unique autoepistemic model. □

Lemma A.2. Let A be an autoepistemic theory,

$$AEC = CIRC(A(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P),$$

and **m** be a model of AEC. Then, $S_{\mathbf{m}} = A \cup \mathbf{m} \cup \{\neg\boldsymbol{L}\xi \mid \boldsymbol{L}\xi \notin \mathbf{m}\}$ is a stable set†.

Proof: Let $M_{\boldsymbol{LP}} = \{\neg\boldsymbol{L}\xi \mid \boldsymbol{L}\xi \notin \mathbf{m}\}$. Then $S_{\mathbf{m}} = A \cup \mathbf{m} \cup M_{\boldsymbol{LP}}$. Since **m** is a model of AEC, $\mathbf{m} \cup M_{\boldsymbol{LP}}$ can be viewed as a model of AEC. It follows that $S_{\mathbf{m}} \not\models \xi$ for any objective atom $\xi \notin \mathbf{m}$. Therefore, for any objective atom ξ, $S_{\mathbf{m}} \models \xi$ iff $\xi \in \mathbf{m}$ iff $\boldsymbol{L}\xi \in \mathbf{m}$. It follows that $\boldsymbol{L}S_{\mathbf{m}} \subseteq S_{\mathbf{m}}$ and $\neg\boldsymbol{L}S_{\mathbf{m}} \subseteq S_{\mathbf{m}}$. □

The above lemma shows that each model of AEC corresponds to a stable set that contains A.

† Note that we identify $S_{\mathbf{m}}$ with its closure of logic consequences.

Proof of Theorem 4.1: (1) Assume T is an epistemically coherent expansion of A. By Lemma A.1 (2), T has a unique autoepistemic model. Let **m** be such an autoepistemic model of T. Then **m** is a model of $P \equiv \boldsymbol{L}P$. We need only to show that **m** is also a model of $CIRC(A; P)$. Assume not, that is, **m** is not a *P*-minimal model of A. Since **m** is a model of A, there exists another model **n** of A such that (a) $\mathbf{n} \subset \mathbf{m}$, and (b) any belief atom $\boldsymbol{L}\xi$ has the same value in both **m** and **n**. Therefore, **n** is a model of $A \cup \boldsymbol{L}T \cup \neg\boldsymbol{L}\overline{T}$. Let T' be $\{\boldsymbol{L}\zeta \mid \boldsymbol{L}\zeta \in \mathbf{n}\} \cup \{\neg\boldsymbol{L}\zeta \mid \boldsymbol{L}\zeta \notin \mathbf{n}\}$, then **n** is a model of $A \cup T'$. Since **n** and **m** have the same interpretation for any belief literal, i.e., T' is the same as $\{\boldsymbol{L}\zeta \mid \boldsymbol{L}\zeta \in \mathbf{m}\} \cup \{\neg\boldsymbol{L}\zeta \mid \boldsymbol{L}\zeta \notin \mathbf{m}\}$, we have $T = \{\phi \mid A \cup T' \models \phi\}$. Thus, for any objective atom ξ in T, ξ is also contained in **n**, which, by Lemma A.1 (1), implies that for any objective atom ξ in **m** ξ is also in **n**. It follows that $\mathbf{m} \subseteq \mathbf{n}$, which contradicts the assumption that $\mathbf{n} \subset \mathbf{m}$.

(2) Assume T is a minimal stable expansion of A, **m** be a model of AEC but not a minimal model of AEC such that **m** is the unique autoepistemic model of T. Then, there exists a minimal model of AEC such that $\mathbf{n} \subset \mathbf{m}$. Let $S_{\mathbf{n}} = A \cup \mathbf{n} \cup \{\neg\boldsymbol{L}\zeta \mid \boldsymbol{L}\zeta \notin \mathbf{n}\}$. By Lemma A.2, S_n is a stable set containing A. However, $S_{\mathbf{n}o} = \{\zeta \mid \zeta \in \mathbf{n}\}$, and $T_o = \{\zeta \mid \zeta \in \mathbf{m}\}$. Since $\mathbf{n} \subset \mathbf{m}$, we have $S_{\mathbf{n}o} \subset T_o$, which contradicts the assumption that T is a minimal stable expansion. □

Theorem 4.2. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory and

$$AEC = CIRC(A'(P, \boldsymbol{L}P); P) \wedge (P \equiv \boldsymbol{L}P),$$

where $A' = A \cup \{\boldsymbol{L}\xi \leftarrow \xi \mid \xi \in P\}$. Then

(1) T is a consistent stable expansion of A' if and only if there exists a model **m** of AEC such that **m** is the unique autoepistemic model of T;

(2) T is a consistent minimal stable expansion of A' if and only if there exists a minimal model **m** of AEC such that **m** is the unique autoepistemic model of T.

Proof: (1) ($\Rightarrow$) We show that a consistent stable expansion of A' is always epistemically coherent. Then the conclusion follows from Theorem 4.1 (1).

Let T be a consistent stable expansion of A'. Then T has a model **m**. For each objective atom ξ, if $\xi \in T$ then $\boldsymbol{L}\xi \in T$, and thus both ξ and $\boldsymbol{L}\xi$ are in **m**. If $\xi \notin T$ then $\neg\boldsymbol{L}\xi \in T$ and $\neg\xi \in T$ due to $\boldsymbol{L}\xi \leftarrow \xi$; thus neither ξ nor $\boldsymbol{L}\xi$ is in **m**. Thus, **m** is an autoepistemic model of T and T is epistemically coherent.

($\Leftarrow$) Let **m** be a model of AEC. Define

$M_P = \{\xi \mid \xi$ is an ordinary atom true in **m**$\}$,

$M_{\boldsymbol{L}P} = \{\boldsymbol{L}\xi \mid \xi \in \mathbf{m}\} \cup \{\neg\boldsymbol{L}\xi \mid \xi \notin \mathbf{m}\}$,

and let $M = M_P \cup M_{\boldsymbol{L}P}$. Then **m** is the only autoepistemic model of M. Further, M is an epistemically coherent stable expansion of itself, because for each ordinary atom ξ, ξ is in M iff $\boldsymbol{L}\xi$ is in M, ξ is not in M iff $\neg\boldsymbol{L}\xi$ is in M, and $M \models \xi$ iff ξ is in M. In addition, since any stable expansion of M must contain M and does not contain any ordinary atom not in M_P, M is the only stable expansion of itself. Therefore, we need only to show that $M \cup A'$ is also a stable expansion of A'.

Since $M_{\boldsymbol{L}P} = \boldsymbol{L}(M \cup A') \cup \neg\boldsymbol{L}(\overline{M \cup A'})$, by Definition 2.2, $M \cup A'$ is a stable expansion of A' iff $M = \{\phi \mid A' \cup M_{\boldsymbol{L}P} \models \phi\}$. Thus, to show that $M \cup A'$ is a stable expansion of A', we need only to show that $A' \cup M_{\boldsymbol{L}P} \models M_P$. Let Ω be obtained from A' by replacing each belief atom with their truth value according to $M_{\boldsymbol{L}P}$. Then Ω is an ordinary theory, and $A' \cup M_{\boldsymbol{L}P} \models M_P$ if M_P is the unique minimal model of Ω.

Clearly, M_P is a minimal model of Ω since **m** is a *P*-minimal model of A', i.e., M_P is the smallest model of A' with the fixed assignment $M_{\boldsymbol{L}P}$ to the belief atoms. Assume M_P is not unique, i.e., there exists another minimal model N of Ω. Then for any $a \notin M_P$, we have $\neg\boldsymbol{L}a \in M_{\boldsymbol{L}P}$, and since $A' \models \boldsymbol{L}\xi \leftarrow \xi$ for each proposition symbol ξ, we have $A' \cup M_{\boldsymbol{L}P} \models \neg a$, and thus $a \notin N$, i.e., $N \subset M_P$, contradicting the assumption.

(2) It follows from (1) and Theorem 4.1 (2). □

Theorem 5.1. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory and $WFT(A)$ be the well-founded constraint theory of A. Then

(i) $WFT(A)$ is contained in every stable expansion of A;

(ii) $WFT(A)$ is contained in every stable $AEL\,(circ)$ expansion of A.

We first show a utility lemma.

Lemma A.3. Let $A(P, \boldsymbol{L}P)$ be an autoepistemic theory and ξ be an objective atom, and m_1 and m_2 be sets of belief literals such that $m_1 \subseteq m_2$. Then

(a) $A \wedge m_2 \models \xi$ if $CIRC(A; P) \wedge m_1 \models \xi$;

(b) $CIRC(A \wedge m_2) \models \neg\xi$ if $CIRC(A; P) \wedge m_1 \models \neg\xi$

Proof: When either m_1 or m_2 is inconsistent, both conclusions trivially hold. Suppose m_1 and m_2 are both consistent.

(a) Let m be a minimal model of $A \wedge m_2$. Then, it is sufficient to show that m is also a model of $CIRC(A; P) \wedge m_1$. Since m is a minimal model of $A \wedge m_2$, and m_2 is a set of belief literals, m is a P-minimal model of A, i.e., a model of $CIRC(A; P)$. Therefore, m is a model of $CIRC(A; P) \wedge m_1$, for $m_2 \models m_1$.

(b) Since $m_1 \subseteq m_2$, any model of $CIRC(A \wedge m_2)$ i.e. any minimal model of A with a fixed assignment to m_2 is a model of $CIRC(A; P) \wedge m_1$. □

Proof of Theorem 5.1:

(i) Let T be a stable expansion of A and $M_{\mathbf{L}P} = \{\mathbf{L}\zeta \mid \zeta \in T\} \cup \{\neg\mathbf{L}\zeta \mid \zeta \notin T\}$. We show the claim by induction on W_i.

Basis: $W^0 \subseteq T$ for W^0 is empty.

Inductive Step: Assume $W^n \subseteq T$. Then $W^n \subseteq M_{\mathbf{L}P}$. Consider W^{n+1}. For each $\mathbf{L}\xi \in W^{n+1} - W^n$, we have

$$CIRC(A; P) \wedge W^n \models \xi$$

and for each $\neg\mathbf{L}\zeta \in W^{n+1} - W^n$, we have

$$CIRC(A; P) \wedge W^n \models \neg\zeta.$$

By Lemma A.3, $A \wedge M_{\mathbf{L}P} \models \xi$ and $CIRC(A \wedge M_{\mathbf{L}P}) \models \neg\zeta$. Note that $T \equiv A \wedge M_{\mathbf{L}P}$. It follows that $\xi \in T$ and $\zeta \notin T$. Thus, $\mathbf{L}\xi \in T$ and $\neg\mathbf{L}\zeta \in T$.

(ii) The proof is the same as the above. Note that $CIRC(A \wedge M_{\mathbf{L}P}) \models \neg\zeta$ implies that $\neg\mathbf{L}\zeta$ is contained in T if T is a stable $AEL(circ)$ expansion of A. □

References

1. M. Gelfond, On Stratified Autoepistemic Theories, in *Proc. AAAI-87*, 1988, 207-211.

2. M. Gelfond and V. Lifschitz, The Stable Model Semantics for Logic Programming, in *Proc. 5th Symposium/Conference on Logic Programming*, R. A. Kowalski and K. A. Bowen (ed.), MIT Press, 1988, 1070-1080.

3. K. Konolige, On the Relation between Default and Autoepistemic Logic, *Artificial Intelligence Vol. 35*, (1988), 343-382.

4. V. Lifschitz, Computing Circumscription, *Proc. IJCAI-85*, Los Angeles, CA, 1985, 121-127.

5. V. Lifschitz, Between Circumscription and Autoepistemic Logic, *Proc. First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, Ont., 1989, 235-244.

6. J. McCarthy, Circumscription—A Form of Non-monotonic Reasoning, *Artificial Intelligence Vol. 13*, (1980), 27-39.

7. J. McCarthy, Applications of Circumscription to Formalizing Common-Sense Reasoning, *Artificial Intelligence Vol. 28*, (1986), 86-116.

8. R. C. Moore, Semantical Considerations on Nonmonotonic Logic, *Artificial Intelligence Vol. 25*, (1985), 75-94.

9. T. C. Przymusinski, Autoepistemic Logics of Closed World Beliefs and Logic Programming, in *Proc. Workshop on Nonmonotonic Reasoning and Logic Programming*, MIT Press, July 1991, 3-20.

10. T. C. Przymusinski, Semantics of Disjunctive Logic Programs and Deductive Databases, in *Proc. the 2nd International Conference on Object Oriented and Deductive Databases*, Dec. 1991, 87-107.

11. R. Reiter, On Closed-World Data Bases, in *Logic and Data Bases*, H. Gallaire and J. Minker (ed.), Plenum Press, 1978.

12. R. C. Stalnaker, A Note on nonmonotonic Modal Logic, Tech Report, Department of Philosophy, Cornell University, 1980 .

13. L. Yuan and J. You, Discriminant Circumscription, in *Proc. 10th Conference on Foundations of Software Technology and Theoretical Computer Science, LNCS 472*, Dec. 1990, 21-32.