An Introspective Logic of Belief

Li-Yan Yuan and Jia-Huai You
Department of Computer Science
University of Alberta
Edmonton, Canada T6G 2H1
{yuan, you}@cs.ualberta.ca

Abstract

This paper investigates the relationship between negative introspection and the expressive power of autoepistemic logic. We first propose a new approach to negative introspection based on classical entailment and then define an introspective logic of belief which is a typical modal logic KDN augmented with a new negative introspection rule. We demonstrate that the logic is capable of characterizing almost all nonmonotonic reasoning semantics and can be easily implemented.

keywords: nonmonotonic reasoning, introspection, autoepistemic logic

1 Introduction

Recently considerable attention has focussed on using autoepistemic logic of belief as a general framework for nonmonotonic reasoning [18, 10, 24, 15, 20, 22, 16, 28, 1, 13]. This is mainly because a nonmonotonic reasoning process can be viewed as an introspective reasoning processing: A rational agent derives conclusions based on his belief, and if the agent's perspective of the world changes he would have to change his conclusions according to revised belief.

Example 1.1 *Consider the following familiar scenario:*

a bird usually flies unless it is abnormal;
a bird either flies or is abnormal;
tweety is a bird.

In Reiter's (propositional) default logic [23], the scenario can be characterized by the following default theory D

$$bird(tweety);$$
$$fly(tweety) \vee abnormal(tweety) \subset bird(tweety);$$
$$\frac{bird(tweety) : \neg abnormal(tweety)}{fly(tweety)}.$$

As expected, D has a unique stable extension which concludes $fly(tweety)$, i.e., tweety flies.

The scenario can also be represented by an autoepistemic logic theory A of belief as follows [18]:

$$bird(tweety);$$
$$fly(tweety) \vee abnormal(tweety) \subset bird(tweety);$$
$$fly(tweety) \subset bird(tweety) \wedge \neg \mathbf{B} abnormal(tweety),$$

where $\mathbf{B}abnormal(tweety)$ stands for believing $abnormal(tweety)$. Since $abnormal(tweety)$ cannot be derived from A in any circumstance, by negative introspection, a rational agent would not believe *abnormal(tweety) and therefore, hold disbelief* $\neg\mathbf{B}abnormal(tweety)$*, which implies that tweety flies. This is reflected in the fact that the unique stable extension of A (in the sense of Moore [18]) implies* $\neg\mathbf{B}anormal(tweety)$ *and* $fly(tweety)$*.*

The nature of nonmonotonicity of introspection is demonstrated by the fact that $fly(tweety)$ *cannot be concluded if* $abnormal(tweety)$ *is added into A.* □

Negative introspection plays an essential role in characterizing nonmonotonic reasoning semantics. Initially, negative introspection is based on classical entailment: Deriving $\neg\mathbf{B}\alpha$ (standing for α is disbelieved) if α is not a logical consequence of the concerned theory. Typical examples are Reiter's default logic and Moore's autoepistemic logic, as demonstrated in the above example. Negative introspection based on classical entailment, however, often leads to (unnecessary) inconsistency because more disbeliefs may be concluded than desired. Consider the simple example.

Example 1.2 *Let A be defined by*

$$fly \subset \neg\mathbf{B}fly.$$

Since A does not imply fly*, in the sense of classical entailment, by negative introspection based on classical entailment, we may derive* $\neg\mathbf{B}fly$*. The derivation of* $\neg\mathbf{B}fly$*, however, implies* fly*, which effectively falsifies the negative introspection.* □

The problem of excessive introspecting, demonstrated in the above example, is well-understood and leads to the introduction of negative introspection based on minimal-model entailment: Deriving $\neg\mathbf{B}\alpha$ if $\neg\alpha$ is true in all minimal models (with fixed beliefs) of the concerned theory [22, 28]. The difficulty of implementing minimal-model entailment, however, is one of the main obstacles to the implementation of nonmonotonic reasoning systems based on autoepistemic logic with minimal-model introspection [19].

1.1 Introspection with respect to all possible belief worlds

Instead of using minimal-model-entailment, we are going to take a different approach to resolving the excessive introspection problem of negative introspection based on classical entailment, as demonstrated in Example 1.2.

Example 1.3 *(Example 1.2 continued)*

A, representing the knowledge base of a rational agent, contains one belief atom $\mathbf{B}fly$*, and, because of incomplete information, the value of* $\mathbf{B}fly$ *is uncertain: There are two possible assignments to values of belief atoms in A, i.e,* $I_1 = \{\mathbf{B}fly\}$ *and* $I_2 = \{\neg\mathbf{B}fly\}$*. Under* I_1*, A becomes* $A_1 = \{fly \vee TRUE\}$*, and under* I_2*, A becomes* $A_2 = \{fly\}$*. Obviously,* A_2 *implies* fly *while* A_1 *does not.*

Because fly *can be derived from A with the assumption of* I_2 *– a possible value assignment of belief atoms in A,* $\neg\mathbf{B}fly$ *should not be derived under negative introspection, despite the fact that* fly *cannot be derived from A. This, naturally, resolves the excessive introspecting problem of negative introspection based on classical entailment.* □

Assume a rational agent uses an autoepistemic theory A as his knowledge base. Then each assignment of values of belief atoms in A represents a possible belief world in the mind of the agent. Because introspection is a process of determining appropriate values for belief atoms according to the knowledge base, we believe introspection should be carried out with respect to to each and every possible belief world of the agent, as demonstrated in the above example. That is, we believe negative introspection should be the following process:

Deriving $\neg \mathbf{B}\alpha$ if α cannot be derived from the concerned theory with respect to each and every possible assignment of values of belief atoms contained in the theory.

For simplicity, we use p-introspection to refer to introspection with respect to all possible belief worlds.

P-introspection was first proposed for autoepistemic logic and logic programs in our earlier works [27, 28], and has been used to characterize the declarative semantics of disjunctive logic programs [20, 22, 21, 3, 4]. However, the previous work has mainly focussed on minimal-model-entailment, which is inherently difficult to compute.

The idea of p-introspection is also shared by the argumentation framework and the acceptability semantics [6, 14]. Introspection with respect to all possible belief worlds is very much similar to the justification of an argument with respect to all consistent arguments.

The difficulty of negative introspection is largely due to incomplete characterization of belief and disbelief of a rational agent's knowledge base. Since each belief assignment represents a possible belief world that is completely characterized, p-introspection is a combination of introspection with respect to each and every belief world, which is intuitive and natural. It is, thus, very interesting to investigate the relationship between p-introspection based on classical entailment and the expressive power of autoepistemic logic.

1.2 Introspective logic of belief

In this paper, we define an introspective logic of belief, which is a classical modal logic KDN augmented with a negative introspection rule NI, based on p-introspection with classical-entailment. The expressive power of the logic is demonstrated by the fact that it can be used to characterize all well-known nonmonotonic reasoning semantics, including default logic, various forms of closed world assumption, parallel circumscription, and declarative semantics of logic programs.

The introspective logic presents a new approach to the implementation of nonmonotonic reasoning systems. One obvious advantage is its enormous expressive power: a proof procedure for the logic can be used for almost any nonmonotonic reasoning semantics. Another advantage is a proof procedure for the logic can be implemented using inference engines for classical entailment. Other particular contributions of the new approach include:

1. The logic specifies a theorem proving method for minimal model reasoning, which provides an alternative approach to minimal-model generators.
2. The logic can be used to define parallel circumscription with varying predicates. Despite a common belief that parallel circumscription with varying predicates are more difficult to compute than that without, our result demonstrates the contrary [19].
3. The logic provides a model-logic based unifying framework for default reasoning [2].
4. The logic specifies a natural skeptical semantics for disjunctive logic programs which can be efficiently implemented. In fact, we show that the computational complexity for the proposed skeptical semantics is P^{NP} which means there exists a polynomial inference algorithm using an oracle for classical entailment. Note that the lower complexity bound for all existing semantics of disjunctive logic programs is Π_2^P [7].

Our results not just present a new approach to the implementation of nonmonotonic reasoning systems, but also provides very much needed insights toward understanding the relationship between the nonmonotonic reasoning semantics and the underlying negative introspection.

The paper is organized as follows. Section 2 presents background material. Section 3 reviews various existing approaches to introspection. P-introspection based on classical entailment and the proposed introspective logic of belief are presented in Section 4 and 5. The expressive power of the logic is demonstrated in Section 6.

2 Language and Syntax

2.1 Language

We consider here a propositional language augmented with a modal operator $\mathbf{B}$. An atomic formula (atom) is either a propositional symbol, or an epistemic atom, also called *belief atom*, $\mathbf{B}\alpha$, where α is a (well-formed) formula defined as usual. The intended meaning of $\mathbf{B}\alpha$ is "α is believed". For convenience, we also use $\mathbf{not}\alpha$, called *disbelief atom*, interchangeably for $\neg\mathbf{B}\alpha$, meaning α is disbelieved. ($\mathbf{not}\alpha$ is also viewed by many authors as a *default negation.*) An *belief theory* (or a theory for short) is a set of well-formed formulae, and a formula (or a theory) is *objective* if it contains no epistemic atoms, otherwise, it is *subjective*. We denote by $P^+(A)$ and $P^-(A)$ the set of all propositions and the set of all negative literals that appear in A, and by $B^+(A)$ and $B^-(A)$ the set of all belief atoms and the set of all disbelief atoms that appear in A, respectively.

The logic has the following axioms and rules of inference.

Axioms.

PL. All propositional tautologies.

K. $\mathbf{B}(\alpha \supset \beta) \supset (\mathbf{B}\alpha \supset \mathbf{B}\beta)$.

D. $\neg\mathbf{B}$ (false)

Inference rules.

Modus Ponens (MP). $$\frac{\alpha \supset \beta, \alpha}{\beta}$$

A rational agent does not belief inconsistent conclusions which is expressed by D. K means that if a conditional and its antecedent are both believed, then so is the consequent. The importance of K is evidenced by the fact that K imposes a constraint of normality on the language: $\mathbf{B}\alpha \equiv \mathbf{B}\beta$ whenever $\alpha \equiv \beta$. (Note that by $\alpha \equiv \beta$ we mean $(\alpha \subset \beta) \wedge (\beta \subset \alpha)$.) MP is a usual inference rule for propositional logic.

Let A be a theory and α a formula. By $A \vdash_{KD} \alpha$ we mean α can be derived from A based on the aforementioned axioms and rules of inference. A is inconsistent if there exists a formula α such that $A \vdash_{KD} \alpha$ and $A \vdash_{KD} \neg\alpha$; otherwise, it is consistent.

2.2 Belief Interpretation

A belief theory A is used to describe the knowledge base of a rational agent. Due to incomplete information, an agent may have to hold a set of possible states of epistemic belief, each of which represents a complete description about the agent's belief. A *belief interpretation* is thus introduced to characterize such a complete state of belief. Formally,

Definition 2.1 *1. A* belief interpretation *of A is a set I of belief atoms and disbelief atoms such that for any formula α in the language of A, either* $\mathbf{B}\alpha \in I$ *or* $\neg\mathbf{B}\alpha \in I$ *(not both).*

2. A belief model *of A is a belief interpretation I of A such that $A \cup I$ is consistent.* □

Obviously, a theory is consistent if and only if it has at least one belief model. Due to unlimited nesting levels of beliefs, a belief model is always infinite, which leads to the following notations.

A *restricted belief interpretation* of A is a set I_A of belief and disbelief atoms such that $I_A = I \cap (B^+(A) \cup B^-(A))$ for some belief interpretation I of A; and a *restricted belief model* of A is a restricted belief interpretation I_A of A such that $A \cup I_A$ is consistent. The following lemma reveals the relationships between belief models and restricted ones.

Lemma 2.2 *Assume A is a belief theory and α a formula.*

1. *There exists a restricted belief model I_A of A such that $A \cup I_A \not\vdash_{KD} \alpha$ if and only if there exists a belief model I of A such that $I_A \subseteq I$ and $A \cup I \not\vdash_{KD} \alpha$.*

2. *there exits a restricted belief model I_A of A such that $A \cup I_A \vdash_{KD} \alpha$ if and only if for any belief model I of A such that $I_A \subseteq I$, $A \cup I \vdash_{KD} \alpha$.* □

For simplicity, from now on, we shall use the belief model and the restricted belief model interchangeably, unless otherwise indicated.

Example 2.1 *Consider a simplified version of the autoepistemic theory in Example 1.1 below*

$$\begin{array}{ll} A: & bird; \\ & fly \vee abnormal \subset bird; \\ & fly \subset bird \wedge \neg \mathbf{B} abnormal. \end{array}$$

Then $I_1 = \{\mathbf{B}bird, \mathbf{B}fly, \neg\mathbf{B}abnormal, \mathbf{B}(fly \vee abnormal), \ldots\}$ is a belief model of A. But

$I_2 = \{\mathbf{B}bird, \neg\mathbf{B}fly, \neg\mathbf{B}abnormal, \mathbf{B}\neg abnormal, \mathbf{B}(fly \vee abnormal), \ldots\}$ is not a belief model of A because, by K, $\mathbf{B}(fly \vee abnormal) \wedge \mathbf{B}\neg abnormal \vdash_{KD} \mathbf{B}fly$ which is not consistent with the fact that $\neg\mathbf{B}fly \in I_2$.

Further, $I_3 = \{\mathbf{B}abnormal\}$ and $I_4 = \{\neg\mathbf{B}abnormal\}$ are only restricted belief models of A. □

Let A be a belief theory and I a belief model of A. An (objective) *perspective theory* of A, denoted by A^I, is defined as an objective theory obtained from A by replacing each belief atom in A with their corresponding truth value in I. Obviously, a belief theory may have more than one perspective theory and each of it represents the agent's perspective with respect to one restricted belief model. For example, A in the above example has two perspective theories, i.e.,

$$\begin{array}{l} A_3 = \{bird; fly \vee abnormal \subset bird\}, \text{ and} \\ A_4 = \{; bird; fly \vee abnormal \subset bird; fly \subset bird\}, \end{array}$$

corresponding to restricted belief models I_3 and I_4 respectively.

3 Introspection Revisited

In this section we first analyze existing approaches to introspection, and then outline the main idea of the proposed approach based on the analysis.

Introspection is a process of revising the agent' belief according to his perspective of the world. For example, Moore [18] uses the stable expansion T of A

$$T = \{\phi \mid A \cup \{\mathbf{B}\alpha \mid \alpha \in T\} \cup \{\neg\mathbf{B}\alpha \mid \alpha \notin T\} \vdash_{KD45} \phi\},$$

where $\vdash_{KD45}$ denotes derivation under logic KD45, to model introspective reasoning. The terms $\{\mathbf{B}\alpha \mid \alpha \in T\}$ and $\{\neg\mathbf{B}\alpha \mid \alpha \notin T\}$ express the positive and negative introspection of an agent respectively.

It is generally agreed that positive introspection is a process of concluding belief $\mathbf{B}\alpha$ if α can be derived while negative introspection is a process of concluding disbelief $\neg\mathbf{B}\alpha$ (or $\mathbf{B}\neg\alpha$) if α cannot be derived. Positive introspection is usually achieved by introducing the necessitation rule: derive $\mathbf{B}\alpha$ if α has been proved, as follows:

Necessitation (N). $\dfrac{\alpha}{\mathbf{B}\alpha}$

The interpretation of non-derivability for negative introspection, however, varies quite diversely. Two typical approaches are:

1. classical entailment based introspection:
   deriving $\neg\mathbf{B}\alpha$ if $A \not\vdash_{KD} \alpha$; and
2. minimal model entailment based p-introspection:
   deriving $\neg\mathbf{B}\alpha$ if $\neg\alpha$ is true in the every minimal model of every perspective theory of A.

The closed world assumption, default logic, and Moore's autoepistemic logic use the classical entailment based negative introspection. This approach usually results stronger negative introspection in that more disbeliefs may be concluded, and as such, many reasonable theories do not possess consistent introspective expansions. This leads to the introduction of the minimal-model based p-introspection outlined below [28, 22].

Let α be an objective formula. By $A \models_{min} \alpha$ we mean α is true in every minimal model of every perspective theory of A. Define the following rule:

(MIN). $\dfrac{\models_{min} \neg\alpha}{\neg\mathbf{B}\alpha}$

Definition 3.1 *We say A MM-introspectively implies α, denoted as $A \vdash_{MMI} \alpha$, if α can be derived from A by the following axioms and rules of inference:*

Axioms*:*	*PL, K, D*
Inference rule:	*MP, N, MIN.*

□

Example 3.1 *(Example 1.3 continued)*

With the formal definition of the perspective theory, we know now that A has two perspective theories, i.e., $A_1 = \emptyset$ and $A_2 = \{fly\}$, corresponding to its two (restricted) belief models $I_1 = \{\mathbf{B}fly\}$ and $I_2 = \{\neg\mathbf{B}fly\}$ of A respectively. Since $\neg fly$ is true in the unique minimal model of A_1 but false in the unique minimal model of A_2, $A \not\vdash_{MMI} \neg\mathbf{B}fly$ and therefore avoids the inconsistent problem of classical-entailment based introspection. □

The basic idea of the MMI is, as we have argued earlier, that introspection shall be carried out with respect to all possible belief worlds. Further, because the MMI utilizes minimal-model entailment, both positive and negative introspection can be characterized cumulatively by a single step application of minimization, and therefore, the approach has attracted considerable attention. For example, the static semantics is defined based on this approach [22, 4].

The minimal-model introspection, however, suffers from the inherent difficulties associated with minimal-model entailment.

First, minimal-model entailment, unlike classical entailment, is syntactically sensitive in that the semantics of a theory may be changed if the theory is presented differently.

Example 3.2 *Consider a simplified version of A in Example 2.1 below:*

$$B_1 = \{fly \subset \neg \mathbf{B} abnormal\}.$$

Further, we introduce a new propositional name normal, standing for $\neg abnormal$, and modify B_1 into a slightly different theory B_2 below:

$$B_2 = \{fly \subset \neg \mathbf{B} abnormal; abnormal \equiv \neg normal\}.$$

Both B_1 and B_2 express essentially the same knowledge base, except that B_2 uses an atom to name the negation of abnormal.

B_1 has two perspective theories: $B_{11} = \{fly\}$ and $B_{12} = \emptyset$, and B_2 also has two perspective theories: $B_{21} = \{fly, abnormal \equiv \neg normal\}$ and $B_{22} = \{abnormal \equiv \neg normal\}$. It is not difficult to see that $B_1 \models_{min} \neg abnormal$ but $B_2 \not\models_{min} \neg abnormal$.

Consequently, $B_1 \vdash_{MMI} \neg \mathbf{B} abnormal \wedge fly$ while neither $B_2 \vdash_{MMI} \neg \mathbf{B} abnormal$ nor $B_2 \vdash_{MMI} fly$. □

Another problem related with the minimal-model-based approach is the computational complexity: the cost for minimal-model entailment is inherently higher than that for classical entailment [8]. Take a disjunctive theory A (to be defined later) as an example. The time complexity of determining if $A \models_{min} \neg\alpha$, where α is an objective atom, is, in general, Π_2^P while that of determining if α cannot be derived from every objective perspective theory of A is only NP.

Recall two approaches to introspection below:

1. deriving $\neg \mathbf{B}\alpha$ if $A \not\vdash_{KD} \alpha$
2. deriving $\neg \mathbf{B}\alpha$ if $A \cup I \models_{min} \neg\alpha$ for every belief model I of A.

The first one ignores the very existence of more than one possible belief world, and therefore, leads to excessive introspecting if the belief world of the knowledge base is not completely characterized.

The p-introspection underlying the second approach indeed resolves the excessive introspecting problem, but the minimal-model entailment causes new problems, as demonstrated above.

Naturally, we argue that introspection shall be carried out using classical entailment with respect to each and every possible belief world (in the form of perspective theory). This leads to the key idea of this paper – the *classical entailment* based possible introspection (or p-classical introspection for short):

Deriving $\neg \mathbf{B}\alpha$ if $A \cup I \not\vdash_{KD} \alpha$ for every belief model I of A.

It turns out, not surprisingly, that the new approach not just provides a natural semantics for non-monotonic reasoning, but can also be used to characterize the minimal-model-based introspection, with help of one single axiom.

4 Introspective Logic of Belief

In this section we formally define introspective logic of belief, based on the p-classical introspection.

First we need to identify the classical entailment with respect to all possible belief worlds.

Definition 4.1 *Let A be a theory and α a formula.*

1. *$A \mid\!\sim \alpha$ if $A \cup I \vdash_{KD} \alpha$ for every belief model I of A, and*
2. *$A \sim\!\mid \alpha$ if $A \cup I \not\vdash_{KD} \alpha$ for every belief model I of A.* □

Note that if A has two belief models I_1 and I_2 such that $A \cup I_1 \vdash_{KD} \alpha$ and $A \cup I_2 \not\vdash_{KD} \alpha$ then neither $A \mid\!\sim \alpha$ nor $A \sim\!\mid \alpha$. Further, if A is inconsistent then $A \mid\!\sim \alpha$ and $A \sim\!\mid \alpha$ for every formula α. The following lemma indicates that, as far as $\mid\!\sim$ and $\sim\!\mid$ are concerned, only restricted belief models are of interest.

Lemma 4.2 *1. $A \mid\!\sim \alpha$ if and only if $A \vdash_{KD} \alpha$.*

2. $A \sim\!\mid \alpha$ if and only if $A \cup I_A \not\vdash_{KD} \alpha$ for every restricted belief model I_A of A. □

Example 4.1 (Example 2.1 continued)
Because I_3 and I_4 are the only restricted belief models of A and abnormal cannot be derived under logic KD from either $A \cup I_3$ nor $A \cup I_4$, we have $A \sim\!\mid abnormal$ □

Now we are in the position to introduce the following two rules of inference for positive and negative introspection with respect to all possible belief worlds respectively.

Positive Introspection (PI). $\dfrac{\mid\!\sim \alpha}{\mathbf{B}\alpha}$

Negative Introspection (NI). $\dfrac{\sim\!\mid \alpha}{\neg\mathbf{B}\alpha}$

PI states that deriving $\mathbf{B}\alpha$ whenever $A \mid\!\sim \alpha$ and NI that deriving $\neg\mathbf{B}\alpha$ whenever $A \sim\!\mid \alpha$. Because $A \mid\!\sim \alpha$ if and only if $A \vdash_{KD} \alpha$, PI is the same as the necessitation rule N.

The following example demonstrates that the introspection based on $\mid\!\sim$ and $\sim\!\mid$ may be obtained cumulatively.

Example 4.2 (Example 4.1 continued)
Obviously $A \mid\!\not\sim fly$ since $A \cup I_3 \not\vdash_{KD} fly$ and $I_3 = \{\mathbf{B}abnormal\}$ is a restricted belief model of A.

However, by NI, we may derive $\neg\mathbf{B}abnormal$, and consequently, $A \cup \{\neg\mathbf{B}abnormal\}$ has a unique belief model $I_4 = \{\mathbf{B}abnormal\}$ which implies that $A \cup \{\neg\mathbf{B}abnormal\} \mid\!\sim fly$. Then, by PI, we may derive $\mathbf{B}fly$. □

Before further discussion, we need to analyze monotonicity and cumulativity of inference rules.

Definition 4.3 *A relation $\vdash$ between a theory T and a formula α is*

1. monotonic *if $F \cup \{\beta\} \vdash \alpha$ whenever $F \vdash \alpha$; and*
2. cumulative *if $F \cup \{\beta\} \vdash \alpha$ whenever $F \vdash \alpha$ and $F \vdash \beta$.*

Note β is any formula. □

Obviously, PI is monotonic and therefore, logic KD augmented with PI is still monotonic. In fact, it is easy to see that logic KD augmented with PI is a modal logic, called KDN, commonly used to model knowledge and belief [12].

The following example demonstrates that NI is not monotonic.

Example 4.3 *Consider $B_1 = \{fly \subset \neg\mathbf{B}abnormal\}$ in Example 3.2 again.*

Obvious, $B_1 \sim\!\mid abnormal$ and thus, B_1 implies $\neg\mathbf{B}abnormal$ under NI. But $B_1 \cup \{abnormal\}$ does not imply $\neg\mathbf{B}abnormal$ under NI simply because $A \cup \{b\} \mid\!\sim abnormal$. □

The following theorem, however, shows that NI is cumulative.

Theorem 4.4 *NI is cumulative.* □

The significance of the theorem is rather obvious: NI can be recursively applied in any ordering, which enable us to define a logic that is nonmonotonic in general but monotonic with respect to all belief and disbelief atoms, as follows.

Definition 4.5 *Assume A is a belief theory and α a formula. Then we say A* introspectively implies *α, denoted as $A \vdash_{IKD} \alpha$ if α can be derived from A by the following axioms and rules of inference:*

Axioms:	*PL, K, D*
Inference rule:	*MP, PI, NI.*

□

Example 4.4 *Consider B_1 in Example 4.3 again. Since $B_1 \sim\mid abnormal$, $B_1 \vdash_{IKD} \neg \mathbf{B}abnormal$. Consequently, $B_1 \vdash_{IKD} fly$ as well as $B_1 \vdash_{IKD} \mathbf{B}fly$.* □

The well-defined-ness of the epistemic entailment is evidenced by the fact that $\vdash_{IKD}$ is belief-monotonic, as described below.

Definition 4.6 *A relation $\vdash$ between a belief theory T and a formula α is said to be* belief-monotonic *if for any formula β*

$$T \cup \{\mathbf{B}\beta\} \vdash \alpha \quad \textit{if} \;\; T \vdash \alpha \quad \textit{and} \quad T \cup \{\neg\mathbf{B}\beta\} \vdash \alpha \quad \textit{if} \quad T \vdash \alpha.$$

□

The introspective implication characterizes both positive and negative introspection, which is naturally nonmonotonic, but still remains belief monotonic.

Theorem 4.7 *$\vdash_{IKD}$ is belief-monotonic.* □

This theorem implies that the computation of the introspective logic can be carried out incrementally in any order, as long as the derived beliefs are preserved in the derivation.

The following example demonstrates that not every consistent theory is also consistent under introspective entailment.

Example 4.5 *Let $A = \{a \subset \neg\mathbf{B}b; \neg a \subset \neg\mathbf{B}b\}$. Since $A \sim\mid b$, $A \vdash_{IKD} \neg\mathbf{B}b$, but $A \cup \{\neg\mathbf{B}b\}$ is not consistent. Note that A is consistent for $\{\mathbf{B}b\}$ is a restricted belief model of A.* □

A theory A is said to be *introspectively consistent* if there exists no formula α such that $A \vdash_{IKD} \alpha$ and $A \vdash_{IKD} \neg\alpha$. Even though it is inherently difficult to check if a given theory is introspectively consistent, there exists a large class of theories that are introspectively consistent. For example, as discussed in Section 6, all belief theories representing disjunctive logic programs with negation are introspectively consistent.

The following observation, a direct consequence of axiom K, demonstrates the normal behavior of introspective logic. That is, for any formulae α and β,

1. $\vdash_{IKD} \mathbf{B}\alpha \wedge \mathbf{B}\beta \equiv \mathbf{B}(\alpha \wedge \beta)$,

2. $\vdash_{IKD} \mathbf{B}\alpha \vee \mathbf{B}\beta \supset \mathbf{B}(\alpha \vee \beta)$.

5 Introspective Expansions

Definition 5.1 *A belief theory T is said to be an* introspective expansion *of A if it satisfies the following fixpoint equation*

$$T = \{\phi \mid A \cup \{\neg \mathbf{B}\alpha \mid T \mathrel{\sim\mid} \alpha\} \vdash_{IKD} \phi\}. \qquad \Box$$

The introspective expansion is to characterize the introspective reasoning process by expanding a given theory A using both rules of PI and NI.

It is worth noting that only the negative introspection $\{\neg \mathbf{B}\alpha \mid T \mathrel{\sim\mid} \alpha\}$ is used in the above fixpoint equation. The use of inappropriate positive introspection in the equation, as indicated by Schwarz [25], may lead to ungrounded expansions.

Among all introspective expansions, the following three are of special interest.

Definition 5.2 *An introspective expansion T of A is said to be*

1. *the* affirmative expansion *of A if it is a subset of any introspective expansion (the least introspective expansion);*
2. *a* stable expansion *of A if T is epistemically complete in that for any formula α, T contains either $\mathbf{B}\alpha$ or $\neg\mathbf{B}\alpha$; and*
3. *a* regular expansion *of A if there exists no introspective expansion T' of A such that $T' \supset T$.*

$\Box$

Obviously, any stable expansion is a regular expansion, but not vice versa.

We use $Cn_{IKD}(A)$ to denoted the set of all formulas introspectively implied by A, i.e., $Cn_{IKD}(A) = \{\alpha \mid A \vdash_{IKD} \alpha\}$.

Example 5.1 *Consider A defined by*

$$bird;\ fly \vee abnormal \subset bird;\ fly \subset bird \wedge \neg\mathbf{B}abnormal$$

in Example 2.1 again. A has exactly one introspective expansions: $T = Cn_{IKD}(A)$, which is also a stable expansion of A. $\Box$

Example 5.2 *Consider $A = \{a \subset \mathbf{B}a, b \subset \neg\mathbf{B}a\}$. Then A has two introspective expansions, that is, $T_1 = Cn_{IKD}(A) = \{\phi | A \vdash_{IKD} \phi\}$ that contains neither $\mathbf{B}a$ nor $\neg\mathbf{B}a$, and $T_2 = Cn_{IKD}(A \cup \{\neg\mathbf{B}a\})$.*

Note that $T_3 = Cn_{IKD}(A \cup \{\mathbf{B}a\})$ is not an introspective expansion of A since $\mathbf{B}a$ cannot be derived from A with any set of disbelief atoms. $\Box$

It turns out that any theory has the affirmative (least) expansion, though not necessarily a consistent one. Furthermore, the affirmative expansion is just the set of all introspective consequences of A.

Theorem 5.3 *Assume A is a belief theory. Then*

1. *If A is introspectively consistent then any introspective expansion of A is consistent.*
2. *$T = \{\phi | A \vdash_{IKD} \phi\}$ is the affirmative expansion of A.*

$\Box$

6 Expressive Power

In this section, we demonstrate the expressive power of introspective logic.

6.1 Circumscription

Let A be a propositional theory, and P and Z are a partition of all propositions in A [1]. Assume m is a model of A, in the sense of classical propositional logic. Then m_P and m_Z denote the sets of all atoms in m that are in P and Z, respectively. A model m of A is called a *P-Z minimal* model if there exists no other model m' of A such that $m_p \subset m'_P$. Thus, $CIR(A; P; Z)$ is a formula whose models are all the P-Z minimal models of A.

Theorem 6.1 *Given a theory A and a formula α, define $B(A, P) = A \cup \{\mathbf{B}a \subset a \mid a \in P\}$. Then $CIR(A; P; Z) \vdash \alpha$ if and only if $B(A, P) \vdash_{IKD} \alpha$.* □

This theorem specifies a theorem proving method for parallel circumscription with varying predicates. Since the belief theory $B(A, P)$ contains $\mathbf{B}a \subset a$ for $a \in P$, the more predicates in A vary, the less belief atoms in $B(A, P)$, and therefore, the more efficient the theorem prover, which demonstrates that parallel circumscription with varying predicates is more efficient to compute than that without varying predicates. Furthermore, assuming the number of predicates to be minimized is small, relative to that of all predicates, the parametric complexity [5] of computing parallel circumscription is the same as that of classical entailment.

6.2 Minimal-Model-Based Introspection

We are going to show that the minimal-model-based introspection can be characterized by the p-classical introspection, augmented with an axiom. Assume A is a belief theory. Then A_{min} is defined as

$$A_{min} = A \cup \{\bar{a} \wedge \mathbf{B}\bar{a} \subset a \mid a \in P^+(A)\} \tag{6.1}$$

Note that $\bar{a}$ is a newly introduced (objective) auxiliary atom. A_{min} can also be viewed as A extended with an axiom for minimal model (MM) below:

MM. $\bar{a} \wedge \mathbf{B}\bar{a} \subset a$; for each objective atom a.

Theorem 6.2 *$A \vdash_{MMI} \alpha$ if and only if $A_{min} \vdash_{IKD} \alpha$.* □

The significants of this theorem is obviously. It demonstrates that the p-classical introspection is more expressive than the minimal-model introspection and that minimal-model entailment can be expressed by classical entailment, augmented with axiom scheme MM.

6.3 Default Logic

Let $\Delta = \langle D, W \rangle$ be a default theory, where W is a set of propositional formulas and D a set of defaults of the form

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

and the extension semantics of Δ is then characterized by its extensions [23]. We define $B(\Delta)$ as a belief theory obtained from W by adding, for each default, a formula

[1] For simplicity, we do not consider the set Q of fixed predicates here.

$$\omega \subset \mathbf{B}\mathbf{B}\alpha \wedge \neg\mathbf{B}\neg\beta_1 \wedge \cdots \wedge \neg\mathbf{B}\neg\beta_n.$$

Then we have the following

Theorem 6.3 *Let Δ be a default theory and $B(\Delta)$ its corresponding belief theory. Then E is an extension of Δ if and only if there exists a stable expansion T of $B(\Delta)$ such that E is the set of all objective formulas contained in T.* □

6.4 Logic Program Semantics

A logic program Π is set of clauses of the form

$$A_1 \vee \cdots \vee A_q \leftarrow B_1, \ldots, B_m, \mathbf{not}C_1, \ldots, \mathbf{not}C_n,$$

where A_i, B_j, C_k are objective atoms and $q \geq 1$, and can be represented by a *disjunctive theory* $A(\Pi)$ defined below:
$A(\Pi) = \{A_1 \vee \cdots A_q \subset B_1 \wedge \cdots \wedge B_m \wedge \neg\mathbf{B}C_1 \wedge \cdots \wedge \neg\mathbf{B}C_n | A_1 \vee \cdots \vee A_q \leftarrow B_1, \ldots, B_m, \mathbf{not}C_1, \ldots, \mathbf{not}C_n \in \Pi\}$.
Various semantics have been proposed in the literature for disjunctive logic programs and deductive databases, such as the well-founded, stable and the static semantics. As indicated in [7], the lower complexity bound of almost all semantics for disjunctive programs is Π_2^P. The following theorem, however, demonstrates that the complexity of inference under the affirmative semantics of disjunctive theory is P^{NP}.

Theorem 6.4 *Let $A(\Pi)$ be a disjunctive theory and α a formula. Deciding if $A(\Pi) \vdash_{IKD} \alpha$ is P^{NP}.* □

This theorem indicates that there exists a polynomial inference algorithm using an oracle for classical entailment under the affirmative semantics.

Theorem 6.5 *Assume Π is a normal program, i.e., $q = 1$ and A_i, B_j, C_k are objective atoms. Then*

1. *the well-founded model [9] of Π coincides with the affirmative expansion of $A(\Pi)$;*
2. *a stable model [11] of Π coincides with a stable expansion of $A(\Pi)$; and*
3. *a regular model [26] of Π coincides with a regular expansion of $A(\Pi)$.* □

The theorem establishes the relationship between the introspective logic of belief with major semantics of normal programs.

Since the static semantics of disjunctive programs is defined using the minimal-model based introspection, by Theorem 6.2, it is straightforward to show that the p-classical introspection is also capable of characterizing the static semantics.

The confirmative expansion of the disjunctive theory characterizes a skeptical semantics for disjunctive logic programs which is quite different from the static semantics.

Example 6.1 (Example 2.1 continued) *A is a typical disjunctive theory, and the confirmative semantics of A implies $\neg\mathbf{B}abnormal$ as well as fly. On the other hand, the static semantics of A does not imply fly due to the very existence of $fly \vee abnormal \subset bird$.* □

In fact, it is easy to show that the confirmative semantics of disjunctive theories satisfies cumulativity while the static semantics does not. Theorem 6.4 shows that the computational complexity of the confirmative semantics is lower than that of the static semantics.

6.5 Closed World Assumption

First, we define a new system of logic that does not contain K.

Definition 6.6 *Assume A is a belief theory and α a formula. Then we say A D-introspectively implies α, denoted as $A \vdash_{ID} \alpha$, if α can be derived from A by the following axioms and rules of inference:*

Axioms: *PL, D*
Inference rule: *MP, PI, NI.* $\square$

The difference between IKD and ID is comparable with that between GCWA [17], and ECWA which is equivalent to circumscription without varying predicates.

Theorem 6.7 *Let $\Pi(P)$, where P is the set of all propositions in Π, be a propositional theory, and $A(\Pi)$ be the belief theory obtained from Π by adding $\mathbf{B}p \subset p$ for every $p \in P$. Then for any objective formula α,*

1. *α is true in $GCWA(\Pi)$ if and only if $A(\Pi) \vdash_{ID} \alpha$;*
2. *α is true in $ECWA(\Pi)$ if and only if $CIR(\Pi; P) \models \alpha$ if and only if $A(\Pi) \vdash_{IKD} \alpha$.* $\square$

7 Conclusion

We propose the p-classical introspection based on a simple but effective idea: deriving $\neg\mathbf{B}\alpha$ if α cannot be derived from the perspective theory of each and every possible belief world. Autoepistemic logic of belief using the p-classical introspection is capable of characterizing almost all nonmonotonic reasoning semantics.

Because the p-classical introspection is based on classical entailment, our results present a new approach to the implementation of efficient nonmonotonic reasoning systems.

References

[1] J.J. Alferes and L.M. Pereira. Belief, provability and logic programs. *Journal of Applied Non-classical Logics*, 5:31–50, 1995.

[2] A. Bondarenko, P.M. Dung, R. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *An manuscript*, 1997.

[3] S. Brass, J. Dix, I. Niemela, and T. Przymusinski. A comparison of the static and the disjunctive well-founded semantics. *An manuscript*, 1996.

[4] S. Brass, J. Dix, and T. Przymusinski. Super logic programs. In *Proceedings of KR96*, 1996.

[5] R.G. Downey and M.R. Fellows. Fixed-parameter tractability and completeness i: Basic results. *SIAM J. Computing*, pages 873–921, 1995.

[6] P. Dung. On the acceptability of argument and its fundamental rule in non-monotonic reasoning and logic programming and n-person game. *Artificial Intelligence*, 76, 1995. A short version appeared in *Proc. IJCAI'93*.

[7] T. Eiter and G. Gottlob. Complexity aspects of various semantics. In *Proceedings of the 12th ACM PODS*, pages 158–167, 1993.

[8] T. Eiter, N. Leone, and D. Sacca. Expressive power and complexity of partial models for disjunctive deductive databases. *Annals of Mathematics and Artificial intelligence*, 1997.

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

[10] M. Gelfond. On stratified autoepistemic theories. In *Proceedings of AAAI-87*, pages 207–211. Morgan Kaufmann Publishers, 1987.

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

[12] J.Y. Halpern and Y.Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[13] Y. J. Jiang and Y. Aramkulchai. On the relationship between assumption-based framework and autoepistemic logic. In *Proc. of 8th Intern. Symposium on Meth. for intelligent Systems, LNCS 869*, pages 571–583, 1994.

[14] A. Kakas, P. Mancarella, and P.M. Dung. The acceptability semantics for logic programs. In *Proc. 11th International Conference on Logic Programming*, pages 504–519, 1994.

[15] V. Lifschitz and G. Schwarz. Extended logic programs as autoepistemic theories. In *Proc. of the 2nd Workshop on LPNR*, pages 101–114, 1993.

[16] W. Marek and M. Truszczunski. *Nonmonotonic Logic*. Springer-Verlag, 1991.

[17] J. Minker. On indefinite databases and the closed world assumption. In *Lecture Notes in Computer Science, 138*, pages 292–308. Springer-Verlag, 1982.

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

[19] I. Niemela. Implementing circumscriptions using a tableau method. In *Proc. of the European Conference on AI*, 1996.

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

[21] T. C. Przymusinski. A knowledge representation framework based on autoepistemic logic of minimal beliefs. In *Proc. of the AAAI'94*, 1994.

[22] T. C. Przymusinski. Static semantics for normal and disjunctive logic programs. *Annals of Mathematics and Artificial intelligence*, 14:323–357, 1995.

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

[24] G. Schwarz. Autoepistemic logic of knowledge. In *Proc. of the first Workshop on LPNR*, pages 260–274, 1991.

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

[26] J.-H. You and L. Y. Yuan. A three-valued semantics of deductive databases and logic programs. *Journal of Computer and System Science*, 49:334–361, 1994. The preliminary version appears in the Proc. of the 9th ACM PODS, page 171-182, 1990.

[27] L. Y. Yuan and J.-H. You. Discriminant circumscription. In *Proceedings of the 10th Conference on Foundation of Software and Theoretical Computer Science, LNCS Vol. 472*, pages 22–32, 1990.

[28] L. Y. Yuan and J.-H. You. Autoepistemic circumscription and logic programming. *Journal of Automated Reasoning*, 10:143–160, 1993.