A Scheme for Weakened Negative Introspection in Autoepistemic Reasoning

Tomi Janhunen and Ilkka Niemelä *

Helsinki University of Technology,
Department of Computer Science, Digital Systems Laboratory,
Otakaari 1, 02150 Espoo, Finland
Email: {Tomi.Janhunen,Ilkka.Niemela}@hut.fi

Abstract. A novel scheme for weakened negative introspection in autoepistemic reasoning is proposed. The scheme is formulated proof-theoretically in terms of enumerations of formulae. The resulting sets of conclusions, cautious expansions, are a generalization of Moore style stable expansions such that each set of premises has a cautious expansion. A finitary characterization is provided for the cautious expansions. Excluding weakly grounded expansions, i.e. considering only $\mathcal{N}$-grounded cautious expansions, is shown to imply attractive properties not present in, e.g., Moore style autoepistemic logic. For every set of premises, the intersection of the $\mathcal{N}$-grounded cautious expansions is a cautious expansion which is the unique least $\mathcal{N}$-grounded cautious expansion under a natural partial order of expansions. The least $\mathcal{N}$-grounded cautious expansion can be constructed as the least fixed point of a monotonic operator. Using the finitary characterization the decidability of decision problems related to cautious expansions is established and the decision problems are related to the polynomial time hierarchy. A comparison indicates that the new principle is stronger than other similar proposals and than that used in the well-founded semantics.

1 Introduction

Nonmonotonic reasoning is a key feature in areas such as databases, logic programming, inheritance hierarchies, reasoning about action, and diagnosis. Several nonmonotonic logics have been devised to provide formal accounts of nonmonotonic reasoning. Autoepistemic logic [11] and default logic [16] belong to the leading approaches. They are closely related [8, 9] and cover a substantial fragment of forms of nonmonotonic reasoning (see e.g. [2, 3, 4]).

This paper raises two issues in the two leading formalizations of nonmonotonic reasoning. Firstly, in both autoepistemic logic and default logic possible sets of correct conclusions for a set of premises are defined nonconstructively using a fixed point condition referring directly to the whole set of conclusions.

* Ilkka Niemelä gratefully acknowledges the support from the following foundations: Emil Aaltonen Foundation, Foundation of Technology, Heikki and Hilma Honkanen Foundation, Alfred Kordelin Foundation, Jenny and Antti Wihuri Foundation.

But there seems to be no iterative method which would construct the set of conclusions step by step. Secondly, there are some sets of premises for which no set of correct conclusions exists. In this paper we develop a variant of autoepistemic logic where reasoning can be understood as an iterative process and where each set of premises induces a set of correct conclusions.

Autoepistemic logic is a modal logic with an operator L which is interpreted as "is believed". Autoepistemic logic models the beliefs of an ideally rational introspective agent. A possible set of correct autoepistemic conclusions from a set of premises, a *stable expansion* of the premises, is defined as a set of beliefs of the agent given the premises as the initial assumptions of the agent. Stable expansions are defined using a fixed point equation capturing the following features of the agent. The agent is capable of using both positive introspection (if χ is a belief, so is $L\chi$) and negative introspection (if χ is not a belief, then $\neg L\chi$ is). The agent's ideal rationality means that the agent believes exactly every logical consequence of the initial assumptions and its beliefs.

There are sets of premises for which there are no stable expansions. A typical example is the premise $\neg Lp \rightarrow p$. Intuitively, the situation comes about because the belief p cannot be adopted by the agent as there is no justification for p, i.e. p is not a logical consequence of the premise and the beliefs of the agent, but then the agent is forced (by negative introspection) to adopt the belief $\neg Lp$. This leads to a non-stable set of beliefs where, on one hand, the agent disbelieves p and, on the other hand, it should believe p as a logical consequence of the premise and its beliefs. This suggests that weakening the negative introspection capability of the agent guarantees a stable set of beliefs for each set of premises.

Przymusinski [15], Kalinski [6, 7], and Bonatti [1] have proposed schemes for weakened negative introspection and established close correspondences with the well-founded semantics [17] of logic programs. However, the notion of negative introspection is rather weak in these approaches as well as in the well-founded semantics and other related proposals such as the stationary default expansions [14]. This aspect is further discussed in Sect. 3.3. In this paper we propose a novel notion of negative introspection. It turns out to be stronger that in the previous approaches and it guarantees a unique least expansion for each set of premises. It also leads to a monotonic operator such that the least expansion can be constructed iteratively as the least fixed point of the operator.

2 Cautious Autoepistemic Logic

To solve the problem of existence of autoepistemic expansions, we propose a generalization of Moore's approach: *cautious autoepistemic logic* CAEL. For the sake of simplicity, we restrict ourselves to the propositional case. Let us first introduce some notation and terminology. The symbol $\mathcal{L}_{ae}$ denotes the *autoepistemic language* which is derived from the *propositional language* $\mathcal{L}$ by introducing an additional rule in the respective grammar: $\phi \in \mathcal{L}_{ae}$ implies $L\phi \in \mathcal{L}_{ae}$. The formulae of the form $L\phi$ are called *quasi-atoms*. *Quasi-literals* are either quasi-atoms or their negations. The propositional consequence relation $\models$ is extended to its autoepistemic counterpart $\models_{ae}$ by treating all $L\phi$ formulae as atomic propositions. We let $Cn(\Gamma)$ denote the set $\{\phi \mid \Gamma \models_{ae} \phi\}$.

Let Σ be the set of premises given as the initial assumptions of the agent. A Moore style *stable expansion* $\Delta \subseteq \mathcal{L}_{ae}$ of Σ representing the set of beliefs of the agent contains exactly all the logical consequences of the premises Σ and the beliefs obtained by positive introspection ($L\Delta$) and negative introspection ($\neg L\overline{\Delta}$) where $\neg\Gamma = \{\neg\phi \mid \phi \in \Gamma\}$, $\overline{\Gamma} = \mathcal{L}_{ae} - \Gamma$ and $L\Gamma = \{L\phi \mid \phi \in \Gamma\}$. Stable expansions Δ of Σ fulfill the following fixed point equation [11]:

$$\Delta = Cn(\Sigma \cup L\Delta \cup \neg L\overline{\Delta}). \tag{1}$$

When formalizing CAEL our intention is to stay close to Moore style expansions. So as a starting point, we adopt the original positive introspection principle but develop a more cautious negative introspection principle. The new principle is based on the following idea: given a belief state of the agent, the agent is cautious and disbelieves a formula ϕ only when it is not possible to extend its current set of beliefs to include ϕ. The strength of the principle depends on what is considered as a possible extension of a set of beliefs. We aim to employ a rather tight notion of an extension and use *enumeration-based introspection*, recently introduced by Niemelä [12], to formalize the notion. In enumeration-based introspection the agent examines formulae in an order determined by an enumeration ε and applies an introspection rule to each formula in ε. The logical consequences of the initial assumptions and the results of previous introspection steps specify a set of beliefs of the agent at each step and the introspection rules employed are based directly on this set: if the next formula ϕ in the enumeration ε belongs to the set of beliefs, $L\phi$ is inferred (positive introspection) and otherwise $\neg L\phi$ is inferred (negative introspection). In this way a set of results of introspection $B^{\varepsilon}(\Sigma)$ is constructed. The corresponding set of beliefs, i.e. the set of logical consequences of Σ and $B^{\varepsilon}(\Sigma)$, is denoted by $EB^{\varepsilon}(\Sigma)$.

Definition 1. [12] Let Σ be a set of formulae of $\mathcal{L}_{ae}$. Let $\varepsilon = \psi_1, \psi_2, ...$ be an enumeration of a subset Γ of $\mathcal{L}_{ae}$ and let $|\varepsilon|$ denote its length. Define $B_0^{\varepsilon}(\Sigma) = \emptyset$ and $B_{i+1}^{\varepsilon}(\Sigma)$ for $i = 0, 1, ..., |\varepsilon|$ as follows:

$$B_{i+1}^{\varepsilon}(\Sigma) = \begin{cases} B_i^{\varepsilon}(\Sigma) \cup \{L\psi_{i+1}\}, & \text{if } \Sigma \cup B_i^{\varepsilon}(\Sigma) \models_{ae} \psi_{i+1} \\ B_i^{\varepsilon}(\Sigma) \cup \{\neg L\psi_{i+1}\}, & \text{if } \Sigma \cup B_i^{\varepsilon}(\Sigma) \not\models_{ae} \psi_{i+1} \end{cases}$$

Finally let

$$B^{\varepsilon}(\Sigma) = \begin{cases} B_{|\varepsilon|}^{\varepsilon}(\Sigma), & \text{if } |\varepsilon| \text{ is finite} \\ \bigcup_{i=0}^{\infty} B_i^{\varepsilon}(\Sigma), & \text{otherwise} \end{cases}$$

and the enumeration-based set of beliefs $EB^{\varepsilon}(\Sigma) = \{\phi \mid \Sigma \cup B^{\varepsilon}(\Sigma) \models_{ae} \phi\}$.

A state of beliefs of the agent is represented by a pair $\langle \Delta^+, \Delta^- \rangle$ where Δ^+ contains the formulae that are *believed* and Δ^- those *disbelieved* by the agent. A possible (final) state of beliefs of the agent given a set of initial assumptions is a state which is not changed by the introspection principles. Formalizing this condition leads to a fixed point equation similar to (1). Depending on the set of premises Σ, there may exist *undetermined* formulae that are neither believed or disbelieved by the agent. In this sense, cautious expansions have a three-valued character, although a two-valued logic is used for monotonic inference. The term *belief status* of ϕ refers to its membership in one of the sets Δ^+, Δ^- or $\overline{\Delta^+ \cup \Delta^-}$.

Definition 2. A pair of subsets $\langle \Delta^+, \Delta^- \rangle$ of $\mathcal{L}_{ae}$ is a cautious autoepistemic expansion of the set of premises Σ iff

$$\langle \Delta^+, \Delta^- \rangle = \langle IP^+_\Sigma(\Delta^+, \Delta^-), IP^-_\Sigma(\Delta^+, \Delta^-) \rangle \tag{2}$$

where $IP^+_\Sigma(\Delta^+, \Delta^-) = Cn(\Sigma \cup L\Delta^+ \cup \neg L\Delta^-)$ and

$$IP^-_\Sigma(\Delta^+, \Delta^-) = \{\phi \mid \text{there is no enumeration } \varepsilon \text{ of } \mathcal{L}_{ae} - (\Delta^+ \cup \Delta^-) \text{ such that } \phi \in EB^\varepsilon(\Sigma \cup L\Delta^+ \cup \neg L\Delta^-)\}.$$

So a possible extension of a set of beliefs Δ^+ in a belief state $\langle \Delta^+, \Delta^- \rangle$ is formalized as a set of beliefs $EB^\varepsilon(\Sigma \cup L\Delta^+ \cup \neg L\Delta^-)$ induced by an enumeration ε of $\mathcal{L}_{ae} - (\Delta^+ \cup \Delta^-)$ given $\Sigma \cup L\Delta^+ \cup \neg L\Delta^-$ as the premises. If $\phi \in EB^\varepsilon(\Sigma \cup L\Delta^+ \cup \neg L\Delta^-)$, we say that ε is a *refuting enumeration* for ϕ. Hence according to the new negative introspection principle the agent disbelieves only irrefutable formulae in each belief state.

Example 1. Consider the set of premises $\Sigma = \{\neg Lp \rightarrow p\}$ and a belief state where p is not believed. In this state the new negative introspection principle does not yield $\neg Lp$ as the set of beliefs can be extended to include p. For instance, an enumeration starting with p refutes p.

Proposition 3. *All stable expansions are cautious expansions.*

Proof. Moore style stable expansions are complete in the following sense: each formula is either believed or disbelieved by the agent. As there are no undetermined formulae, $\Delta^+ = \overline{\Delta^-}$ holds implying that $\mathcal{L}_{ae} - (\Delta^+ \cup \Delta^-) = \emptyset$. Under these two equalities the requirements of Definition 2 reduce to (1). □

2.1 Finitary Characterization

A cautious expansion consists of two infinite sets of formulae which leads to representational problems. Furthermore, cautious expansions have a nonconstructive nature due to self-reference. Thus cautious expansions may seem tedious to handle and reason about. A solution is to provide a *finitary characterization* for cautious expansions. Niemelä [13] has developed *full sets* as a finitary characterization for Moore style stable expansions. In the following, we develop a similar framework for cautious expansions.

Two *recursive consequence relations* $\models_L$ and $=\!\mid_L$ are introduced in Definition 4. They provide the finitary mechanism for deciding whether a formula ϕ is believed or disbelieved by the agent, respectively. The *subbeliefs* $SB_{\langle\Sigma,\Lambda\rangle}(\phi)$ of ϕ are determined before deciding the belief status of ϕ by both relations. Niemelä used $\not\models_L$ for negative introspection in conjunction with stable expansions. However, we introduce an *irrefutability relation* $=\!\mid_L$ to capture the weakened negative introspection capability of the agent. In general, the two relations $=\!\mid_L$ and $\models_L$ are nonmonotonic. We let $Sf^L(\phi)$ denote the set of quasi-atoms that appear in a formula ϕ and define a function q-atom such that q-atom(χ) = q-atom$(\neg\chi) = \chi$ for any quasi-atom χ. These two operations are naturally extended to take sets of formulae as their arguments. The *full sets* of Definition 5 form the finitary characterization for cautious expansions. The two conditions are self-referential implying that Σ-full sets have still a nonconstructive nature.

Definition 4. For a set of premises Σ, a set of quasi-literals Λ, and $\phi \in \mathcal{L}_{ae}$:

1. $\langle\Sigma, \Lambda\rangle \models_L \phi$ iff $\Sigma \cup \Lambda \cup SB_{\langle\Sigma,\Lambda\rangle}(\phi) \models_{ae} \phi$
2. $\langle\Sigma, \Lambda\rangle =\!\mid_L \phi$ iff there is no enumeration ε of $\{\chi \mid L\chi \in Sf^L(\Sigma \cup \{\phi\}) - \text{q-atom}(\Lambda \cup SB_{\langle\Sigma,\Lambda\rangle}(\phi))\}$ such that $\phi \in EB^{\varepsilon}(\Sigma \cup \Lambda \cup SB_{\langle\Sigma,\Lambda\rangle}(\phi))$

where
$SB_{\langle\Sigma,\Lambda\rangle}(\phi) = \{L\chi \in Sf^L(\phi) \mid \langle\Sigma, \Lambda\rangle \models_L \chi\} \cup \{\neg L\chi \in \neg Sf^L(\phi) \mid \langle\Sigma, \Lambda\rangle =\!\mid_L \chi\}$.

Definition 5. Given a set of premises Σ, a set $\Lambda \subseteq Sf^L(\Sigma) \cup \neg Sf^L(\Sigma)$ is Σ-full iff the following conditions hold for every $L\chi \in Sf^L(\Sigma)$:

1. $L\chi \in \Lambda$ iff $\langle\Sigma, \Lambda\rangle \models_L \chi$ 2. $\neg L\chi \in \Lambda$ iff $\langle\Sigma, \Lambda\rangle =\!\mid_L \chi$

Theorem 6 gives a fundamental result: cautious expansions of Σ are in one-to-one correspondence with Σ-full sets. The function FS_Σ maps cautious expansions of Σ to the corresponding Σ-full sets. Given a Σ-full set Λ, the recursive consequence relations $\models_L$ and $=\!\mid_L$ provide means of determining the belief status of any formula of $\mathcal{L}_{ae}$ with respect to the corresponding cautious expansion. This is how the function CE_Σ reconstructs the cautious expansion.

Theorem 6. *For the set of premises Σ, let $\mathcal{CE}$ be the set of cautious expansions of Σ and $\mathcal{FS}$ the set of Σ-full sets. Then $FS_\Sigma(\langle\Delta^+, \Delta^-\rangle) = \{\phi \in L\Delta^+ \cup \neg L\Delta^- \mid \text{q-atom}(\phi) \in Sf^L(\Sigma)\}$ is a function $FS_\Sigma : \mathcal{CE} \rightarrow \mathcal{FS}$ and $CE_\Sigma(\Lambda) = \langle\{\phi \mid \langle\Sigma, \Lambda\rangle \models_L \phi\}, \{\phi \mid \langle\Sigma, \Lambda\rangle =\!\mid_L \phi\}\rangle$ is a function $CE_\Sigma : \mathcal{FS} \rightarrow \mathcal{CE}$. Moreover, FS_Σ and CE_Σ are bijections and inverses of each other.*

Proof sketch. (i) The enumerations employed by the negative introspection can be significantly reduced. We let $\Pi_\Gamma(\varepsilon)$ denote the enumeration obtained from an enumeration ε by preserving the elements that appear in Γ. Quite similarly, $\Pi_\Gamma(\Lambda)$ denotes the set $\{\phi \in \Lambda \mid \text{q-atom}(\phi) \in \Gamma\}$ where Λ is a set of quasi-literals and Γ is a set of quasi-atoms. Let Σ be a set of premises such that $Sf^L(\Sigma)$ is finite. Let Λ be a consistent set of quasi-literals and ε an enumeration of the formulae in $\mathcal{L}_{ae} - \{\psi \mid L\psi \in \text{q-atom}(\Lambda)\}$. It can be proved by induction on the introspection sequence that $\phi \in EB^{\varepsilon}(\Sigma \cup \Lambda)$ iff $\phi \in EB^{\varepsilon'}(\Sigma \cup \Pi_{Sf^L(\Sigma\cup\{\phi\})}(\Lambda))$ where $\varepsilon' = \Pi_{\{\psi \mid L\psi \in Sf^L(\Sigma\cup\{\phi\})\}}(\varepsilon)$. (ii) Let $\langle\Delta^+, \Delta^-\rangle$ be a cautious expansion of Σ and $\Lambda = \Pi_{Sf^L(\Sigma)}(L\Delta^+ \cup \neg L\Delta^-)$. By the L-depth of a formula ϕ we mean the maximum nesting level of belief operators appearing in ϕ. Using the reduction of enumerations, one can prove by induction on the L-depth of formulae that for every $L\chi \in Sf^L(\Sigma)$: $\langle \Sigma, \Lambda \rangle \models_L \chi$ iff $\chi \in IP^+_\Sigma(\Delta^+, \Delta^-)$ and $\langle \Sigma, \Lambda \rangle \dashv_L \chi$ iff $\chi \in IP^-_\Sigma(\Delta^+, \Delta^-)$. These together imply that Λ is Σ-full and FS_Σ is a function as claimed above. (iii) Let Σ be the set of premises and Λ a Σ-full set. Let $\Delta^+ = \{\phi \mid \langle \Sigma, \Lambda \rangle \models_L \phi\}$ and $\Delta^- = \{\phi \mid \langle \Sigma, \Lambda \rangle \dashv_L \phi\}$. It follows that $\Lambda = \Pi_{Sf^L(\Sigma)}(L\Delta^+ \cup \neg L\Delta^-)$ and for every $\phi \in \mathcal{L}_{ae}$, $SB_{\langle \Sigma, \Lambda \rangle}(\phi) = \Pi_{Sf^L(\phi)}(L\Delta^+ \cup \neg L\Delta^-)$. Using these properties and the reduction of enumerations it can be proven that $\langle \Delta^+, \Delta^- \rangle$ is a cautious expansion of Σ by induction on the L-depth of formulae. Thus CE_Σ is a well-defined function. (iv) By (iii) for a Σ-full set Λ, $\Lambda = \Pi_{Sf^L(\Sigma)}(L\Delta^+ \cup \neg L\Delta^-)$ where $\langle \Delta^+, \Delta^- \rangle = CE_\Sigma(\Lambda)$. This implies that $\Lambda = FS_\Sigma(CE_\Sigma(\Lambda))$. It follows that the functions FS_Σ and CE_Σ are bijections and inverses of each other. Consequently, for a cautious expansion $\mathcal{E}$, $\mathcal{E} = CE_\Sigma(FS_\Sigma(\mathcal{E}))$ holds. □

Next we present some exemplary sets of premises and give their cautious expansions. By Theorem 6 we need to consider the consistent subsets of $Sf^L(\Sigma) \cup \neg Sf^L(\Sigma)$ and check whether Σ-fullness conditions are satisfied. Notice that only formulae χ such that $L\chi \in Sf^L(\Sigma)$ have to be considered.

Example 2. The set of premises $\Sigma = \{\neg Lp \rightarrow p\}$ has no stable expansions. Yet Σ has a cautious expansion $CE_\Sigma(\emptyset)$ in which the belief status of p is undetermined.

Example 3. The set of premises $\Sigma = \{Lp \rightarrow p\}$ has two cautious expansions, namely $CE_\Sigma(\{\neg Lp\})$ and $CE_\Sigma(\{Lp\})$.

Example 4. The set of premises $\Sigma = \{\neg Lp \rightarrow q, \neg Lq \rightarrow p, \neg Lp \wedge \neg Lq \rightarrow r\}$ has three cautious expansions: $CE_\Sigma(\{\neg Lp, Lq\})$, $CE_\Sigma(\{Lp, \neg Lq\})$ and $CE_\Sigma(\emptyset)$. The first two represent complementary states of beliefs. In the third, the belief status of both p and q is undetermined because of the two competing choices.

3 Grounded Cautious Expansions

As the previous examples show, there are sets of premises with multiple cautious expansions. A standard approach is to treat the intersection of the expansions as the set of correct conclusions. The intersection of Moore style stable expansions cannot be a stable expansion but Example 4 provides a set of premises for which the intersection of the cautious expansions is a cautious expansion. However, the intersection of the cautious expansions in Example 3 shows that not every set of premises enjoys this property. It turns out that the intersection fails to be a cautious expansion only when there is an expansion containing self-supporting beliefs. Such expansions are possible because we have adopted the original positive introspection principle which leaves room for weakly grounded beliefs. In Example 3, the expansion $CE_\Sigma(\{Lp\})$ of $\Sigma = \{Lp \rightarrow p\}$ contains self-supporting beliefs such as p. We show that if a more strongly grounded form of positive introspection is employed, then for the resulting subclass of cautious expansions, the intersection of the expansions is a cautious expansion and, moreover, it is the least expansion with respect to a natural partial order on expansions.

Let us first make precise the partial order on cautious expansions and what is meant by the least cautious expansion. A cautious expansion $\mathcal{E}_1 = \langle \Delta_1^+, \Delta_1^- \rangle$ is smaller than or equal to $\mathcal{E}_2 = \langle \Delta_2^+, \Delta_2^- \rangle$, denoted by $\mathcal{E}_1 \sqsubseteq \mathcal{E}_2$, iff $\Delta_1^+ \subseteq \Delta_2^+$ and $\Delta_1^- \subseteq \Delta_2^-$. Let Σ be a set of premises and $\mathcal{C}E$ the set of cautious expansions of Σ. A cautious expansion $\mathcal{E} \in \mathcal{C}E$ is the least cautious expansion of $\mathcal{C}E$ iff $\mathcal{E} \sqsubseteq \mathcal{E}'$ for every $\mathcal{E}'$ in $\mathcal{C}E$. If the least cautious expansion exists in $\mathcal{C}E$, it is contained in every other cautious expansion. By the properties of subset inclusion, it is the intersection of all cautious expansions in $\mathcal{C}E$.

The directly self-supporting beliefs can be eliminated by imposing a special $\mathcal{N}$-groundedness condition on cautious expansions. The symbol $\mathcal{N}$ refers to a nonstandard modal logic, i.e. the propositional calculus equipped with the necessitation rule $\phi \vdash L\phi$. We let $Cn_{\mathcal{N}}(\Sigma)$ denote the set $\{\phi \mid \Sigma \vdash_{\mathcal{N}} \phi\}$ where $\vdash_{\mathcal{N}}$ denotes derivability in $\mathcal{N}$. Marek and Truszczyński [10] have used a similar condition for stable expansions. The $\mathcal{N}$-groundedness property can be captured in the finitary characterization by exploiting a condition similar to that used by Gottlob [5] in a finitary characterization for Moore style $\mathcal{N}$-expansions.

Definition 7. Let $\langle \Delta^+, \Delta^- \rangle$ be a cautious expansion of a set of premises Σ. It is $\mathcal{N}$-grounded iff it satisfies the equation $\Delta^+ = Cn_{\mathcal{N}}(\Sigma \cup \neg L\Delta^-)$.

Definition 8. Let Λ be a Σ-full set. It is $\mathcal{N}$-grounded iff it satisfies for all $L\chi \in Sf^L(\Sigma)$: $L\chi \in \Lambda$ iff $\Sigma \cup [\Lambda]^- \vdash_{\mathcal{N}} \chi$ where $[\Lambda]^- = \{\neg L\chi \mid \neg L\chi \in \Lambda\}$.

By Theorem 9 the $\mathcal{N}$-grounded Σ-full sets and $\mathcal{N}$-grounded cautious expansions of Σ are in one-to-one correspondence. The subset inclusion of $\mathcal{N}$-grounded full sets coincides with the partial order on $\mathcal{N}$-grounded cautious expansions as stated in Proposition 10. Consequently, the least $\mathcal{N}$-grounded cautious expansion corresponds to the least $\mathcal{N}$-grounded full set.

Theorem 9. *Let Σ be a set of premises, $\mathcal{C}E$ the set of $\mathcal{N}$-grounded cautious expansions of Σ and $\mathcal{F}S$ the set of $\mathcal{N}$-grounded Σ-full sets. Let FS_Σ and FS_Σ be defined similarly as in Theorem 6. Then $FS_\Sigma : \mathcal{C}E \to \mathcal{F}S$ and $CE_\Sigma : \mathcal{F}S \to \mathcal{C}E$ are functions. Furthermore, they are bijections and inverses of each other.*

Proof sketch. The derivability relation $\vdash_{\mathcal{N}}$ satisfies the following reduction property. Let Λ be a set of quasi-literals satisfying the consistency requirement: $\neg L\chi \in \Lambda$ implies $\Sigma \cup \Lambda \nvdash_{\mathcal{N}} \chi$. Then $\Sigma \cup \Lambda \vdash_{\mathcal{N}} \phi$ implies $\Sigma \cup \Lambda' \vdash_{\mathcal{N}} \phi$ where $\Lambda' = \Pi_{Sf^L(\Sigma \cup \{\phi\})}(\Lambda)$. This implies that, given an $\mathcal{N}$-grounded cautious expansion $\mathcal{E}$ of Σ, $FS_\Sigma(\mathcal{E})$ is also $\mathcal{N}$-grounded. On the other hand, given an $\mathcal{N}$-grounded Σ-full set Λ, it can be proven that $CE_\Sigma(\Lambda)$ is $\mathcal{N}$-grounded by induction on the length of minimal proofs in the logic $\mathcal{N}$. The bijective relationship follows then by Theorem 6. □

Proposition 10. *Let Σ be a set of premises. Let Λ_1 and Λ_2 be two $\mathcal{N}$-grounded Σ-full sets and $\mathcal{E}_1$ and $\mathcal{E}_2$ the corresponding $\mathcal{N}$-grounded cautious expansions $CE_\Sigma(\Lambda_1)$ and $CE_\Sigma(\Lambda_2)$, respectively. Then $\Lambda_1 \subseteq \Lambda_2$ if and only if $\mathcal{E}_1 \sqsubseteq \mathcal{E}_2$.*

Proof sketch. Let $\Lambda_1 \subseteq \Lambda_2$ hold. It can be shown by induction of the L-depth of formulae that for any $\phi \in \mathcal{L}_{ae}$, $\langle\Sigma, \Lambda_1\rangle \models_L \phi$ implies $\langle\Sigma, \Lambda_2\rangle \models_L \phi$ and $\langle\Sigma, \Lambda_1\rangle \dashv_L \phi$ implies $\langle\Sigma, \Lambda_2\rangle \dashv_L \phi$. The claim of the proposition follows from these implications. The latter implication can be proved using some auxiliary results on enumeration-based introspection. We say that a set of quasi-literals Γ is *generated by an enumeration* from the set of premises Σ if $\Gamma = B^{\varepsilon}(\Sigma)$ where ε is an enumeration of $\{\psi \mid L\psi \in \text{q-atom}(\Gamma)\}$. A Σ-full set Λ is $\mathcal{N}$-grounded iff it can be generated by an enumeration of $\{\psi \mid L\psi \in \text{q-atom}(\Lambda)\}$. As Λ_2 is an $\mathcal{N}$-grounded Σ-full set, for any $\phi \in \mathcal{L}_{ae}$ the set of quasi-literals $\Lambda_2 \cup SB_{\langle\Sigma,\Lambda_2\rangle}(\phi)$ has a generating enumeration. By the inductive hypothesis $\Lambda_1 \cup SB_{\langle\Sigma,\Lambda_1\rangle}(\phi) \subseteq \Lambda_2 \cup SB_{\langle\Sigma,\Lambda_2\rangle}(\phi)$ yielding that the difference $\Lambda_2 \cup SB_{\langle\Sigma,\Lambda_2\rangle}(\phi) - \Lambda_1 \cup SB_{\langle\Sigma,\Lambda_1\rangle}(\phi)$ can be generated by an enumeration, say ε_d. As a consequence, a refuting enumeration of ϕ from $\Sigma \cup \Lambda_1 \cup SB_{\langle\Sigma,\Lambda_1\rangle}(\phi)$ can be constructed from a refuting enumeration ε_r of ϕ from $\Sigma \cup \Lambda_2 \cup SB_{\langle\Sigma,\Lambda_2\rangle}(\phi)$ by simply concatenating the two enumerations ε_d and ε_r. Hence, $\langle\Sigma, \Lambda_2\rangle \not\dashv_L \phi$ implies $\langle\Sigma, \Lambda_1\rangle \not\dashv_L \phi$. □

3.1 Least Grounded Expansion

In this section we provide an iterative method for constructing the least $\mathcal{N}$-grounded Σ-full set. The consequence relations $\models_{ae}$ and $\dashv_{ae}$ introduced in Definition 11 are the non-recursive counterparts of $\models_L$ and $\dashv_L$. The iterative construction is based on these relations. A parametrized operator FCAE_Σ is given in Definition 12. The argument Γ represents a Σ-full set under construction. At each step of the iteration, introspection is carried out on the basis of the previous step.

Definition 11. For a set of premises Σ, a set of quasi-literals Γ, and $\phi \in \mathcal{L}_{ae}$:

1. $\langle\Sigma, \Gamma\rangle \models_{ae} \phi$ iff $\Sigma \cup \Gamma \models_{ae} \phi$
2. $\langle\Sigma, \Gamma\rangle \dashv_{ae} \phi$ iff there is no enumeration ε of $\{\chi \mid L\chi \in Sf^L(\Sigma) - \text{q-atom}(\Gamma)\}$ such that $\phi \in EB^{\varepsilon}(\Sigma \cup \Gamma)$.

Definition 12. For a set of premises Σ, define an operator FCAE_Σ taking a consistent set of quasi-literals $\Gamma \subseteq Sf^L(\Sigma) \cup \neg Sf^L(\Sigma)$ as its argument:

$$FCAE_\Sigma(\Gamma) = \{L\psi \in Sf^L(\Sigma) \mid \langle\Sigma, \Gamma\rangle \models_{ae} \psi\} \cup \{\neg L\psi \in \neg Sf^L(\Sigma) \mid \langle\Sigma, \Gamma\rangle \dashv_{ae} \psi\}.$$

Then define the iterative sequence $\Gamma_0, \Gamma_1, \ldots$ by letting $\Gamma_0 = \emptyset$ and $\Gamma_{i+1} = FCAE_\Sigma(\Gamma_i)$ when $i \geq 0$.

We have collected three important properties of the iterative construction in Proposition 13. By Theorem 14 the fixed point of FCAE_Σ is the least $\mathcal{N}$-grounded Σ-full set. Consequently, every set of premises has a cautious expansion. This result is in contrast with the original autoepistemic logic.

Proposition 13. *Let Σ be a finite set of sentences of $\mathcal{L}_{ae}$. The sequence given in Definition 12 is non-decreasing, i.e. $\Gamma_i \subseteq \Gamma_{i+1}$ for all $i \geq 0$. Furthermore, the sequence reaches its fixed point in a finite number of steps. Let fp(FCAE$_\Sigma$) denote the fixed point of FCAE$_\Sigma$ when the iteration is started from $\Gamma_0 = \emptyset$. The set of quasi-literals fp(FCAE$_\Sigma$) is Σ-full.*

Proof sketch. The iterative sequence can be proven to be non-decreasing by induction on i, i.e. for all $L\chi \in Sf^L(\Sigma)$, $\langle \Sigma, \Gamma_i \rangle \models_{ae} \chi$ implies $\langle \Sigma, \Gamma_{i+1} \rangle \models_{ae} \chi$ and $\langle \Sigma, \Gamma_i \rangle =\!|_{ae} \chi$ implies $\langle \Sigma, \Gamma_{i+1} \rangle =\!|_{ae} \chi$. The former implication follows by the monotonicity of $\models_{ae}$. To prove the latter, notice that a generating enumeration ε_{i+1} of Γ_{i+1} from Σ can be constructed from the generating enumeration ε_i of Γ_i by concatenating the elements of $\{\psi \mid L\psi \in \text{q-atom}(\Gamma_{i+1} - \Gamma_i)\}$ to ε_i in an arbitrary order. The difference of the two enumerations ε_i and ε_{i+1} can be used for constructing refuting enumerations so that $\langle \Sigma, \Gamma_{i+1} \rangle \not=\!|_{ae} \chi$ implies $\langle \Sigma, \Gamma_i \rangle \not=\!|_{ae} \chi$. As the iterative sequence is non-decreasing, at each step of the iteration *either* new quasi-literals are added to the Σ-full set under construction *or* the fixed point fp(FCAE$_\Sigma$) has been reached. Since $Sf^L(\Sigma)$ is finite, additions can be made only finitely many times. The fixed point Γ = fp(FCAE$_\Sigma$) satisfies for all $L\chi \in Sf^L(\Sigma)$: (i) $L\chi \in \Gamma$ iff $\langle \Sigma, \Gamma \rangle \models_{ae} \chi$ and (ii) $\neg L\chi \in \Gamma$ iff $\langle \Sigma, \Gamma \rangle =\!|_{ae} \chi$. This implies that fp(FCAE$_\Sigma$) is Σ-full. □

Theorem 14. *Let Σ be a finite set of premises. Then Λ = fp(FCAE$_\Sigma$) is the unique least $\mathcal{N}$-grounded Σ-full set and $CE_\Sigma(\Lambda)$ is the unique least $\mathcal{N}$-grounded cautious expansion.*

Proof sketch. The set of quasi-literals fp(FCAE$_\Sigma$) has a generating enumeration by the proof of Proposition 13. Consequently, fp(FCAE$_\Sigma$) is an $\mathcal{N}$-grounded Σ-full set. To show the leastness of fp(FCAE$_\Sigma$), one has to prove that fp(FCAE$_\Sigma$) is contained in any other $\mathcal{N}$-grounded Σ-full set Λ. This can be done by induction on the sequence of Definition 12. However, some auxiliary results are needed. The sets of quasi-literals Γ_i have generating enumerations ε_i from Σ. Moreover, these enumerations can be extended to an enumeration generating the Σ-full set Λ. As a consequence $\Gamma_i \subseteq \Lambda$ for all $i \geq 0$ and fp(FCAE$_\Sigma$) $\subseteq \Lambda$. By Theorem 9 and Proposition 10, CE_Σ(fp(FCAE$_\Sigma$)) is the least $\mathcal{N}$-grounded cautious expansion of Σ. The uniqueness follows by the antisymmetry of $\subseteq$ and $\sqsubseteq$. □

3.2 Decidability and Complexity Results

In Definition 15 we list some problems of interest concerning the decidability and computational complexity of CAEL. All these problems are decidable. Theorem 17 gives upper bounds on the complexity by relating the various problems to the polynomial time hierarchy (PH). The corresponding completeness results remain open. Similar reasoning problems in Moore style autoepistemic logic are complete problems on the second level of PH [5, 13]. Note that the main decision problem in CAEL, **GCAE**, is polynomial time (deterministic) Turing reducible to a problem on the second level of PH.

Definition 15. Let Σ be a finite set of formulae and ϕ a formula. Then define the problem **BAE** of determining whether ϕ belongs to one of the cautious expansions of Σ and the problem **CAE** of determining whether a formula ϕ belongs to all cautious expansions of Σ. Let **GBAE** and **GCAE** denote the respective problems for the $\mathcal{N}$-grounded cautious expansions of Σ.

Theorem 16. *The problems **BAE**, **CAE**, **GBAE**, and **GCAE** are decidable.*

Proof sketch. The consequence relation $\models_{ae}$ is decidable as $\models$ is. The set $Sf^L(\Sigma)$ is finite implying that the decisions involving the relations $\dashv_L$ and $\dashv_{ae}$ can be made by considering only finitely many enumerations of formulae. □

Theorem 17. *The problems **BAE** and **GBAE** are in the class Σ_3^p. The problems **CAE** and **GCAE** are in the classes Π_3^p and Δ_3^p, respectively.*

Proof sketch. The consequence relation $\models_{ae}$ needed in the enumeration-based introspection is in Σ_1^p. A nondeterministic Turing machine can be used for finding refuting enumerations. This implies that the relations $\not\dashv_L$ and $\not\dashv_{ae}$ belong to Σ_2^p. The theorem follows directly from these results. Note that the operator FCAE_Σ provides a polynomial time reduction of **GCAE** to a problem on the second level of PH. □

3.3 Comparison with Other Approaches

In this section we compare cautious autoepistemic logic with the other approaches to weakened negative introspection. It seems unavoidable that when using weakened negative introspection some conclusions warranted by the original autoepistemic logic are lost. For instance, consider the set of premises in Example 4 which has two stable expansions corresponding to Σ-full sets $\{\neg Lp, Lq\}$ and $\{Lp, \neg Lq\}$. The intersection of the stable expansions contains, for instance, $p \vee q$ which does not belong to the least $\mathcal{N}$-grounded cautious expansion $CE_\Sigma(\emptyset)$. Therefore, conclusions based on the least $\mathcal{N}$-grounded cautious expansion are weaker than the ones based on the intersection of the stable expansions.

Przymusinski [15] proposes a form of weakened negative introspection based on minimal models for a limited subset of autoepistemic logic corresponding to logic programs. Kalinski's weak autoepistemic reasoning [6, 7] exploits the modal logic KD45 without restrictions to specific fragments of autoepistemic logic. Bonatti [1] develops a 3-valued autoepistemic logic. Each of the variants is shown to capture the *well-founded semantics* [17], a leading formal semantics for logic programs: if a general logic program is translated into a set of autoepistemic formulae using the Gelfond-Lifschitz translation [4] where negation-as-failure is interpreted as disbelief, then the least expansion of the translated program in each of the three approaches corresponds exactly to the unique well-founded model of the program. Consider a logic program $\{p \leftarrow \mathit{not}\, q; q \leftarrow \mathit{not}\, p; r \leftarrow \mathit{not}\, p, \mathit{not}\, q\}$. The well-founded (partial) model of the program is empty, i.e. no truth value is assigned to p, q, and r. The autoepistemic translation of the program is $\Sigma = \{\neg Lp \rightarrow q, \neg Lq \rightarrow p, \neg Lp \wedge \neg Lq \rightarrow r\}$ and correspondingly p, q, and r

are neither believed nor disbelieved in the least expansion of Σ in the three approaches. But in CAEL a stronger conclusion is obtained as r is disbelieved, i.e. $\neg Lr$ belongs to the least $\mathcal{N}$-grounded cautious expansion $CE_{\Sigma}(\emptyset)$. This seems quite intuitive, as both p and q are never disbelieved *simultaneously* by the agent.

Kalinski's and Bonatti's approaches are also applicable for premises not corresponding to general logic programs under the Gelfond-Lifschitz translation. But also here the approaches differ from CAEL. Given the premises $\Sigma = \{Lp \rightarrow p\}$ in Example 3 the belief status of p is left undetermined in Kalinski's approach. A similar observation can be made about Bonatti's approach. However, the least $\mathcal{N}$-grounded cautious expansion of Σ contains the formula $\neg Lp$.

4 Conclusions

We have introduced a novel negative introspection scheme for autoepistemic reasoning. The underlying idea is that a cautious autoepistemic agent disbelieves a formula in a belief state only when it is not possible to extend its set of beliefs to include the formula. A tight notion of a possible extension of a set of beliefs is formalized using enumeration-based introspection [12]. Given a set of initial premises, possible (final) belief states of the agent are called cautious expansions of the premises. When the positive introspection principle of the original autoepistemic logic is adopted, the resulting cautious expansions are shown to be a generalization of Moore style stable expansions. This positive introspection rule leaves room for self-supporting beliefs and we have studied the notion of $\mathcal{N}$-groundedness for eliminating them. The resulting expansions are shown to satisfy attractive properties which do not hold for Moore style expansions. The intersection of $\mathcal{N}$-grounded expansions is an $\mathcal{N}$-grounded expansion. It is the least $\mathcal{N}$-grounded expansion under a natural partial order on expansions and it can be constructed iteratively as the least fixed point of a monotonic operator. So in contrast to the original autoepistemic logic, every set of premises induces a set of beliefs of agent and the novel form of autoepistemic reasoning can be understood more constructively as a step by step process. We have provided a finitary characterization for cautious expansions by adapting Niemelä style full sets [13] and shown that the characterization implies the decidability of propositional cautious autoepistemic logic. A comparison indicates that the new introspection scheme delivers stronger conclusions than other autoepistemic logics employing weakened negative introspection and the closely related well-founded semantics.

There are many promising topics for further research. It is interesting to study the applicability of the new introspection principle to semantics of logic programs, inheritance reasoning, truth maintenance systems, and abduction. A semantical treatment of cautious expansions is of importance. A first-order generalization of the principle should be developed. The exact characterization of the complexity of cautious autoepistemic reasoning, e.g. in terms of a completeness result with respect to the polynomial time hierarchy, remains open.

References

1. Bonatti, P.: Autoepistemic logics as a unifying framework for the semantics of logic programs. In *Proceedings of the 1992 Joint International Conference and Symposium on Logic Programming*, Washington, D.C., USA, November 1992.
2. Elkan, C.: A rational reconstruction of nonmonotonic truth maintenance systems. *Artificial Intelligence*, 43:219–234, 1990.
3. Etherington, D.: Formalizing nonmonotonic reasoning systems. *Artificial Intelligence*, 31:41–85, 1987.
4. Gelfond, M. and Lifschitz, V.: The stable model semantics for logic programming. In *Proceedings of the 5th International Conference on Logic Programming*, pages 1070–1080, Seattle, USA, August 1988. The MIT Press.
5. Gottlob, G.: Complexity results for nonmonotonic logics. *Journal of Logic and Computation*, 2(3):397–425, June 1992.
6. Kalinski, J.: Autoepistemic expansions with incomplete belief introspection. In *Proceedings of the 1st International Workshop on Nonmonotonic and Inductive Logic*, pages 223–232, Karlsruhe, Germany, December 1990. Springer-Verlag. Lecture Notes in Artificial Intelligence 543.
7. Kalinski, J.: Weak autoepistemic reasoning. Presented at the European Workshop on Theoretical Foundations of Knowledge Representation and Reasoning held in conjunction with ECAI'92, August 3–4, Vienna, Austria, 1992.
8. Konolige, K.: On the relation between default and autoepistemic logic. *Artificial Intelligence*, 35:343–382, 1988.
9. Marek, W. and Truszczyński, M.: Relating autoepistemic and default logics. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, pages 276–288, Toronto, Canada, May 1989.
10. Marek, W. and Truszczyński, M.: Modal logic for default reasoning. *Annals of Mathematics and Artificial Intelligence*, 1:275–302, 1990.
11. Moore, R.: Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25:75–94, 1985.
12. Niemelä, I.: Constructive tightly grounded autoepistemic reasoning. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 399–404, Sydney, Australia, August 1991. Morgan Kaufmann Publishers.
13. Niemelä, I.: On the decidability and complexity of autoepistemic reasoning. *Fundamenta Informaticae*, 17(1,2):117–155, 1992.
14. Przymusinska, H. and Przymusinski, T.: Stationary default extensions. In *Working Notes of the 4th International Workshop on on Nonmonotonic Reasoning*, pages 179–193, Plymouth, Vermont, USA, May 1992.
15. Przymusinski, T.: Autoepistemic logics of closed beliefs and logic programming. In *Proceedings of the 1st International Workshop on Logic Programming and Nonmonotonic Reasoning*, pages 3–20, Washington, D.C., USA, July 1991. The MIT Press.
16. Reiter, R.: A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.
17. Van Gelder, A., Ross, K., and Schlipf, J.: The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):620–650, July 1991.