# Representing Autoepistemic Introspection in Terms of Default Rules *

Tomi Janhunen
Helsinki University of Technology,
Digital Systems Laboratory,
P.O.Box 1100, FIN-02015 HUT, Finland

April 10, 1997

**Abstract**

This paper addresses the relationship of Reiter's default logic, Moore's autoepistemic logic as well as Marek and Truszczyński's strong autoepistemic logic from a new point of view. Earlier research results on their interconnection support the view that default reasoning is a special case of autoepistemic reasoning. To the contrary, this paper shows how autoepistemic theories can be faithfully translated into default theories. Consequently, autoepistemic reasoning can be seen a form of default reasoning. This indicates – together with the previous research results – that autoepistemic logic and default logic are of equal generality. As a practical application of the presented translation, it is demonstrated how default decision procedures can be systematically used as autoepistemic decision procedures.

## 1 Introduction

This paper takes the interconnection of two leading formalizations of *non-monotonic reasoning* – Reiter's *default logic* [14] and Moore's *autoepistemic logic* [10] – into reconsideration. In light of earlier research results, default reasoning is often seen as a special case of autoepistemic reasoning. This view is supported, e.g., by the results of Gottlob [3], Konolige [4, 5, 6], Marek and Truszczyński [8, 9], Niemelä [12] and Truszczyński [15]. To the contrary, we demonstrate how autoepistemic reasoning can be seen as a form of default reasoning. This seemingly straightforward result has important consequences as it (i) indicates that autoepistemic logic and default logic are of equal generality, (ii) reveals

*This article has first appeared in the Proceedings of the 12th European Conference on Artificial Intelligence, Budabest Hungary, pages 70-74, August 1996. John Wiley & Sons, Ltd.

that much of non-monotonic reasoning is based on very simple primitives and (iii) allows the use of default decision procedures to solve autoepistemic decision problems in a systematic way.

The relationship of the two logics has been approached in terms of various *translation functions* that map a default theory into an autoepistemic theory and establish a correspondence between the *extensions* of a default theory and the *stable expansions* of its translation. In this way, it turned out that autoepistemic logic is capable of capturing default reasoning, but in a rather non-trivial way. A reasonable correspondence follows by composing translations provided by Truszczyński [15] and Gottlob [3]. This involves an intermediary representation of a default theory as a theory in *strong autoepistemic logic* proposed by Marek and Truszczyński [8]. Strong autoepistemic logic is a variant that considers only the *iterative expansions* of an autoepistemic theory that form a subclass of its stable expansions.

This paper proceeds in the opposite direction – starting from a characterization of stable expansions due to Marek, Schwarz and Truszczyński [7] which indicates that standard autoepistemic reasoning can be captured by means of strong autoepistemic reasoning. We exploit this characterization to extract a suitably reduced translation function which supports converting autoepistemic decision problems into decision problems of strong autoepistemic logic. A novel translation function is devised to capture strong autoepistemic reasoning in terms of default reasoning. The basic idea is to represent autoepistemic introspection in terms of default rules. Consequently, iterative expansions of the autoepistemic theory coincide with the extensions of the translation. Moreover, propositionally consistent stable expansions can be captured by composing the presented translation functions.

The remaining sections of the paper are organized as follows. In Sections 2 and 3, major definitions of default and autoepistemic logic are briefly reviewed. The reduction of the translation by Marek et al. [7] takes place in Section 4. In Section 5, a novel translation of autoepistemic theories into default theories is presented. It is shown that iterative expansions of an autoepistemic theory correspond exactly to the extensions of its translation. Practical applications of the results are considered in Section 6. Basic algorithms for solving autoepistemic decision problems using default decision procedures are developed. The correctness of the algorithms is also addressed. Finally, we present our conclusions in Section 7.

## 2 Default Theories

Reiter [14] introduced *default logic* as a formalization of default reasoning. A *default theory* consists of a pair $\langle D, W\rangle$ where $D$ is a set of *default rules* of the form $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma}$ and $W$ is a propositional theory. The idea behind a default rule is that the *consequent* $\gamma$ can be inferred given the *prerequisite* $\alpha$ can and it is consistent to assume the *justifications* $\beta_1, ..., \beta_n$.

The possible sets of conclusions based on a default theory $\langle D, W\rangle$ are called *extensions* of the default theory. Marek and Truszczyński [9] define them as follows. The *reduction* $D_S$ of a set of default rules $D$ with respect to a theory $S$ is the set of inference rules $\frac{\alpha}{\gamma}$ for which $\frac{\alpha:\beta_1,...,\beta_n}{\gamma} \in D$ and every $\beta_i$, $0 < i \leq n$, is consistent with $S$. Given a set $R$ of *inference rules* of the form $\frac{\alpha}{\gamma}$ and a propositional theory $W$, the theory $\mathrm{Cn}^R(W)$ is the closure of $W$ under propositional consequence and inference by the rules of $R$. Extensions of a default theory are simply propositional theories satisfying the fix-point condition of Definition 1.

**Definition 1 (Marek and Truszczyński [9])** *Let $\langle D, W\rangle$ be a default theory. A theory $S$ is an extension of $\langle D, W\rangle$ if and only if $S = \mathrm{Cn}^{D_S}(W)$.*

A default theory can have several extensions and there are even default theories without extensions. Thus it is not always possible to assign a unique set of conclusions to a default theory. One possibility to approach this problem is to consider the intersection of extensions as the correct set of conclusions associated with the default theory. This corresponds to *cautious* reasoning strategy in contrast to *brave* one where one of the extensions is selected, e.g., using an external criterion.

# 3 Autoepistemic Theories

Moore [10] introduced *autoepistemic logic* that models the possible sets of beliefs of an introspective agent given a set of premises as its initial assumptions. The agent is capable of *positive and negative introspection*: it decides what it believes and disbelieves on the basis of the logical consequences of its premises, beliefs and disbeliefs. The premises of the agent form an *autoepistemic theory*, i.e. a subset of an *autoepistemic language* $\mathcal{L}_{\mathrm{B}}$. This language extends a propositional language $\mathcal{L}$ by sentences of the form $\mathbf{B}\phi$ meaning that "$\phi$ is believed"[1]. Such sentences referred to as *belief atoms* in the sequel. In autoepistemic logic, the logical consequences $\mathrm{Cn}(\Sigma)$ of an autoepistemic theory $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$ are simply its propositional consequences $\{\phi \mid \Sigma \models \phi\}$ when belief atoms are treated as atomic propositions. Moore [10] defines the sets of beliefs of the agent as *stable expansions* of the premises.

**Definition 2 (Moore [10])** *A theory $\Delta \subseteq \mathcal{L}_{\mathrm{B}}$ is a* stable expansion *of an autoepistemic theory $\Sigma$ if and only if*

$$\Delta = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \neg\mathbf{B}\overline{\Delta})$$

*where $\mathbf{B}\Gamma = \{\mathbf{B}\phi \mid \phi \in \Gamma\}$, $\neg\Gamma = \{\neg\phi \mid \phi \in \Gamma\}$ and $\overline{\Gamma} = \mathcal{L}_{\mathrm{B}} - \Gamma$ for all $\Gamma \subseteq \mathcal{L}_{\mathrm{B}}$.*

[1] We use $\mathbf{B}$ as a modality for belief, instead of the conventional $\mathbf{L}$.

Because the beliefs $\mathbf{B}\Delta$ of the agent are present in the definition, there are autoepistemic theories having stable expansions in which beliefs support themselves. Such stable expansions are commonly known as expansions that are *weakly grounded* in the premises. In the following, the set of sentences $\Delta_1$ serves as an example of an expansion having this property.

**Example 3** *The set of premises* $\Sigma = \{\mathbf{B}\mathrm{p} \rightarrow \mathrm{p}\}$ *has two stable expansions:* $\Delta_1 = \{\mathrm{p}, \mathbf{B}\mathrm{p}, ...\}$ *and* $\Delta_2 = \{\neg\mathbf{B}\mathrm{p}, ...\}$. *In the first expansion, the sentence p is believed (*$\mathrm{p} \in \Delta_1$*) and in the other it is disbelieved (*$\mathrm{p} \notin \Delta_2$*). The expansion* $\Delta_1$ *is weakly grounded in* $\Sigma$*: the belief in* p *is supported by itself.*

To exclude such stable expansions, Marek and Truszczyński [8] propose a groundedness condition for stable expansions – stipulating that the beliefs $\Delta$ should be dependent on the premises $\Sigma$ and the negative beliefs $\neg\mathbf{B}\overline{\Delta}$ only. This is formalized in Definition 4 as the notion of *iterative expansions* of an autoepistemic theory. Here we use the notion of **B**-*proofs* introduced by Gottlob [2]. **B**-proofs serve as a characterization of provability in a non-standard modal logic N [8] which is effectively propositional calculus equipped with the necessitation rule $\phi \vdash \mathbf{B}\phi$ for the modality $\mathbf{B}$. More formally, $\Sigma \vdash_{\mathbf{B}} \phi$ if and only if there exists a finite sequence $\mathbf{B}\phi_1, ..., \mathbf{B}\phi_n$ of belief atoms such that for each $0 < i \leq n$ it holds that $\Sigma \cup \{\mathbf{B}\phi_1, ..., \mathbf{B}\phi_{i-1}\} \models \phi_i$ and $\Sigma \cup \{\mathbf{B}\phi_1, ..., \mathbf{B}\phi_n\} \models \phi$. A sequence of the above kind is a **B**-proof of $\phi$ from $\Sigma$. The idea is of course that each step of a **B**-proof corresponds to an application of the necessitation rule $\phi \vdash \mathbf{B}\phi$.

**Definition 4 (Marek and Truszczyński [8])**
*A theory* $\Delta \subseteq \mathcal{L}_{\mathrm{B}}$ *is an iterative expansion of an autoepistemic theory* $\Sigma$ *if and only if* $\Delta = \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \neg\mathbf{B}\overline{\Delta})$.

Every iterative expansion of an autoepistemic theory $\Sigma$ is also a stable expansion of $\Sigma$, but not necessarily vice versa. *Strong autoepistemic reasoning* is based on the iterative expansions of an autoepistemic theory. In analogy to default logic, an autoepistemic theory may not have a unique stable/iterative expansion. Therefore the cautious and brave reasoning strategies are useful in autoepistemic reasoning as well.

# 4 Capturing Stable Expansions with Iterative Expansions

Let us now consider the possibility of capturing standard autoepistemic reasoning in terms of strong autoepistemic reasoning. Marek, Schwarz and Truszczyński [7] characterize stable expansions in terms of expansions in various nonmonotonic modal logics. Theorem 5 is implied by their results. Here **B** refers to the closure with respect to **B**-provability (or equivalently the necessitation rule $\phi \vdash \mathbf{B}\phi$) and **5** is the standard axiom schema $\neg\mathbf{B}\neg\mathbf{B}\phi \rightarrow \mathbf{B}\phi$.

**Theorem 5 (Marek, Schwarz and Truszczyński [7])**
*A propositionally consistent theory* $\Delta \subseteq \mathcal{L}_{\mathrm{B}}$ *is a stable expansion of* $\Sigma$ *if and only if* $\Delta = \mathrm{Cn}_{\mathbf{B5}}(\Sigma \cup \neg\mathbf{B}\overline{\Delta})$.

Theorem 5 provides a reduction of standard autoepistemic reasoning into strong autoepistemic reasoning provided that propositionally inconsistent expansions are not considered. This is achieved simply by augmenting an autoepistemic theory $\Sigma$ with all instances of the schema **5**. However, such an operation seems rather impractical from the point of view of automating autoepistemic reasoning as it makes even finite autoepistemic theories infinite. In the following, we show that the instances $\neg\mathbf{B}\neg\mathbf{B}\phi \rightarrow \mathbf{B}\phi$ of the schema **5** for which the belief atom $\mathbf{B}\phi$ appears in $\Sigma$ are sufficient. To give an exact definition of the translation function, we let $\mathrm{Rba}(\phi)$ denote the set of belief atoms that recursively appear in a sentence $\phi$ and $\mathrm{Rba}(\Sigma) = \bigcup\{\mathrm{Rba}(\phi) \mid \phi \in \Sigma\}$. For instance, $\mathrm{Rba}(\{\mathbf{B}(\neg\mathbf{B}\mathrm{p} \wedge \mathrm{q}), \mathbf{B}\mathrm{q}\}) = \{\mathbf{B}(\neg\mathbf{B}\mathrm{p} \wedge \mathrm{q}), \mathbf{B}\mathrm{p}, \mathbf{B}\mathrm{q}\}$. The translation function is given in Definition 6. In Proposition 7, we show that the stable expansions of the autoepistemic theories $\Sigma$ and $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$ coincide.

**Definition 6** *A standard autoepistemic theory* $\Sigma$ *is translated into a strong autoepistemic theory*

$$\mathrm{Tr}_{\mathrm{SAL}}(\Sigma) = \Sigma \cup \{\neg\mathbf{B}\neg\mathbf{B}\phi \rightarrow \mathbf{B}\phi \mid \mathbf{B}\phi \in \mathrm{Rba}(\Sigma)\}.$$

**Proposition 7** *A propositionally consistent theory* $\Delta \subseteq \mathcal{L}_{\mathrm{B}}$ *is a stable expansion of an autoepistemic theory* $\Sigma$ *if and only if* $\Delta$ *is a stable expansion of* $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$.

*Proof.* This follows easily because the sentences $\neg\mathbf{B}\neg\mathbf{B}\phi \rightarrow \mathbf{B}\phi$ satisfying $\mathbf{B}\phi \in \mathrm{Rba}(\Sigma)$ are instances of the schema **5**. Thus the theory $\mathrm{Cn}_{\mathbf{B5}}(\Sigma \cup \neg\mathbf{B}\overline{\Delta})$ coincides with the theory $\mathrm{Cn}_{\mathbf{B5}}(\mathrm{Tr}_{\mathrm{SAL}}(\Sigma) \cup \neg\mathbf{B}\overline{\Delta})$. The rest follows by Theorem 5. □

The propositionally consistent stable expansions of the translation $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$ coincide with the propositionally consistent iterative expansions of $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$:

**Proposition 8** *Let* $\Sigma$ *be an autoepistemic theory. A propositionally consistent theory* $\Delta \subseteq \mathcal{L}_{\mathrm{B}}$ *is a stable expansion of* $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$ *if and only if* $\Delta$ *is an iterative expansion of* $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$.

*Proof sketch.* ($\Rightarrow$) For any sentence $\phi$, the *belief depth* $\mathrm{Bd}(\phi)$ gives the maximum nesting level of belief operators in $\phi$. It can be proved by induction on $\mathrm{Bd}(\phi)$ that $\phi \in \Delta$ implies $\phi \in \mathrm{Cn}_{\mathbf{B}}(\mathrm{Tr}_{\mathrm{SAL}}(\Sigma) \cup \neg\mathbf{B}\overline{\Delta})$. Thus $\Delta \subseteq \mathrm{Cn}_{\mathbf{B}}(\mathrm{Tr}_{\mathrm{SAL}}(\Sigma) \cup \neg\mathbf{B}\overline{\Delta})$. The converse holds, because $\Delta = \mathrm{Cn}_{\mathbf{B5}}(\mathrm{Tr}_{\mathrm{SAL}}(\Sigma) \cup \neg\mathbf{B}\overline{\Delta})$ by Theorem 5. ($\Leftarrow$) Iterative expansions $\Delta$ of $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$ are always stable expansions of $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$. □

Theorem 9 is obtained as a corollary of Propositions 7 and 8. It asserts that the translation function of Definition 6 is faithful provided that only propositionally consistent stable expansions are considered.

**Theorem 9** *A propositionally consistent theory $\Delta \subseteq \mathcal{L}_{\mathrm{B}}$ is a stable expansion of an autoepistemic theory $\Sigma$ if and only if $\Delta$ is an iterative expansion of* $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$.

In Example 10, an autoepistemic theory having a propositionally inconsistent stable expansion is presented. The translation $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$ is

$$\Sigma \cup \{\neg\mathbf{B}\neg\mathbf{B}\mathrm{p} \rightarrow \mathbf{B}\mathrm{p}, \neg\mathbf{B}\neg\mathbf{B}\mathrm{q} \rightarrow \mathbf{B}\mathrm{q}\}.$$

The theories $\Delta_1$, $\Delta_2$ and $\Delta_3$ (of the ones listed below) are iterative expansions of $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$, but $\Delta_4$ is not. This is because $\neg\mathbf{B}\overline{\Delta_4}$ is empty and the two instances of the schema **5** are not sufficient to **B**-prove $\mathbf{B}\mathrm{p}$, $\mathbf{B}\mathrm{q}$ and $\mathrm{r} \wedge \neg\mathrm{r}$ from $\Sigma$.

**Example 10** *Consider an autoepistemic theory*

$$\Sigma = \{\mathbf{B}\mathrm{p} \rightarrow \mathrm{p} \wedge \mathrm{r}, \mathbf{B}\mathrm{q} \rightarrow \mathrm{q} \wedge \neg\mathrm{r}\}$$

*which has the following four stable expansions:*

$$\begin{aligned}
\Delta_1 &= \{\neg\mathbf{B}\mathrm{p}, \neg\mathbf{B}\mathrm{q}, \ldots\}, \\
\Delta_2 &= \{\mathbf{B}\mathrm{p}, \mathrm{p}, \mathrm{r}, \neg\mathbf{B}\mathrm{q}, \ldots\}, \\
\Delta_3 &= \{\neg\mathbf{B}\mathrm{p}, \mathbf{B}\mathrm{q}, \mathrm{q}, \neg\mathrm{r}, \ldots\} \textit{ and} \\
\Delta_4 &= \{\mathbf{B}\mathrm{p}, \mathbf{B}\mathrm{q}, \mathrm{r} \wedge \neg\mathrm{r}, \ldots\} = \mathcal{L}_{\mathrm{B}}.
\end{aligned}$$

*Of these, the expansion $\Delta_1$ is the only iterative expansion of $\Sigma$ and the expansion $\Delta_4$ is propositionally inconsistent.*

# 5 Capturing Iterative Expansions with Default Extensions

In this section, we concentrate on embedding strong autoepistemic logic in default logic. Before presenting the actual translation function, we make a crucial remark on the languages corresponding to theories. Every autoepistemic theory $\Sigma$ has a set of atomic propositions $\mathcal{P}_{\mathrm{B}}$ and an autoepistemic language $\mathcal{L}_{\mathrm{B}}$ associated with it. When an autoepistemic theory $\Sigma$ is translated into a default theory, the autoepistemic language $\mathcal{L}_{\mathrm{B}}$ is viewed as a *propositional language* based on a set of atomic propositions $\mathcal{P} = \mathcal{P}_{\mathrm{B}} \cup \{\mathbf{B}\psi \mid \psi \in \mathcal{L}_{\mathrm{B}}\}$. In this setting, the two languages $\mathcal{L}_{\mathrm{B}}$ and $\mathcal{L}$ coincide. This is actually what the convention of "treating belief atoms as atomic propositions" means in autoepistemic logic.

The novel translation function is given in Definition 11. The autoepistemic theory $\Sigma$ remains unchanged but it is seen as a propositional theory $\Sigma \subseteq \mathcal{L}$ in the way explained above. The default rules of the form $\frac{\psi:}{\mathbf{B}\psi}$ capture the positive introspection on the default side. If $\psi$ can be inferred $\psi$ is believed and $\mathbf{B}\psi$ can be inferred as an indication of belief in $\psi$. In a similar way, the purpose of the default rules of the form $\frac{:\neg\psi}{\neg\mathbf{B}\psi}$ is to capture the negative introspection. If it is consistent to assume $\neg\psi$ in an extension $S$ (a propositionally closed theory), it holds that $\psi \notin S$. In this case, it is natural to disbelieve $\psi$ and infer the sentence $\neg\mathbf{B}\psi$. This is how the autoepistemic introspection principles are represented as default rules in the translation.

**Definition 11** *An autoepistemic theory* $\Sigma$ *is translated into a default theory*

$$\mathrm{Tr}_{\mathrm{DL}}(\Sigma) = \langle\{\tfrac{\psi:}{\mathbf{B}\psi} \mid \psi \in \mathcal{L}_{\mathrm{B}}\} \cup \{\tfrac{:\neg\psi}{\neg\mathbf{B}\psi} \mid \psi \in \mathcal{L}_{\mathrm{B}}\}, \Sigma\rangle.$$

In Proposition 12, it is shown how $\mathbf{B}$-provability can be simulated in terms of the inference rules $R = \{\frac{\psi}{\mathbf{B}\psi} \mid \psi \in \mathcal{L}_{\mathrm{B}}\}$. In fact, these inference rules are instances of the necessitation rule and they are also present (as default rules) in the set of default rules of Definition 11.

**Proposition 12** *If* $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$ *and* $R = \{\frac{\psi}{\mathbf{B}\psi} \mid \psi \in \mathcal{L}_{\mathrm{B}}\}$, *then* $\mathrm{Cn}_{\mathbf{B}}(\Sigma) = \mathrm{Cn}^{R}(\Sigma)$.

*Proof sketch.* ($\subseteq$) It can be proved by induction on the lengths of $\mathbf{B}$-proofs that $\Sigma \vdash_{\mathbf{B}} \phi$ implies $\phi \in \mathrm{Cn}^{R}(\Sigma)$. Thus $\mathrm{Cn}_{\mathbf{B}}(\Sigma) \subseteq \mathrm{Cn}^{R}(\Sigma)$.

($\supseteq$) By the results of Marek and Truszczyński [9] $\mathrm{Cn}^{R}(\Sigma) = \mathrm{Cn}(\Sigma \cup \mathbf{A}^{R,\Sigma} \uparrow^{\omega} (\emptyset))$ where the operator $\mathbf{A}^{R,\Sigma}$ is defined by

$$\mathbf{A}^{R,\Sigma}(\Gamma) = \Gamma \cup \{\gamma \mid \frac{\alpha}{\gamma} \in R \text{ and } \alpha \in \mathrm{Cn}(\Sigma \cup \Gamma)\}$$

for subsets $\Gamma$ of $\{\gamma \mid \frac{\alpha}{\gamma} \in R\}$. It can be proved by induction on $i$ that $\Gamma_i = \mathbf{A}^{R,\Sigma} \uparrow^{i} (\emptyset)$ is a subset of $\mathrm{Cn}_{\mathbf{B}}(\Sigma)$ and $\mathrm{Cn}(\Sigma \cup \Gamma_i) \subseteq \mathrm{Cn}_{\mathbf{B}}(\Sigma)$. Thus $\Gamma_\omega \subseteq \mathrm{Cn}_{\mathbf{B}}(\Sigma)$ and $\mathrm{Cn}^{R}(\Sigma) = \mathrm{Cn}(\Sigma \cup \Gamma_\omega) \subseteq \mathrm{Cn}_{\mathbf{B}}(\Sigma)$. $\square$

By our main result — Theorem 13 — the iterative expansions of an autoepistemic theory $\Sigma$ and the extensions of its translation $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$ are exactly the same theories. This indicates that equipping an autoepistemic theory with the default rules of Definition 11 yields exactly the strong autoepistemic logic [8] proposed by Marek and Truszczyński.

**Theorem 13** *A theory* $\Delta \subseteq \mathcal{L}_{\mathrm{B}}$ *is an iterative expansion of an autoepistemic theory* $\Sigma$ *if and only if* $\Delta \subseteq \mathcal{L}$ *is an extension of* $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$.

*Proof.* Let $D$ be the set of default rules of Definition 11 and $R$ the set of inference rules of Proposition 12. As $\Delta$ is closed with respect to propositional consequence, the reduction $D_\Delta = R \cup \{\frac{\top}{\neg\mathbf{B}\psi} \mid \psi \in \overline{\Delta}\}$. It follows that $\mathrm{Cn}^{D_\Delta}(\Sigma) = \mathrm{Cn}^{R}(\Sigma \cup \neg\mathbf{B}\overline{\Delta}) = \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \neg\mathbf{B}\overline{\Delta})$ by Proposition 12. Under this equality, the fix-point conditions of Definitions 1 and 4 coincide. $\square$

Example 14 demonstrates how the iterative expansions of an autoepistemic theory $\Sigma$ can be determined in terms of its translation $\mathrm{Tr}_{\mathrm{DL}}(\Sigma) = \langle D, W\rangle$. We emphasize that $\langle D, W\rangle$ does not have an extension $S'$ containing p. To verify this, notice that $D_{S'} = \{\frac{\mathrm{p}}{\mathbf{B}\mathrm{p}}, \ldots\}$ and $\mathrm{Cn}^{D_{S'}}(W)$ does not contain p because p cannot be inferred by the inference rule $\frac{\mathrm{p}}{\mathbf{B}\mathrm{p}}$ and the premise $\mathbf{B}\mathrm{p} \rightarrow \mathrm{p}$. Therefore $S' \neq \mathrm{Cn}^{D_{S'}}(W)$ is not an extension of $\langle D, W\rangle$ as expected by Theorem 13 and Example 3.

**Example 14** *Consider an autoepistemic theory* $\Sigma$ *consisting of the sentence* $\mathbf{B}\mathrm{p} \rightarrow \mathrm{p}$ *only. The translation* $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$ *is a default theory* $\langle D, W\rangle$ *where* $W = \{\mathbf{B}\mathrm{p} \rightarrow \mathrm{p}\}$ *and* $D$ *is a set of default rules* $\{\frac{\mathrm{p}:}{\mathbf{B}\mathrm{p}}, \frac{:\neg\mathrm{p}}{\neg\mathbf{B}\mathrm{p}}, ...\}$*. This default theory has only one extension* $S$ *that does not contain* $\mathrm{p}$*. To see this, notice that* $D_S = \{\frac{\mathrm{p}}{\mathbf{B}\mathrm{p}}, \frac{\top}{\neg\mathbf{B}\mathrm{p}}, ...\}$ *and* $\mathrm{Cn}^{D_S}(W)$ *contains* $\neg\mathbf{B}\mathrm{p}$ *but not* $\mathrm{p}$*. It follows that* $S = \mathrm{Cn}^{D_S}(W)$ *and* $S$ *is an extension of* $\langle D, W\rangle$*.*

If we consider the composed translation $\mathrm{Tr}_{\mathrm{DL}}(\mathrm{Tr}_{\mathrm{SAL}}(\Sigma))$, we obtain a default theory $\langle D, W \cup \{\neg\mathbf{B}\neg\mathbf{B}\mathrm{p} \rightarrow \mathbf{B}\mathrm{p}\}\rangle$ having an additional extension

$$\{\neg\mathbf{B}\neg\mathbf{B}\mathrm{p}, \mathbf{B}\mathrm{p}, \mathrm{p}, \ldots\}.$$

Note that $D$ contains the default rule $\frac{:\neg\neg\mathbf{B}\mathrm{p}}{\neg\mathbf{B}\neg\mathbf{B}\mathrm{p}}$ by means of which $\neg\mathbf{B}\neg\mathbf{B}\mathrm{p}$ is inferred. As suggested by this example, we combine Theorems 9 and 13 to the following corollary.

**Corollary 15** *A propositionally consistent theory* $\Delta \subseteq \mathcal{L}_{\mathrm{B}}$ *is a stable expansion of an autoepistemic theory* $\Sigma$ *if and only if* $\Delta$ *is an extension of* $\mathrm{Tr}_{\mathrm{DL}}(\mathrm{Tr}_{\mathrm{SAL}}(\Sigma))$.

## 6 Decision Procedures for Autoepistemic Logic

This section provides a demonstration how the results of previous sections apply in practice. The translations of Definitions 6 and 11 provide a systematic way of solving autoepistemic decision problems in terms of default decision procedures. Therefore, let us assume that there is available a decision procedure $\mathsf{DD1}$ for default logic – following the cautious strategy. The call $\mathsf{DD1}(D, W, \phi)$ returns the answer $\mathsf{yes}$ (of the possible answers $\mathsf{yes}$ and $\mathsf{no}$) if and only if the sentence $\phi$ belongs to every extension of the *finite* default theory $\langle D, W\rangle$. Given a *finite* autoepistemic theory $\Sigma$ and a sentence $\phi$ as input, a decision procedure $\mathsf{AD1}$ for strong autoepistemic logic acts as follows. Two sets of default rules $D^+$ (positive introspection) and $D^-$ (negative introspection) are constructed according to Definition 11, but a restriction to the belief atoms of $\mathrm{Rba}(\Sigma \cup \{\phi\})$ is imposed. This is to keep the set of default rules finite. As the final operation in $\mathsf{AD1}$, the procedure $\mathsf{DD1}$ is consulted with the arguments $D^+ \cup D^-$, $\Sigma$ and $\phi$.

Function $\mathsf{AD1}(\Sigma, \phi)$: answer;
Where
  $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$ is finite and $\phi \in \mathcal{L}_{\mathrm{B}}$;
Begin
  $D^+ := \{\frac{\chi:}{\mathbf{B}\chi} \mid \mathbf{B}\chi \in \mathrm{Rba}(\Sigma \cup \{\phi\})\}$;
  $D^- := \{\frac{:\neg\chi}{\neg\mathbf{B}\chi} \mid \mathbf{B}\chi \in \mathrm{Rba}(\Sigma \cup \{\phi\})\}$;
  $\mathsf{AD1} := \mathsf{DD1}(D^+ \cup D^-, \Sigma, \phi)$
End

Let us justify the use of the default theory $\langle D^+ \cup D^-, \Sigma\rangle$ in $\mathsf{AD1}$. Proposition 16 gives a close correspondence between the extensions of this default theory and $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$. In fact, Proposition 16 provides a kind of a *finitary characterization* for autoepistemic logic. The extensions of the default theory $\langle D^+ \cup D^-, \Sigma\rangle$ have a close connection to *full sets* introduced by Niemelä [11] and Gottlob [2]. The correctness of $\mathsf{AD1}$ is addressed in Theorem 17.

**Proposition 16** *There is an extension $S$ of $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$ if and only if there is an extension $S' \subseteq S$ of $\langle D^+ \cup D^-, \Sigma\rangle$. Given such a pair of extensions and a sentence $\psi$ such that $\mathrm{Rba}(\psi) \subseteq \mathrm{Rba}(\Sigma \cup \{\phi\})$, $\psi \in S$ if and only if $\psi \in S'$.*

*Proof sketch.* For a set of belief atoms $\Gamma$ and a set of belief literals $\Lambda$, the set $\{\mathbf{B}\phi \in \Lambda \mid \mathbf{B}\phi \in \Gamma\} \cup \{\neg\mathbf{B}\phi \in \Lambda \mid \mathbf{B}\phi \in \Gamma\}$ is denoted by $\Pi_\Gamma(\Lambda)$. ($\Rightarrow$) Given an extension $S$ of $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$, the corresponding extension of $\langle D^+ \cup D^-, \Sigma\rangle$ is

$$S' = \mathrm{Cn}(\Sigma \cup \Pi_{\mathrm{Rba}(\Sigma\cup\{\phi\})}(\mathbf{B}S \cup \neg\mathbf{B}\overline{S})).$$

This is proved by showing that (i) $S' \subseteq S$, (ii) $\psi \in S$ iff $\psi \in S'$ for all $\psi$ satisfying $\mathrm{Rba}(\psi) \subseteq \mathrm{Rba}(\Sigma \cup \{\phi\})$, (iii) $\Pi_{\mathrm{Rba}(\Sigma\cup\{\phi\})}(\mathbf{B}S \cup \neg\mathbf{B}\overline{S}) = \Pi_{\mathrm{Rba}(\Sigma\cup\{\phi\})}(\mathbf{B}S' \cup \neg\mathbf{B}\overline{S'})$ and (iv) $S' = \mathrm{Cn}^{D^+{}_{S'}}(\Sigma \cup \Pi_{\mathrm{Rba}(\Sigma\cup\{\phi\})}(\neg\mathbf{B}\overline{S'})) = \mathrm{Cn}^{(D^+\cup D^-)_{S'}}(\Sigma)$.

($\Leftarrow$) Given an extension $S'$ of $\langle D^+ \cup D^-, \Sigma\rangle$, define the sets

$$\begin{aligned}&\Lambda^+ = \Pi_{\mathrm{Rba}(\Sigma\cup\{\phi\})}(\mathbf{B}S'),\\ &\Lambda^- = \Pi_{\mathrm{Rba}(\Sigma\cup\{\phi\})}(\neg\mathbf{B}\overline{S'}) \text{ and}\\ &\Lambda = \Lambda^+ \cup \Lambda^-.\end{aligned}$$

Then it can be shown that $S' = \mathrm{Cn}^{(D^+\cup D^-)_{S'}}(\Sigma) = \mathrm{Cn}^{D^+{}_{S'}}(\Sigma \cup \Lambda^-) = \mathrm{Cn}(\Sigma \cup \Lambda)$. It follows that $\Pi_{\mathrm{Rba}(\Sigma)}(\Lambda)$ is $\Sigma$-full [11] and $\Pi_{\mathrm{Rba}(\Sigma)}(\Lambda)$ satisfies the groundedness condition of Gottlob [2]. Consequently, there is an iterative expansion $S$ of $\Sigma$ which is also an extension of $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$ by Theorem 13. To prove $S' \subseteq S$, one has to show that $\Lambda \subseteq S$. This is easy after proving by induction on $\mathrm{Bd}(\psi)$ that $\psi \in S'$ if and only if $\psi \in S$ for all $\psi$ satisfying $\mathrm{Rba}(\psi) \subseteq \mathrm{Rba}(\Sigma \cup \{\phi\})$. $\square$

**Theorem 17** *Let $\Sigma$ be a finite autoepistemic theory and $\phi$ a sentence. The call* $\mathsf{AD1}(\Sigma, \phi)$ *returns* $\mathsf{yes}$ *if and only if $\phi$ belongs to every iterative expansion of $\Sigma$.*

*Proof.* By Proposition 16, the sentence $\phi$ belongs to every extension of $\langle D^+ \cup D^-, \Sigma\rangle$ if and only if $\phi$ belongs to every extension of $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$. Therefore – by the correctness of $\mathsf{DD1}$ – the call $\mathsf{DD1}(D^+ \cup D^-, \Sigma, \phi)$ returns $\mathsf{yes}$ if and only if $\phi$ belongs to every extension of $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$. By Theorem 13 this is the case if and only if $\phi$ belongs to every iterative extension of $\Sigma$. $\square$

The decision procedure $\mathsf{AD1}$ serves as basis when devising a similar decision procedure $\mathsf{AD2}$ for conventional autoepistemic logic. Because Theorem 9 is restricted to consistent stable expansions, we have to make an additional restriction on $\mathsf{DD1}$ – leading to another decision procedure $\mathsf{DD2}$. The call

$\mathsf{DD2}(D, W, \phi)$ returns $\mathsf{yes}$ if and only if $\phi$ belongs to every *propositionally consistent* extension of $\langle D, W\rangle$. The decision procedure $\mathsf{AD2}$ is shown correct in Theorem 18.

Function $\mathsf{AD2}(\Sigma, \phi)$: answer;
Where
  $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$ is finite and $\phi \in \mathcal{L}_{\mathrm{B}}$;
Begin
  $\Sigma' := \Sigma \cup \{\neg\mathbf{B}\neg\mathbf{B}\chi \rightarrow \mathbf{B}\chi \mid \mathbf{B}\chi \in \mathrm{Rba}(\Sigma)\}$;
  $D^{+} := \{\frac{\chi:}{\mathbf{B}\chi} \mid \mathbf{B}\chi \in \mathrm{Rba}(\Sigma' \cup \{\phi\})\}$;
  $D^{-} := \{\frac{:\neg\chi}{\neg\mathbf{B}\chi} \mid \mathbf{B}\chi \in \mathrm{Rba}(\Sigma' \cup \{\phi\})\}$;
  $\mathsf{AD2} := \mathsf{DD2}(D^{+} \cup D^{-}, \Sigma', \phi)$
End

**Theorem 18** *Let $\Sigma$ be a finite autoepistemic theory and $\phi$ a sentence. The call $\mathsf{AD2}(\Sigma, \phi)$ returns* $\mathsf{yes}$ *if and only if $\phi$ belongs to every propositionally consistent stable expansion of $\Sigma$.*

*Proof.* Clearly, $\Sigma'$ is the translation $\mathrm{Tr}_{\mathrm{SAL}}(\Sigma)$. By Proposition 16, the sentence $\phi$ belongs to every *propositionally consistent* extension of $\langle D^{+} \cup D^{-}, \Sigma'\rangle$ if and only if $\phi$ belongs to every *propositionally consistent* extension of $\mathrm{Tr}_{\mathrm{DL}}(\Sigma')$. By the correctness of $\mathsf{DD2}$, the call $\mathsf{DD2}(D^{+} \cup D^{-}, \Sigma', \phi)$ returns $\mathsf{yes}$ if and only if $\phi$ belongs to every propositionally consistent extension of $\mathrm{Tr}_{\mathrm{DL}}(\Sigma') = \mathrm{Tr}_{\mathrm{DL}}(\mathrm{Tr}_{\mathrm{SAL}}(\Sigma))$. By Corollary 15 this is the case if and only if $\phi$ belongs to every propositionally consistent stable extension of $\Sigma$. $\square$

The translation functions used in this paper enjoy some nice properties. First of all, they are *modular*, i.e. local modifications in the theory being translated cause only local modifications in its translation. Secondly, the size of the translation grows at most *linearly* in the size of the translated theory. Moreover, the required default rules are of two generic forms allowing representation in constant space.

# 7 Discussion and Conclusions

In this paper, we demonstrate how autoepistemic introspection can be represented in terms of default rules. This leads to natural translations of autoepistemic theories into default theories such that standard and strong autoepistemic reasoning are reduced to default reasoning. We present the following conclusions:

*Autoepistemic logic and default logic are of equal generality.* We propose a new interpretation of the earlier research results which unanimously suggest that default reasoning is a special case of autoepistemic reasoning. The results of this paper imply that the converse claim is also justified. This indicates that the two logics are equally general: things expressible in autoepistemic logic are expressible in default logic and vice versa. This view is in harmony with the recent complexity results (see e.g. [2, 11]) on non-monotonic reasoning: the major decision problems of the two logics belong to same classes of the polynomial time hierarchy PH [1]. In fact, existence of polynomial time transformations between the decision problems of the two logics is already implied by the complexity results. A contribution of this paper is to provide an actual translation function having reasonable properties.

*A very simple default reasoning system is sufficient to cover autoepistemic and default reasoning.* In a simplified setting, such a system consists of a propositional theory $W$ augmented with inference rules of the form $\frac{\alpha}{\gamma}$ and $\frac{\sim\beta}{\gamma}$. The latter form has a natural interpretation that "$\gamma$ can be inferred in the case $\beta$ cannot"[2]. Arbitrary default theories are also reducible to the simplified form by composing Truszczyński's translation function [15] with the one presented in this paper. This suggests that the default rules of the original form have "syntactic sugar" and they can be replaced by rules of a simpler form whenever necessary. Taking this interpretation of our results, we conclude that default logic as well as autoepistemic logic are based on very simple primitives. This is of theoretical importance, since these formalizations cover a substantial fragment of non-monotonic reasoning.

*At the level of decision procedures, there is a systematic way of solving autoepistemic decision problems in terms of default decision procedures.* This has practical importance – existing default reasoning systems can be used for autoepistemic reasoning. For the same reason, it is also important that the developed translations are modular, linear and allow succinct representation of default rules. On the other hand, we are not claiming that decision procedures AD1 and AD2 are necessarily optimal. We expect that dedicated decisions procedures such as those developed by Niemelä [13] are more efficient.

# 8 Acknowledgments

The author thanks Dr Ilkka Niemelä and anonymous referees for their comments and suggestions to improve the paper. The financial support from Alfred Kordelin Foundation, Emil Aaltonen Foundation and Foundation of Technology is also acknowledged with gratitude.

# References

[1] J.L. Balcázar, I. Díaz, and J. Gabarró, *Structural Complexity I*, Springer-Verlag, Berlin, 1988.

[2] A word of caution is needed here: this kind of inference rules must be applied globally (in analogy to Definition 1 where a set of default rules is globally reduced). Otherwise the application order of such inference rules becomes significant.

[2] G. Gottlob, 'Complexity results for nonmonotonic logics', *Journal of Logic and Computation*, **2**(3), 397–425, (June 1992).

[3] G. Gottlob, 'The power of beliefs or translating default logic into standard autoepistemic logic', in *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pp. 570–575, Chambéry, France, (August 1993). Morgan Kaufmann Publishers.

[4] K. Konolige, 'On the relation between default and autoepistemic logic', *Artificial Intelligence*, **35**, 343–382, (1988).

[5] K. Konolige, 'Errata: On the relation between default and autoepistemic logic [Artificial Intelligence 35 (1988) 343-382]', *Artificial Intelligence*, **41**, 115, (1989).

[6] K. Konolige, 'Ideal introspective belief', in *Proceedings of the 10th National Conference on Artificial Intelligence*, pp. 635–641, San Jose, CA, USA, (July 1992). The MIT Press.

[7] W. Marek, G.F. Schwarz, and M. Truszczyński, 'Modal nonmonotonic logics: Ranges, characterization, computation', *Journal of the ACM*, **40**(4), 963–990, (1993).

[8] W. Marek and M. Truszczyński, 'Modal logic for default reasoning', *Annals of Mathematics and Artificial Intelligence*, **1**, 275–302, (1990).

[9] W. Marek and M. Truszczyński, *Nonmonotonic Logic: Context-Dependent Reasoning*, Springer-Verlag, Berlin, 1993.

[10] R.C. Moore, 'Semantical considerations on nonmonotonic logic', in *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pp. 272–279, Karlsruhe, FRG, (August 1983). Morgan Kaufmann Publishers.

[11] I. Niemelä, 'On the decidability and complexity of autoepistemic reasoning', *Fundamenta Informaticae*, **17**(1,2), 117–155, (1992).

[12] I. Niemelä, 'A unifying framework for nonmonotonic reasoning', in *Proceedings of the 10th European Conference on Artificial Intelligence*, pp. 334–338, Vienna, Austria, (August 1992). John Wiley.

[13] I. Niemelä, 'A decision method for nonmonotonic reasoning based on autoepistemic reasoning', *Journal of Automated Reasoning*, **14**, 3–42, (1995).

[14] R. Reiter, 'A logic for default reasoning', *Artificial Intelligence*, **13**, 81–132, (1980).

[15] M. Truszczyński, 'Embedding default logic into modal nonmonotonic logics', in *Proceedings of the 1st International Workshop on Logic Programming and Non-monotonic Reasoning*, pp. 151–165, Washington, D.C., USA, (July 1991). The MIT Press.