Separating Disbeliefs from Beliefs in Autoepistemic Reasoning *

Tomi Janhunen

Helsinki University of Technology, Digital Systems Laboratory,
P.O.Box 1100, FIN-02015 HUT, Finland
E-mail: Tomi.Janhunen@hut.fi

Abstract. This paper investigates separated autoepistemic logic which is a generalization of Moore's autoepistemic logic with separate modalities for belief and disbelief. Along the separation of beliefs and disbeliefs, the relationship between autoepistemic logic and default logic becomes very intuitive. Straightforward ways of translating default theories into separated autoepistemic theories and back are presented. These translations are shown to preserve a variety of semantics of default theories such as those based on default extensions, weak extensions and stationary extensions. These classes of extensions are captured by their analogs in separated autoepistemic logic, and vice versa. A particular novelty of the approach is that a reasonable notion of separated stationary expansions can be established.

1 Introduction

The interconnection of Reiter's *default logic* [26] and Moore's *autoepistemic logic* [17] has been intensively studied since late eighties. Numerous *translations* from default theories into autoepistemic theories have been proposed (see e.g. [8, 13, 19, 28, 31]) to establish a correspondence between *extensions* of a default theory and *expansions* of the translation. So far, the results cover Reiter's *extensions* [26] Marek and Truszczyński's *weak extensions* [26] that have corresponding notions of expansions in autoepistemic logic. However, there are promising notions of extensions of default theories without a reasonable counterpart in autoepistemic logic. An example is the notion of *stationary extensions* [22] by Przymusinska and Przymusinski. This paper introduces *separated autoepistemic logic* – a reconstruction of autoepistemic logic that captures Reiter's extensions, weak extensions and stationary extensions of default theories as well as stable expansions [17] of conventional autoepistemic theories through natural translations. Our results provide new insight to default and autoepistemic logics and the principles behind stationary extensions in particular. A summary on the results of the paper follows.

In Section 2, we introduce autoepistemic theories with separate modalities for belief and disbelief and generalize the notions of stable [17] and iterative [13]

* The reported research has been financed by Academy of Finland and Helsinki Graduate School in Computer Science and Engineering.

expansions. The relationship to conventional autoepistemic logic [17] is investigated: it turns out that Moore's stable expansions are faithfully captured by their counterparts in the novel variant. This result is established using an axiom schema forcing the consistency of beliefs. In Section 3, we devise a finitary characterization of separated stable and iterative expansions. The full set approach used by Niemelä [18] and Gottlob [6] is generalized for separated stable and iterative expansions in a straightforward way.

In Section 4, we study how the separation of beliefs and disbeliefs affects the relationship to Reiter's default logic [26]. In light of earlier results by Konolige [8], Marek and Truszczyński [14], Niemelä [19] and others, it is non-trivial to translate default theories into autoepistemic theories such that Reiter's extensions are faithfully captured. By separate modalities for beliefs and disbeliefs, these difficulties are avoided. Reiter's extensions and weak extensions are preserved by a Konolige-style translation [8] of default theories into separated autoepistemic theories. Surprisingly, there is also a translation in the other direction. The translation of standard autoepistemic theories proposed by the author [7] generalizes for separated autoepistemic theories in a straightforward way. Via this translation, separated stable and iterative expansions are captured by their counterparts in default logic.

Section 5 is devoted to Przymusinska and Przymusinski's *stationary extensions* [22] of default theories. It is shown that stationary extensions are captured by their analogs in separated autoepistemic under the presented translations. The same applies to the well-founded semantics [29] of normal logic programs which is captured by stationary default extensions under a suitable translation of normal logic programs into default theories. A key property is that separated autoepistemic logic is capable of handling consistently belief states where something is believed and disbelieved simultaneously. McDermott's non-monotonic logics [15] – including Marek and Truszczyński's strong autoepistemic logic [13] – are shown to lack this property as well as a reasonable notion of stationary expansions.

In Section 6, we return to semantical questions. We address Moore's critique [17] on McDermott and Doyle's Non-monotonic modal logic I [16] by identifying sets of beliefs that are propositionally consistent, but inconsistent *at the level of beliefs*. This distinction is useful. It is possible to exclude belief inconsistent sets of beliefs if necessary. Moreover, even belief inconsistent sets of beliefs may provide useful information and can be used for drawing inferences. This is in contrast with Moore's autoepistemic logic where belief inconsistent sets of beliefs are inherently propositionally inconsistent. Finally, belief consistent separated stable expansions are characterized in terms of modal axioms in analogy with the results of Marek, Schwarz and Truszczyński [12].

In Section 7, a comparison is made with related work. We consider approaches that (i) capture the well-founded semantics of normal logic programs [3, 23, 24, 25, 31] (ii) are non-monotonic logics based on two modalities [10, 11] and (iii) provide yet alternative notion of extensions for default theories [2, 30]. The conclusions of the paper are presented in Section 8.

2 Separated Autoepistemic Theories

In the sequel, we introduce *separated autoepistemic theories* and define sets of beliefs associated with them. Let $\mathcal{L}_{\mathrm{BD}}$ denote an autoepistemic language based on a set of atomic propositions $\mathcal{P}$, propositional connectives and modalities $\mathbf{B}$ (belief) and $\mathbf{D}$ (disbelief). The corresponding conventional autoepistemic language $\mathcal{L}_{\mathrm{B}}$ (without the modality $\mathbf{D}$) is a sub-language of $\mathcal{L}_{\mathrm{BD}}$. The sentences of the forms $\mathbf{B}\phi$ and $\mathbf{D}\phi$ are treated as atomic propositions, so that the total set of atomic propositions in $\mathcal{L}_{\mathrm{BD}}$ is $\mathcal{P}_{\mathrm{BD}} = \mathcal{P} \cup \{\mathbf{B}\phi \mid \phi \in \mathcal{L}_{\mathrm{BD}}\} \cup \{\mathbf{D}\phi \mid \phi \in \mathcal{L}_{\mathrm{BD}}\}$. *Models* of separated autoepistemic theories $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ are simply subsets of $\mathcal{P}_{\mathrm{BD}}$. The set of *logical consequences* $\mathrm{Cn}(\Sigma)$ of $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ is the set $\{\phi \in \mathcal{L}_{\mathrm{BD}} \mid \Sigma \models \phi\}$.

Given a set of initial assumptions $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ we define the resulting sets of beliefs below. An ideal and rational agent believes exactly the logical consequences of Σ, its beliefs $\mathbf{B}\Delta = \{\mathbf{B}\phi \mid \phi \in \Delta\}$ obtained by *positive introspection* and its disbeliefs $\mathbf{D}\overline{\Delta} = \{\mathbf{D}\phi \mid \phi \in \mathcal{L}_{\mathrm{BD}} - \Delta\}$ obtained by *negative introspection.* This definition is analogous to that of *stable expansions* by Moore [17] where the equation $\Delta = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \neg\mathbf{B}\overline{\Delta})$ is used. The difference is that we use an extended language $\mathcal{L}_{\mathrm{BD}}$ instead of $\mathcal{L}_{\mathrm{B}}$ and express disbeliefs as atoms $\mathbf{D}\phi$ rather than negative literals $\neg\mathbf{B}\phi$. The same applies to autoepistemic theories $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ of our interest: referring to disbeliefs takes place with $\mathbf{D}$.

Definition 1. A set $\Delta \subseteq \mathcal{L}_{\mathrm{BD}}$ is a separated stable expansion of an autoepistemic theory $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ if and only if $\Delta = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$.

We extend the same idea of separation to the *strong autoepistemic logic* proposed by Marek and Truszczyński [13]. Their logic considers only *iterative expansions* of an autoepistemic theory Σ that are stable expansions of Σ satisfying also the equation $\Delta = \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \neg\mathbf{B}\overline{\Delta})$[2]. In Definition 2, we have rephrased the notion of iterative expansions for our framework with separate modalities. With iterative expansions, positive introspection is defined in terms of the necessitation rule. This excludes stable expansions with self-supporting beliefs. For example, $\Delta = \{\mathrm{p}, \ldots\}$ is a separated stable expansion but not an separated iterative expansion of $\Sigma = \{\mathbf{B}\mathrm{p} \rightarrow \mathrm{p}\}$. It is easy to show that separated iterative expansions of an autoepistemic theory $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ are separated stable expansions of Σ, but not vice versa.

Definition 2. A set $\Delta \subseteq \mathcal{L}_{\mathrm{BD}}$ is a separated iterative expansion of an autoepistemic theory $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ if and only if $\Delta = \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \mathbf{D}\overline{\Delta})$.

The separate modality $\mathbf{D}$ for disbeliefs makes autoepistemic logic more flexible. For example, we can express that our beliefs with respect to a sentence ϕ are *complete* using the sentence $\mathbf{B}\phi \vee \mathbf{D}\phi$. Definitions 1 and 2 ensure that separated stable/iterative expansions contain this sentence for every $\phi \in \mathcal{L}_{\mathrm{BD}}$.

[2] Here $\mathrm{Cn}_{\mathbf{B}}(\Gamma)$ is simply the closure of Γ with respect to propositional consequence and the *necessitation rule* $\phi \vdash \mathbf{B}\phi$. We let $\Gamma \vdash_{\mathbf{B}} \phi$ denote that ϕ is $\mathbf{B}$-provable from Γ in Gottlob's sense [6] so that $\Gamma \vdash_{\mathbf{B}} \phi$ if and only if $\phi \in \mathrm{Cn}_{\mathbf{B}}(\Gamma)$.

This not necessarily the case for separated stationary expansions to be introduced in Section 5. Another important sentence is $\neg(\mathbf{B}\phi \wedge \mathbf{D}\phi)$ expressing that our beliefs about ϕ are *consistent*, i.e. ϕ is not believed and disbelieved *simultaneously*. A shorthand $\mathbf{C}\phi$ is also used for this sentence. The corresponding sentences $\mathbf{B}\phi \vee \neg\mathbf{B}\phi$ and $\neg(\mathbf{B}\phi \wedge \neg\mathbf{B}\phi)$ are propositional tautologies in conventional autoepistemic logic. This indicates that new aspects of beliefs become distinguishable separate modalities for belief and disbelief.

Enforcing the belief consistency of each $\phi \in \mathcal{L}_{\mathrm{BD}}$ is necessary to capture stable expansions [17] of conventional autoepistemic theories $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$ in the new variant. Therefore the translation function of Theorem 3 extends a conventional autoepistemic theory Σ with $\mathbf{C}\phi$ for each $\phi \in \mathcal{L}_{\mathrm{BD}}$. We let $[\phi]$ denote a rewrite of $\phi \in \mathcal{L}_{\mathrm{BD}}$ such that $\mathbf{D}$ is recursively replaced by $\neg\mathbf{B}$, i.e. $[\mathbf{D}\phi] = \neg\mathbf{B}[\phi]$. The translation function Tr_{C} is shown to be faithful in Theorem 3 which relates stable expansions of $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$ and separated stable expansions of $\mathrm{Tr}_{\mathrm{C}}(\Sigma)$. In fact, a one-to-one correspondence results.

Theorem 3. *Let* $\mathrm{Tr}_{\mathrm{C}}(\Sigma)$ *be the translation* $\Sigma \cup \{\mathbf{C}\phi \mid \phi \in \mathcal{L}_{\mathrm{BD}}\}$ *of a conventional autoepistemic theory* $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$*. (i) If a set* $\Delta \subseteq \mathcal{L}_{\mathrm{B}}$ *is a stable expansion of* $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$*, then* $\Delta' = \{\phi' \in \mathcal{L}_{\mathrm{BD}} \mid [\phi'] \in \Delta\}$ *is a separated stable expansion of* $\mathrm{Tr}_{\mathrm{C}}(\Sigma)$*. (ii) If a set* $\Delta' \subseteq \mathcal{L}_{\mathrm{BD}}$ *is a separated stable expansion of* $\mathrm{Tr}_{\mathrm{C}}(\Sigma)$*, then* $\Delta = [\Delta']$ *is a stable expansion of* $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$*.*

Proof sketch. (i) For any $\Gamma \subseteq \mathcal{L}_{\mathrm{BD}}$ and $\phi \in \mathcal{L}_{\mathrm{BD}}$, $\Gamma \models \phi$ implies $[\Gamma] \models [\phi]$. Thus $\mathrm{Tr}_{\mathrm{C}}(\Sigma) \cup \mathbf{B}\Delta' \cup \mathbf{D}\overline{\Delta'} \models \phi$ implies $[\mathrm{Tr}_{\mathrm{C}}(\Sigma) \cup \mathbf{B}\Delta' \cup \mathbf{D}\overline{\Delta'}] \models [\phi]$. It follows that $\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta} \models [\phi]$, $[\phi] \in \Delta$ and $\phi \in \Delta'$. For the converse, let $\mathrm{Tr}_{\mathrm{C}}(\Sigma) \cup \mathbf{B}\Delta' \cup \mathbf{D}\overline{\Delta'} \not\models \phi$ hold, i.e. there is a model $\mathcal{M}$ of $\mathrm{Tr}_{\mathrm{C}}(\Sigma) \cup \mathbf{B}\Delta' \cup \mathbf{D}\overline{\Delta'}$ such that $\mathcal{M} \not\models \phi$. Then $\mathcal{M}' = \mathcal{M} \cap (\mathcal{P} \cup \mathbf{B}\mathcal{L}_{\mathrm{B}})$ is a model of $\Sigma \cup \mathbf{B}\Delta \cup \neg\mathbf{B}\overline{\Delta}$ so that $\mathcal{M}' \not\models [\phi]$. Thus $\Sigma \cup \mathbf{B}\Delta \cup \neg\mathbf{B}\overline{\Delta} \not\models [\phi]$, $[\phi] \notin \Delta$ and $\phi \notin \Delta'$. (ii) It can be shown by structural induction that $\phi' \leftrightarrow [\phi'] \in \mathrm{Cn}(\mathrm{Tr}_{\mathrm{C}}(\Sigma) \cup \mathbf{B}\Delta' \cup \mathbf{D}\overline{\Delta'})$ for all $\phi' \in \mathcal{L}_{\mathrm{BD}}$. It follows that $\phi' \in \Delta'$ iff $\mathrm{Tr}_{\mathrm{C}}(\Sigma) \cup \mathbf{B}\Delta' \cup \mathbf{D}\overline{\Delta'} \models \phi'$ iff $[\mathrm{Tr}_{\mathrm{C}}(\Sigma) \cup \mathbf{B}\Delta' \cup \mathbf{D}\overline{\Delta'}] \models [\phi']$ iff $\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta} \models [\phi']$. This implies for $\phi \in \mathcal{L}_{\mathrm{B}}$ that $\phi \in \Delta$ iff $\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta} \models \phi$. $\square$

3 Finitary Characterization

Separated stable and iterative expansions of an autoepistemic theory $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ are infinite sets of sentences. In order to implement reasoning based on this kind of expansions, a crucial step is establishing a finitary characterization for separated stable and iterative expansions. In this section, we give a finitary characterization of separated stable and iterative expansions based on *full sets* introduced by Niemelä [18] and Gottlob [6]. Decision procedures of separated autoepistemic logic can be based on algorithms computing full sets, but we leave this as a task of further research. Nevertheless, we need full sets to prove existence of expansions in Sections 4 and 5.

Generally speaking, the idea behind full sets is that we restrict our attention to sentences ϕ for which $\mathbf{B}\phi$ or $\mathbf{D}\phi$ appears in the premises Σ (and a possible query ϕ). More formally, we let $\mathrm{RBa}(\phi)$ denote the set of belief atoms that appear in ϕ recursively. For instance, $\mathrm{RBa}(\mathbf{D}(\mathbf{B}\mathrm{p} \vee \mathbf{D}\mathrm{q}))$ is $\{\mathbf{D}(\mathbf{B}\mathrm{p} \vee \mathbf{D}\mathrm{q}), \mathbf{B}\mathrm{p}, \mathbf{D}\mathrm{q}\}$. For a theory $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$, the set $\mathrm{RBa}(\Sigma)$ is $\bigcup_{\phi \in \Sigma} \mathrm{RBa}(\phi)$. In Definition 4 we adopt Niemelä's conditions [18] to provide a finitary characterization of separated stable expansions.

Definition 4. Let $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ be an autoepistemic theory. A set $\Lambda \subseteq \mathrm{RBa}(\Sigma)$ is Σ-full if and only if for every $\mathbf{B}\chi \in \mathrm{RBa}(\Sigma)$ and $\mathbf{D}\chi \in \mathrm{RBa}(\Sigma)$:

$$1.\ \mathbf{B}\chi \in \Lambda \ \text{ iff } \ \Sigma \cup \Lambda \models \chi \quad \text{and} \quad 2.\ \mathbf{D}\chi \in \Lambda \ \text{ iff } \ \Sigma \cup \Lambda \not\models \chi.$$

Given a separated stable expansion Δ of Σ the corresponding full set is easily obtained by a simple intersection as described in Theorem 5. However, reconstructing a separated stable expansion from a full set is a more tedious task. For this purpose, we define for each i a sublanguage $\mathcal{L}^i_{\mathrm{BD}} = \{\phi \mid \mathrm{BD}(\phi) < i\}$. Here the *belief depth* $\mathrm{BD}(\phi)$ of ϕ is the maximum nesting level of belief operators $\mathbf{B}$ and $\mathbf{D}$ in ϕ. For $\mathbf{D}(\mathbf{B}\mathrm{p} \vee \mathbf{D}\mathrm{q})$ above, the belief depth is 2. Note that $\mathcal{L}^0_{\mathrm{BD}} = \emptyset$ and $\mathcal{L}^1_{\mathrm{BD}} = \mathcal{L}$. A layer-wise reconstruction of separated stable expansions is provided in Definition 6 and Theorem 7.

Theorem 5. *Let $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$. If $\Delta \subseteq \mathcal{L}_{\mathrm{BD}}$ is a separated stable expansion of Σ, then $\Lambda = \mathrm{SFS}_\Sigma(\Delta) = (\mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta}) \cap \mathrm{RBa}(\Sigma)$ is Σ-full.*

Proof sketch. It can be shown for a set of belief atoms Γ and a sentence $\chi \in \mathcal{L}_{\mathrm{BD}}$ that $\Sigma \cup \Gamma \models \chi$ if and only if $\Sigma \cup \Lambda \models \chi$ where $\Lambda = \Gamma \cap \mathrm{RBa}(\Sigma \cup \{\chi\})$. Consider $\Gamma = \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta}$ and any $\mathbf{B}\chi \in \mathrm{RBa}(\Sigma)$, so that $\mathrm{RBa}(\Sigma \cup \{\chi\}) = \mathrm{RBa}(\Sigma)$ and $\Lambda = \Gamma \cap \mathrm{RBa}(\Sigma)$. Now $\mathbf{B}\chi \in \Lambda$ iff $\chi \in \Delta$ iff $\Sigma \cup \Gamma \models \chi$ iff $\Sigma \cup \Lambda \models \chi$. Similarly, $\mathbf{D}\chi \in \Lambda$ iff $\Sigma \cup \Lambda \not\models \chi$ holds for $\mathbf{D}\chi \in \mathrm{RBa}(\Sigma)$. □

Definition 6. Let $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ and $\Lambda \subseteq \mathrm{RBa}(\Sigma)$ a set of belief atoms. Define $\Delta^0 = \emptyset$ and $\Delta^{i+1} = \mathrm{Cn}(\Sigma \cup \Lambda \cup \mathbf{B}\Delta^i \cup \mathbf{D}\overline{\Delta^i}) \cap \mathcal{L}^i_{\mathrm{BD}}$ where $\overline{\Delta^i} = \mathcal{L}^i_{\mathrm{BD}} - \Delta^i$ for all $i \geq 0$. Finally, define $\mathrm{SSE}_\Sigma(\Lambda) = \bigcup_{i=0}^{\infty} \Delta^i$.

Theorem 7. *Let $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$. If $\Lambda \subseteq \mathrm{RBa}(\Sigma)$ is Σ-full, then $\Delta = \mathrm{SSE}_\Sigma(\Lambda)$ is a separated stable expansion of Σ such that $\Lambda = \mathrm{SFS}_\Sigma(\Delta)$.*

Proof sketch. Define Δ^i for $i \geq 0$ as in Definition 6. It can be shown by induction on i that (i) $(\mathbf{B}\Delta^i \cup \mathbf{D}\overline{\Delta^i}) \cap \mathrm{RBa}(\Sigma) \subseteq \Lambda$, (ii) $\Delta^i = \Delta^{i+1} \cap \mathcal{L}^i_{\mathrm{BD}}$ and (iii) $\overline{\Delta^i} = \overline{\Delta^{i+1}} \cap \mathcal{L}^i_{\mathrm{BD}}$. For the limit $\Delta = \bigcup_{i=0}^{\infty} \Delta^i$ we obtain $\Delta = \mathrm{Cn}(\Sigma \cup \Lambda \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$. Since also $\Lambda = (\mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta}) \cap \mathrm{RBa}(\Sigma)$, it follows that $\Delta = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$, i.e. Δ is a separated stable expansion of Σ. □

Gottlob [6] extends Niemelä's full set approach for iterative expansions of conventional autoepistemic theories. In Definition 8, we have reformulated Gottlob's condition to fit it in our framework. In the definition, $[\Lambda]^{\mathrm{D}}$ denotes the set $\{\mathbf{D}\chi \mid \mathbf{D}\chi \in \Lambda\}$. In Theorem 9, it is shown that the condition really characterizes separated iterative expansions.

Definition 8. Let $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$. A set $\Lambda \subseteq \mathrm{RBa}(\Sigma)$ is $\mathbf{B}$-ground if and only if for all $\mathbf{B}\chi \in \mathrm{RBa}(\Sigma)$: $\mathbf{B}\chi \in \Lambda$ iff $\Sigma \cup [\Lambda]^{\mathrm{D}} \vdash_{\mathbf{B}} \chi$.

Theorem 9. *Let $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$. (i) If $\Delta \subseteq \mathcal{L}_{\mathrm{BD}}$ is a separated iterative expansion of Σ, then $\Lambda = \mathrm{SFS}_{\Sigma}(\Delta)$ is Σ-full and $\mathbf{B}$-ground. (ii) If $\Lambda \subseteq \mathrm{RBa}(\Sigma)$ is Σ-full and $\mathbf{B}$-ground, then $\Delta = \mathrm{SSE}_{\Sigma}(\Lambda)$ is a separated iterative expansion of Σ such that $\Lambda = \mathrm{SFS}_{\Sigma}(\Delta)$.*

Proof sketch. (i) It can be shown for a set of belief atoms Γ and a sentence $\chi \in \mathcal{L}_{\mathrm{BD}}$ that $\Sigma \cup \Gamma \vdash_{\mathbf{B}} \chi$ if and only if $\Sigma \cup \Lambda \vdash_{\mathbf{B}} \chi$ where $\Lambda = \Gamma \cap \mathrm{RBa}(\Sigma \cup \{\chi\})$. Consider $\Gamma = \mathbf{D}\overline{\Delta}$ and any $\mathbf{B}\chi \in \mathrm{RBa}(\Sigma)$, so that $\mathrm{RBa}(\Sigma \cup \{\chi\}) = \mathrm{RBa}(\Sigma)$ and $[\Lambda]^{\mathrm{D}} = \Gamma \cap \mathrm{RBa}(\Sigma)$. Now $\mathbf{B}\chi \in \Lambda$ iff $\chi \in \Delta$ iff $\Sigma \cup \mathbf{D}\overline{\Delta} \vdash_{\mathbf{B}} \chi$ iff $\Sigma \cup [\Lambda]^{\mathrm{D}} \vdash_{\mathbf{B}} \chi$. (ii) It can be shown by induction on the lengths of $\mathbf{B}$-proofs for every separated stable expansion Δ that $\mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \mathbf{D}\overline{\Delta}) \subseteq \Delta$. The inclusion in the other direction follows by augmenting the proof of Theorem 7 by showing that $\Delta^i \subseteq \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \mathbf{D}\overline{\Delta})$. For the limit $\Delta = \bigcup_{i=0}^{\infty} \Delta^i$, we obtain $\Delta \subseteq \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \mathbf{D}\overline{\Delta})$. Thus Δ is a separated iterative expansion of Σ. □

4 Relationship to Default Logic

We have designed separated autoepistemic logic so that a straightforward relationship with Reiter's *default logic* [26] is established. Conventional autoepistemic theories and default theories are already closely related, but a variety of *translations* and *groundedness conditions* have been proposed to capture Reiter's extensions with stable expansions. In this section, we capture Reiter's *extensions* [26] and Marek and Truszczyński's *weak extensions* [14] of a default theory with separated iterative and separated stable expansions of a translation, and vice versa. Most importantly, these correspondences are established using very intuitive translations involving the modality $\mathbf{D}$ in a major role and with a groundedness condition corresponding to separated iterative expansions only. This makes the connection between separated autoepistemic logic and default logic considerably simpler compared to the connection between standard autoepistemic logic and default logic.

Let us now introduce Reiter's default theories [26]. Such a theory is a pair $\langle D, W \rangle$ where D is a set of *default rules* of the form $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma}$ [3] and W is a propositional theory in a propositional language $\mathcal{L}$. For a set of default rules D, we let $\mathrm{P}(D)$ and $\mathrm{J}(D)$ denote the sets of prerequisites α and justifications β that appear in the rules of D. A default rule can be applied, if the prerequisite α has been derived and the justifications $\beta_1, \ldots, \beta_n$ can be consistently assumed. The sets of conclusions associated with a default theory are typically called *extensions*. For the moment, we shall consider only Reiter's *extensions* [26] and *weak extensions* [14] as defined by Marek and Truszczyński [14]. They use two

[3] A default rule contains a prerequisite α, justifications $\beta_1, \ldots, \beta_n$ and a consequent γ. A default rule is prerequisite-free if $\alpha = \top$. In this case, α is often omitted.

	Translation	Reference
Tr_{K}	$\mathbf{B}\alpha \wedge \neg\mathbf{B}\neg\beta_1 \wedge ... \wedge \neg\mathbf{B}\neg\beta_n \rightarrow \gamma$	Konolige [8]
$\mathrm{Tr}_{\mathrm{MT}}$	$\mathbf{B}\alpha \wedge \neg\mathbf{BB}\neg\beta_1 \wedge ... \wedge \neg\mathbf{BB}\neg\beta_n \rightarrow \gamma$	Marek and Truszczyński [13]
Tr_{T}	$\mathbf{B}\alpha \wedge \mathbf{B}\neg\mathbf{B}\neg\beta_1 \wedge ... \wedge \mathbf{B}\neg\mathbf{B}\neg\beta_n \rightarrow \gamma$	Truszczyński [28]

Table 1. Translations of $\frac{\alpha:\beta_1,...,\beta_n}{\gamma}$ into an autoepistemic sentence

reductions of a set of default rules D with respect to a theory E. The reduct D_E is the set of *inference rules*

$$\{\frac{\alpha}{\gamma} \mid \frac{\alpha:\beta_1,...,\beta_n}{\gamma} \in D \text{ and } \neg\beta_1 \notin E, ..., \neg\beta_n \notin E\}$$ [4]

and the reduct D^E is the set of *sentences*

$$\{\gamma \mid \frac{\alpha:\beta_1,...,\beta_n}{\gamma} \in D, \alpha \in E \text{ and } \neg\beta_1 \notin E, ..., \neg\beta_n \notin E\}.$$

For a set R of inference rules, the theory $\mathrm{Cn}^R(W)$ is the closure of a propositional theory W under propositional consequence and the rules of R.

Definition 10 (Marek and Truszczyński [14]). Let $\langle D, W\rangle$ be a default theory. A set $E \subseteq \mathcal{L}$ is an extension (resp. a weak extension) of $\langle D, W\rangle$ iff $E = \mathrm{Cn}^{D_E}(W)$ (resp. $E = \mathrm{Cn}(W \cup D^E)$).

4.1 Embedding Default Logic in Separated Autoepistemic Logic

Since the introduction of autoepistemic logic, its relationship to default logic has received much attention. It turned out soon that embedding default logic into autoepistemic logic is not as trivial as one would expect. A number of translations were proposed. In Table 1 we have listed three different ways of translating a default rule into an autoepistemic sentence. There are also other possibilities [4, p. 275 in Konolige's article], but we have listed only the ones relevant for our forthcoming discussion. We let Tr_{K}, $\mathrm{Tr}_{\mathrm{MT}}$ and Tr_{T} denote the corresponding translation functions for defaults, respectively. Given a default theory $\langle D, W\rangle$ and a translation function Tr for defaults of D, the translation $\mathrm{Tr}(\langle D, W\rangle)$ of the whole theory $\langle D, W\rangle$ is simply $W \cup \{\mathrm{Tr}(d) \mid d \in D\}$.

The goal is to establish a bijective relationship between extensions E of $\langle D, W\rangle$ and a stable expansions Δ of the translation $\mathrm{Tr}(\langle D, W\rangle)$. Unfortunately, a translation $\mathrm{Tr}(\langle D, W\rangle)$ tends to have more stable expansions than $\langle D, W\rangle$ has extensions. This problem is typically approached by considering only stable expansions of $\mathrm{Tr}(\langle D, W\rangle)$ that satisfy an additional *groundedness condition.* Konolige's translation [8] requires a groundedness condition that is based on a

[4] Note that β is consistent with a propositionally closed theory E iff $\neg\beta \notin E$.

reduction of the premises. This makes the involved notion of groundedness syntax dependent: propositionally equivalent premises may have different strongly grounded stable expansions. Marek and Truszczyński [13] restrict to *iterative expansions* of $\mathrm{Tr}_{\mathrm{MT}}(\langle D, W\rangle)$. The notion of groundedness corresponding to iterative expansions is attractive, since it is syntax independent. Truszczyński [28] investigates a range of non-monotonic modal logics N − SF4 to establish correspondence for Tr_{T}. Expansions in the weakest non-monotonic modal logic in this range are exactly iterative expansions.

We agree with Marek and Truszczyński that iterative expansions form a natural counterpart of default extensions on the autoepistemic side. This is mainly because reasoning with inference rules of the form $\frac{\alpha}{\gamma}$ is naturally captured by autoepistemic implications of the form $\mathbf{B}\alpha \rightarrow \gamma$ when the necessitation rule $\phi \vdash \mathbf{B}\phi$ is present. For example we can infer α_n from $\{\alpha_0\} \cup \{\mathbf{B}\alpha_0 \rightarrow \alpha_1, ..., \mathbf{B}\alpha_{n-1} \rightarrow \alpha_n\}$ through n applications of the necessitation rule. This captures inferring α_n from α_0 by the inference rules $\frac{\alpha_0}{\alpha_1}, ..., \frac{\alpha_{n-1}}{\alpha_n}$. More precisely, the following can be established:

Proposition 11. *Let W be a propositional theory and R a set of inference rules. Then* $\mathrm{Cn}^R(W) = \mathrm{Cn}_{\mathbf{B}}(W \cup \{\mathbf{B}\alpha \rightarrow \gamma \mid \frac{\alpha}{\gamma} \in R\}) \cap \mathcal{L}$.

The translation functions Tr_{K}, $\mathrm{Tr}_{\mathrm{MT}}$ and Tr_{T} behave differently what comes to capturing extensions of $\langle D, W\rangle$ *with* iterative expansions of the translation but *without* further restrictions. Example 1 differentiates Tr_{K} from $\mathrm{Tr}_{\mathrm{MT}}$ and Tr_{T} which avoid the unwanted iterative expansion. Example 2 shows that the translation Tr_{T} is faithful if only *propositionally consistent* iterative expansions are considered. Niemelä [19] uses Tr_{T}, but rules out the iterative expansion of Example 2 using a notion of **B**-*hierarchic expansions*.

Example 1 (Marek and Truszczyński). Let $D = \{\frac{\neg\neg \mathrm{p}:}{\mathrm{p}}, \frac{:\neg \mathrm{p}}{\mathrm{p}}\}$ and $W = \emptyset$. The default theory $\langle D, W\rangle$ does not have extensions. On the other hand, the translation $\mathrm{Tr}_{\mathrm{K}}(\langle D, W\rangle) = \{\mathbf{B}\neg\neg\mathrm{p} \rightarrow \mathrm{p}, \mathbf{B}\top \wedge \neg\mathbf{B}\neg\neg\mathrm{p} \rightarrow \mathrm{p}\}$ has an iterative expansion containing p, since p is **B**-provable from $\mathrm{Tr}_{\mathrm{K}}(\langle D, W\rangle)$ directly.

Example 2 (Niemelä [19]). Let $D = \{\frac{\neg \mathrm{b}:}{\neg \mathrm{a}}, \frac{:\mathrm{b}}{\neg \mathrm{b}}\}$ and $W = \{\mathrm{a}\}$. The default theory $\langle D, W\rangle$ has no extensions at all, whereas the translation $\mathrm{Tr}_{\mathrm{T}}(\langle D, W\rangle)$ i.e. the theory $\{\mathrm{a}, \mathbf{B}\neg\mathrm{b} \rightarrow \neg\mathrm{a}, \mathbf{B}\top \wedge \mathbf{B}\neg\mathbf{B}\neg\mathrm{b} \rightarrow \neg\mathrm{b}\}$ has a propositionally inconsistent iterative expansion, since $\perp$ is **B**-provable from $\mathrm{Tr}_{\mathrm{T}}(\langle D, W\rangle)$.

It is clear by these examples that the way in which the consistency conditions $\beta_1, ..., \beta_n$ of a default rule $\frac{\alpha:\beta_1,...,\beta_n}{\gamma}$ are translated is of crucial importance. In our opinion, Konolige's translation is the most intuitive one, because the consistency of a justification is naturally expressed as $\neg\mathbf{B}\neg\beta$ on the autoepistemic side. The problems demonstrated in Examples 1 and 2 can be avoided if we express $\neg\mathbf{B}\neg\beta$ by $\mathbf{D}\neg\beta$. This leads to the natural translation function presented in Definition 12. In Theorem 13, a two-level correspondence between extensions and separated expansions of the translation is established. First of all, the separated stable expansions of $\mathrm{Tr}^{\mathrm{D}}_{\mathrm{K}}(\langle D, W\rangle)$ correspond to the weak extensions of

$\langle D, W\rangle$. This result was established for Tr_{K} and standard autoepistemic logic by Marek and Truszczyński [14]. Moreover, the Reiter's extensions of $\langle D, W\rangle$ are captured if the separated iterative expansions of $\mathrm{Tr}_{\mathrm{K}}^{\mathrm{D}}(\langle D, W\rangle)$ are considered. Let us yet emphasize that Theorem 13 is not restricted to propositionally consistent extensions/expansions.

Definition 12. A default theory $\langle D, W\rangle$ is translated into an autoepistemic theory $\mathrm{Tr}_{\mathrm{K}}^{\mathrm{D}}(\langle D, W\rangle)=W \cup \{\mathbf{B}\alpha \wedge \mathbf{D}\neg\beta_1 \wedge \ldots \wedge \mathbf{D}\neg\beta_n \rightarrow \gamma \mid \frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma} \in D\}$.

Theorem 13. *A set $E \subseteq \mathcal{L}$ is an extension (resp. a weak extension) of a default theory $\langle D, W\rangle$ iff there is a separated iterative expansion (resp. separated stable expansion) Δ of $\Sigma = \mathrm{Tr}_{\mathrm{K}}^{\mathrm{D}}(\langle D, W\rangle)$ such that $E = \Delta \cap \mathcal{L}$.*

Proof sketch. ($\Rightarrow$) For an extension (resp. a weak extension) E of $\langle D, W\rangle$, let Λ be the set of belief atoms containing for all $\alpha \in \mathrm{P}(D)$ and $\beta \in \mathrm{J}(D)$ (i) the atom $\mathbf{B}\alpha$ whenever $\alpha \in E$ and (ii) the atom $\mathbf{D}\neg\beta$ whenever $\neg\beta \notin E$. Now Λ is Σ-full. Moreover, if E is an extension of $\langle D, W\rangle$, the set Λ is also $\mathbf{B}$-ground. Consequently, there is a separated iterative (resp. separated stable) expansion of Σ such that $E = \Delta \cap \mathcal{L}$. ($\Leftarrow$) If Δ is a separated iterative expansion of Σ, the set $E = \Delta \cap \mathcal{L} = \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \mathbf{D}\overline{\Delta}) \cap \mathcal{L} = \mathrm{Cn}_{\mathbf{B}}(W \cup \{\mathbf{B}\alpha \rightarrow \gamma \mid \frac{\alpha}{\gamma} \in D_E\}) \cap \mathcal{L} = \mathrm{Cn}^{D_E}(W)$ is an extension of $\langle D, W\rangle$ by Proposition 11. If Δ is a separated stable expansion of Σ, the set $E = \Delta \cap \mathcal{L} = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta}) \cap \mathcal{L} = \mathrm{Cn}(W \cup D^E)$ is a weak extension of $\langle D, W\rangle$. $\square$

Let us now reconsider Examples 1 and 2. In Example 3, it is essential that the translation Σ *does not entail* $\mathbf{B}\neg\neg\mathrm{p} \vee \mathbf{D}\neg\neg\mathrm{p}$. To the contrary, the sentence $\mathbf{B}\neg\neg\mathrm{p} \vee \neg\mathbf{B}\neg\neg\mathrm{p}$ is a tautology in conventional autoepistemic logic. Thus p is $\mathbf{B}$-provable from the standard Konolige-style translation $\{\mathbf{B}\neg\neg\mathrm{p} \rightarrow \mathrm{p}, \mathbf{B}\top \wedge \neg\mathbf{B}\neg\neg\mathrm{p} \rightarrow \mathrm{p}\}$ and the only stable expansion is also an iterative one.

Example 3. Consider the theory $\Sigma = \{\mathbf{B}\neg\neg\mathrm{p} \rightarrow \mathrm{p}, \mathbf{B}\top \wedge \mathbf{D}\neg\neg\mathrm{p} \rightarrow \mathrm{p}\}$ which is the translation $\mathrm{Tr}_{\mathrm{K}}^{\mathrm{D}}(\langle\{\frac{\neg\neg\mathrm{p}:}{\mathrm{p}}, \frac{:\neg\mathrm{p}}{\mathrm{p}}\}, \emptyset\rangle)$. The theory Σ has a separated stable expansion Δ containing p, but Δ is not a separated iterative expansion, because p is not $\mathbf{B}$-provable from $\Sigma \cup \mathbf{D}\overline{\Delta}$. Here note especially that $\neg\neg\mathrm{p} \notin \overline{\Delta}$.

In Example 4, the $\mathbf{B}$-proof of $\perp$ from the translation is no longer possible, because $\mathbf{D}\neg\mathrm{b}$ has replaced $\mathbf{B}\neg\mathbf{B}\neg\mathrm{b}$ and $\mathbf{D}\neg\mathrm{b}$ cannot be concluded. In both examples, there is a *weak* extension $E = \Delta \cap \mathcal{L}$ of the default theory in question as implied by Theorem 13.

Example 4. Consider the theory $\Sigma = \{\mathrm{a}, \mathbf{B}\neg\mathrm{b} \rightarrow \neg\mathrm{a}, \mathbf{B}\top \wedge \mathbf{D}\neg\mathrm{b} \rightarrow \neg\mathrm{b}\}$ which is the translation $\mathrm{Tr}_{\mathrm{K}}^{\mathrm{D}}(\langle\{\frac{\neg\mathrm{b}:}{\neg\mathrm{a}}, \frac{:\mathrm{b}}{\neg\mathrm{b}}\}, \{\mathrm{a}\}\rangle)$. The theory Σ has a propositionally inconsistent separated stable expansion $\Delta = \mathcal{L}_{\mathrm{BD}}$, but Δ is not separated iterative, since $\perp$ is not $\mathbf{B}$-provable from $\Sigma \cup \mathbf{D}\overline{\Delta} = \Sigma$.

4.2 Embedding Separated Autoepistemic Logic in Default Logic

The author has recently proposed a translation of standard autoepistemic theories into default theories [7]. The translation is based on an intuitive principle of representing autoepistemic introspection in terms of default rules. To apply the same principle in the context of separated autoepistemic theories, we augment an autoepistemic theory with default rules of the forms $\frac{\phi:}{\mathbf{B}\phi}$ and $\frac{:\neg\phi}{\mathbf{D}\phi}$ to capture positive and negative introspection of $\phi \in \mathcal{L}_{\mathrm{BD}}$. Definition 14 gives the resulting translation $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$. The language of the translation is a propositional language $\mathcal{L}$ based on the set of atomic propositions $\mathcal{P}_{\mathrm{BD}} = \mathcal{P} \cup \{\mathbf{B}\phi \mid \phi \in \mathcal{L}_{\mathrm{BD}}\} \cup \{\mathbf{D}\phi \mid \phi \in \mathcal{L}_{\mathrm{BD}}\}$[5]. From this point of view, the languages $\mathcal{L}_{\mathrm{BD}}$ and $\mathcal{L}$ coincide. It is established in Theorem 15 that $\mathrm{Tr}_{\mathrm{DL}}$ preserves separated stable expansions and separated iterative expansions.

Definition 14. A autoepistemic theory $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ is translated into a default theory $\mathrm{Tr}_{\mathrm{DL}}(\Sigma) = \langle \{\frac{\phi:}{\mathbf{B}\phi} \mid \phi \in \mathcal{L}_{\mathrm{BD}}\} \cup \{\frac{:\neg\phi}{\mathbf{D}\phi} \mid \phi \in \mathcal{L}_{\mathrm{BD}}\}, \Sigma\rangle$.

Theorem 15. *A set $\Delta \subseteq \mathcal{L}_{\mathrm{BD}}$ is separated iterative expansion (resp. separated stable expansion) of $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ iff $\Delta \subseteq \mathcal{L}$ is an extension (resp. a weak extension) of the translation $\mathrm{Tr}_{\mathrm{DL}}(\Sigma)$.*

Proof sketch. Let D be the set of default rules of Definition 14 and R the set of inference rules $\{\frac{\psi}{\mathbf{B}\psi} \mid \psi \in \mathcal{L}_{\mathrm{BD}}\}$. Since Δ is closed with respect to propositional consequence, the reduct $D_\Delta = R \cup \{\frac{\top}{\mathbf{D}\psi} \mid \psi \in \overline{\Delta}\}$ and the reduct $D^\Delta = \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta}$. It follows that $\mathrm{Cn}^{D_\Delta}(\Sigma) = \mathrm{Cn}^{R}(\Sigma \cup \mathbf{D}\overline{\Delta}) = \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \mathbf{D}\overline{\Delta})$ and $\mathrm{Cn}(\Sigma \cup D^\Delta) = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$. Under these equalities, the definitions of extensions (resp. weak extensions) and separated iterative expansions (resp. separated stable expansions) coincide. □

Few remarks on the presented translations follow. (i) The translation $\mathrm{Tr}_{\mathrm{DL}}$ is not computationally feasible, since an infinite set of default rules is involved. However, this set of default rules can be significantly reduced if we aim at capturing full sets rather than complete expansions. In this setting, the required translation is $\mathrm{Tr}^{\mathrm{f}}_{\mathrm{DL}}(\Sigma) = \langle \{\frac{\chi:}{\mathbf{B}\chi} \mid \mathbf{B}\chi \in \mathrm{RBa}(\Sigma)\} \cup \{\frac{:\neg\chi}{\mathbf{D}\chi} \mid \mathbf{D}\chi \in \mathrm{RBa}(\Sigma)\}, \Sigma\rangle$. which finite for finite autoepistemic theories $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$. Using this technique, (separated) autoepistemic reasoning can be implemented [7] using an existing default reasoning system. (ii) The presented translation functions are *modular*, i.e. local changes in a theory cause only local changes in the translation. We take this as a strong indication that separated autoepistemic logic and default logic are of equal expressive power under the notions of expansions and extensions considered in this paper. (iii) The obvious compositions of the translations $\mathrm{Tr}^{\mathrm{D}}_{\mathrm{K}}$, $\mathrm{Tr}_{\mathrm{DL}}$ and $\mathrm{Tr}^{\mathrm{f}}_{\mathrm{DL}}$ lead to self-embeddings of separated autoepistemic logic and default logic. Especially, every default theory $\langle D, W\rangle$ can be translated to a default theory $\mathrm{Tr}^{\mathrm{f}}_{\mathrm{DL}}(\mathrm{Tr}^{\mathrm{D}}_{\mathrm{K}}(\langle D, W\rangle))$ which contains only default rules of the simple forms $\frac{\chi:}{\mathbf{B}\chi}$ and $\frac{:\neg\chi}{\mathbf{D}\chi}$ introduced by $\mathrm{Tr}^{\mathrm{f}}_{\mathrm{DL}}$.

[5] Alternatively, belief atoms $\mathbf{B}\phi$ and $\mathbf{D}\phi$ can be rewritten as new propositional atoms b_ϕ and d_ϕ, respectively.

5 Capturing Stationary Extensions of Default Theories

In Example 1 we encountered a default theory without extensions. In a way, the fix-point condition of Definition 10 is too strict to be satisfied in some cases. One possibility is to loosen the condition somehow. This is what Przymusinska and Przymusinski [22] do when they propose *stationary extensions* for default theories. In Definition 16, we have defined stationary extensions using Marek and Truszczyński's reduction of default rules. This is already a third kind of extensions considered in this paper. One of the attractive properties of stationary extensions is that every default theory has the least stationary extension which can be iteratively constructed as the least fix-point of a monotonic operator [22]. In this section, we introduce a natural counterpart of stationary extensions for separated autoepistemic logic, namely *separated stationary expansions.* It is then shown that such expansions are in a one-to-one correspondence with stationary extensions under the translation functions that were introduced in Section 4. This is a further indication of the close relationship of separated autoepistemic logic and default logic.

Definition 16. A pair $\langle E_1, E_2\rangle \in \mathbf{2}^{\mathcal{L}} \times \mathbf{2}^{\mathcal{L}}$ is a stationary extension of a default theory $\langle D, W\rangle$ iff $E_1 = \mathrm{Cn}^{D_{E_2}}(W)$, $E_2 = \mathrm{Cn}^{D_{E_1}}(W)$ and $E_1 \subseteq E_2$.

In Definition 17, we have phrased an analogous definition for separated autoepistemic theories. A *separated stationary expansion* $\langle \Delta_1, \Delta_2\rangle$ partitions $\mathcal{L}_{\mathrm{BD}}$ into three disjoint sets of sentences. The sentences of Δ_1 are considered to be beliefs while the sentences of $\overline{\Delta_2}$ are considered to be disbeliefs. The sentences in $\Delta_2 - \Delta_1$ remain *undefined.* In this way, separated stationary expansions have a three-valued interpretation at the level of beliefs despite expansions have been defined in terms of two-valued propositional logic.

Definition 17. A pair $\langle \Delta_1, \Delta_2\rangle \in \mathbf{2}^{\mathcal{L}_{\mathrm{BD}}} \times \mathbf{2}^{\mathcal{L}_{\mathrm{BD}}}$ is a separated stationary expansion of an autoepistemic theory $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ iff $\Delta_1 = \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \mathbf{D}\overline{\Delta_2})$, $\Delta_2 = \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \mathbf{D}\overline{\Delta_1})$ and $\Delta_1 \subseteq \Delta_2$.

Example 5 provides the least stationary expansion for a separated autoepistemic theory. In this example, Δ_2 contains both $\mathbf{B}\mathrm{p}$ and $\mathbf{D}\mathrm{p}$, since assuming $\mathbf{D}\mathrm{p}$ allows deriving p and $\mathbf{B}\mathrm{p}$ using the implication $\mathbf{D}\mathrm{p} \to \mathrm{p}$ and the necessitation rule. Despite of this inconsistency at the level of beliefs Δ_2 remains propositionally consistent. We take this as a crucial property which is achieved by separating disbeliefs from beliefs in terms of the second modality $\mathbf{D}$ and selecting a suitable class of (propositional) models.

Example 5. Consider $\Sigma = \{\mathbf{D}\mathrm{p} \to \mathrm{p}, \mathbf{D}\mathrm{q} \to \mathrm{r}\}$ having no separated stable expansions. This theory has a separated stationary expansion $\langle \Delta_1, \Delta_2\rangle$ where $\Delta_1 = \{\mathbf{D}\mathrm{q}, \mathrm{r}, \mathbf{B}\mathrm{r}, \ldots\}$ and $\Delta_2 = \Delta_1 \cup \{\mathbf{D}\mathrm{p}, \mathrm{p}, \mathbf{B}\mathrm{p}, \ldots\}$. The sentence p is undefined while r is believed and q is disbelieved.

The construction of Definition 17 is not successful for McDermott-style non-monotonic modal logics [15] which are based on monotonic modal logics S with the modality $\mathbf{B}$ subject to the necessitation rule. In these logics, expansions are solutions to $\Delta = \mathrm{Cn}_S(\Sigma \cup \neg\mathbf{B}\overline{\Delta})$. Defining stationary expansions in the spirit of Definition 17 leads to conditions $\Delta_1 = \mathrm{Cn}_S(\Sigma \cup \neg\mathbf{B}\overline{\Delta_2})$, $\Delta_2 = \mathrm{Cn}_S(\Sigma \cup \neg\mathbf{B}\overline{\Delta_1})$ and $\Delta_1 \subseteq \Delta_2$. Three cases arise. (i) If $\mathrm{Cn}_S(\Sigma)$ is inconsistent, then $\langle \mathcal{L}_\mathrm{B}, \mathcal{L}_\mathrm{B} \rangle$ is the least stationary expansion. Then assume that $\Delta = \mathrm{Cn}_S(\Sigma)$ is consistent. (ii) In practical cases, we can pick a sentence $\phi \in \overline{\Delta}$ such that $\neg\mathbf{B}\phi \in \overline{\Delta}$. Then both $\mathbf{B}\neg\mathbf{B}\phi$ and $\neg\mathbf{B}\neg\mathbf{B}\phi$ belong to $\mathrm{Cn}_S(\Sigma \cup \neg\mathbf{B}\overline{\Delta})$ which is therefore inconsistent and $\langle \Delta, \mathcal{L}_\mathrm{B} \rangle$ is the least stationary expansion. (iii) Otherwise, $\Delta = \mathrm{Cn}_S(\Sigma) = \mathrm{Cn}_S(\Sigma \cup \neg\mathbf{B}\overline{\Delta})$ and $\langle \Delta, \Delta \rangle$ is the least stationary expansion of Σ. By the cases (i – iii), the stationary semantics collapses to $\mathrm{Cn}_S(\Sigma)$ for non-monotonic modal logics S and stationary extensions of default theories cannot be captured under any reasonable translation. This is also the case with Marek and Truszczyński's strong autoepistemic logic [13] based on the iterative expansions of the premises.

Theorems 13 and 15 generalize for the translation functions $\mathrm{Tr}_\mathrm{K}^\mathrm{D}$ and Tr_DL and the notions of extensions and expansions introduced in Definitions 16 and 17. Przymusinska and Przymusinski [22] have shown that stationary extensions capture the well-founded semantics of normal logic programs [29]. On the basis of Theorem 18, the same holds for separated stationary expansions.

Theorem 18. *A pair* $\langle E_1, E_2 \rangle \in \mathbf{2}^{\mathcal{L}} \times \mathbf{2}^{\mathcal{L}}$ *is a stationary extension of a default theory* $\langle D, W \rangle$ *iff there is a separated stationary expansion* $\langle \Delta_1, \Delta_2 \rangle \in \mathbf{2}^{\mathcal{L}_\mathrm{BD}} \times \mathbf{2}^{\mathcal{L}_\mathrm{BD}}$ *of* $\mathrm{Tr}_\mathrm{K}^\mathrm{D}(\langle D, W \rangle)$ *such that* $E_i = \Delta_i \cap \mathcal{L}$ *for* $i \in \{1, 2\}$.

Theorem 19. *A pair* $\langle \Delta_1, \Delta_2 \rangle \in \mathbf{2}^{\mathcal{L}_\mathrm{BD}} \times \mathbf{2}^{\mathcal{L}_\mathrm{BD}}$ *is separated stationary expansion of* $\Sigma \subseteq \mathcal{L}_\mathrm{BD}$ *iff* $\langle \Delta_1, \Delta_2 \rangle$ *is a stationary extension of* $\mathrm{Tr}_\mathrm{DL}(\Sigma)$.

Separated stationary expansions are appealing *approximations* of separated iterative expansions. In analogy to the least stationary extension [22], the *least separated stationary expansion* of $\Sigma \subseteq \mathcal{L}_\mathrm{BD}$ can be iteratively constructed as the least fixed point $\mathrm{lfp}(\mathrm{SIE}^2_\Sigma, \emptyset)$ of a monotonic operator SIE^2_Σ which corresponds to applying twice the operator SIE_Σ defined by $\mathrm{SIE}_\Sigma(\Delta) = \mathrm{Cn}_\mathbf{B}(\Sigma \cup \mathbf{D}\overline{\Delta})$ for $\Delta \subseteq \mathcal{L}_\mathrm{BD}$. In addition, $\mathrm{lfp}(\mathrm{SIE}^2_\Sigma, \emptyset)$ is contained in all iterative expansions of Σ. This suggests that $\mathrm{lfp}(\mathrm{SIE}^2_\Sigma, \emptyset)$ should be used as the first approximation when the separated iterative expansions of Σ are determined. The finitary characterization separated iterative expansions presented in Section 3 extends for separated stationary expansions. This allows representing separated stationary extensions of a finite autoepistemic theory Σ in a finite space. Moreover, there is a finitary counterpart of the operator SIE^2_Σ that constructs a pair of full sets characterizing the least separated stationary expansion. This leads to a basic algorithm for deciding whether a sentence belongs to the least separated stationary expansion. It is to be expected that the resulting algorithm is tractable in restricted cases (e.g. premises containing Horn-clauses only) like in [22].

6 Characterizing Separated Stable Expansions

Moore proposed autoepistemic logic as a reconstruction of Non-monotonic modal logic I (NML I) by McDermott and Doyle [16]. In this section, we study how Moore's criticism should be taken into account with separated autoepistemic theories. This leads us to identify separated stable expansions that are consistent at the level of beliefs. Starting from this notion of consistency, we devise a characterization for separated stable expansions using modal axiom schemas involving the modalities **B** and **D**.

According to Moore, expansions in NML I are semantically unsatisfactory. Much of this criticism can be demonstrated by a theory $\{\mathrm{p}, \mathbf{M}\neg\mathrm{p}\}$ in NML I where the modality **M** expresses propositional consistency (with respect to the current set of beliefs). The point is that the given theory has a consistent expansion in NML I despite having p true and $\neg p$ consistent simultaneously is contradictory. Moore singles out the lack of positive introspection as a reason for this semantical problem. This is why he defines stable expansions using the equation $\Delta = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \neg\mathbf{B}\overline{\Delta})$. In fact, the corresponding conventional autoepistemic theory $\{\mathrm{p}, \neg\mathbf{B}\mathrm{p}\}$ has only a propositionally inconsistent stable expansion. The corresponding separated autoepistemic theory follows.

Example 6. The theory $\Sigma = \{\mathrm{p}, \mathbf{D}\mathrm{p}\}$ has a propositionally consistent separated stable expansion $\Delta = \{\mathrm{p}, \mathbf{B}\mathrm{p}, \mathbf{D}\mathrm{p}, \mathbf{B}\mathrm{p} \wedge \mathbf{D}\mathrm{p} \ldots\}$.

As it is apparent from Definition 1, positive introspection is performed with separated autoepistemic theories. In spite of this, the separated stable expansion Δ of Example 6 remains propositionally consistent. This is clearly due to separated modalities for representing the results of positive and negative introspection. On the other hand, the expansion Δ is inconsistent *at the level of beliefs*, i.e. is contains a sentence $\mathbf{B}\phi \wedge \mathbf{D}\phi$ for some $\phi \in \mathcal{L}_{\mathrm{BD}}$. In Definition 20, we distinguish separated stable expansions that are belief consistent with respect to every $\phi \in \mathcal{L}_{\mathrm{BD}}$ *simultaneously.* A belief consistent separated stable expansion is always propositionally consistent. Example 6 is a counter-example for the converse result.

Definition 20. A separated stable expansion Δ of $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ is belief consistent if and only if $\Delta \cup \{\mathbf{C}\phi \mid \phi \in \mathcal{L}_{\mathrm{BD}}\}$ is propositionally consistent.

On the basis of the criticism presented by Moore, we might argue that a theories like $\{\mathrm{p}, \mathbf{D}\mathrm{p}\}$ above should have only propositionally inconsistent expansions associated with them. This would require a more strict class of models compared to our choice of propositional models made in Section 2. Unfortunately, this would invalidate the results of Section 5 on capturing stationary default extensions with separated stationary expansions. Therefore we prefer the notion of belief consistency of Definition 20. Moreover, even belief inconsistent separated expansions can be sources of useful information, e.g. for debugging knowledge bases. In Example 6, we can understand $\mathbf{D}\mathrm{p}$ as an integrity constraint requiring that p should not be believed. This is violated by the fact p (in a more realistic setting, p could follow from a larger theory). In this way, detecting that an expansion is belief consistent amounts to finding out that an integrity constraint is not violated. In contrast to this, an autoepistemic theory entailing p and $\neg\mathbf{B}$p can have only inconsistent expansions in conventional autoepistemic logic. By this inevitable inconsistency, all beliefs about the remaining theory are lost. In Theorem 21, we identify a subclass of separated autoepistemic theories for which propositional consistency and belief consistency of separated stable expansions coincide. Note that translations $\mathrm{Tr}_{\mathrm{K}}^{\mathrm{D}}(\langle D, W\rangle)$ of default theories $\langle D, W\rangle$ belong to this class.

Theorem 21. *Assume $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ contains only sentences of the form $\mathbf{B}\phi_1 \wedge \ldots \wedge \mathbf{B}\phi_n \wedge \mathbf{D}\psi_1 \wedge \ldots \wedge \mathbf{D}\psi_m \rightarrow \chi$ where $\chi \in \mathcal{L}$. Then every propositionally consistent separated stable expansion Δ of Σ is also belief consistent.*

Proof sketch. Let Δ be any propositionally consistent separated stable expansion of Σ, i.e. $\Delta = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$ and there is a model $\mathcal{M}$ of Δ. It can be shown that also $\mathcal{M}' = (\mathcal{M} \cap \mathcal{P}) \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta}$ is also a model of Δ. In this model, all instances of the schema $\mathbf{C}$ are true. Thus $\mathcal{M}'$ is also a model of $\mathrm{Cn}_{\mathbf{C}}(\Delta)$, i.e. the separated stable expansion Δ is belief consistent. $\square$

Stalnaker's stability conditions [27] are easily modified to fit our approach. The conditions are (i) $\Delta \models \phi$ implies $\phi \in \Delta$, (ii) $\phi \in \Delta$ implies $\mathbf{B}\phi \in \Delta$ and (iii) $\phi \notin \Delta$ implies $\mathbf{D}\phi \in \Delta$ with the last condition obviously modified. From Definition 1 we can verify that these conditions are met by separated stable expansions. Moreover, our notion of belief consistency provides a way of characterizing separated stable expansions in terms of modal axiom schemas (see below). These axiom schemas are obtained from *propositional equivalents* of the standard axiom schemas **K**, **D**, **4** and **5** by replacing $\neg\mathbf{B}$ by **D** and they describe the relationship of the modality **D** to the propositional connectives $\rightarrow$ and $\neg$ and the modality **B**.

$$
\begin{array}{ll}
\mathbf{K}^{\mathrm{C}} : \mathbf{C}\phi \wedge \mathbf{C}\psi \rightarrow & \mathbf{K}' : \mathbf{B}\phi \wedge \mathbf{D}\psi \rightarrow \mathbf{D}(\phi \rightarrow \psi) \\
\quad (\mathbf{B}\phi \wedge \mathbf{D}\psi \rightarrow \mathbf{D}(\phi \rightarrow \psi)) & \\
\mathbf{D}^{\mathrm{C}} : \mathbf{C}\phi \rightarrow (\mathbf{B}\phi \rightarrow \mathbf{D}\neg\phi) & \mathbf{D}' : \mathbf{B}\phi \rightarrow \mathbf{D}\neg\phi \\
\mathbf{4}^{\mathrm{C}} : \mathbf{C}\mathbf{B}\phi \rightarrow (\mathbf{D}\mathbf{B}\phi \rightarrow \mathbf{D}\phi) & \mathbf{4}' : \mathbf{D}\mathbf{B}\phi \rightarrow \mathbf{D}\phi \\
\mathbf{5}^{\mathrm{C}} : \mathbf{C}\mathbf{D}\phi \rightarrow (\mathbf{D}\mathbf{D}\phi \rightarrow \mathbf{B}\phi) & \mathbf{5}' : \mathbf{D}\mathbf{D}\phi \rightarrow \mathbf{B}\phi
\end{array}
$$

Theorem 22. *A separated stable expansion Δ of $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ contains all instances of axiom schemas $\mathbf{K}^{\mathrm{C}}$, $\mathbf{D}^{\mathrm{C}}$, $\mathbf{4}^{\mathrm{C}}$ and $\mathbf{5}^{\mathrm{C}}$.*

Proof sketch. It is sufficient to consider only propositionally consistent separated stable expansions $\Delta = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$ of Σ. By modified Stalnaker's conditions it is easy to see that $\mathbf{B}\phi \vee \mathbf{D}\phi \in \Delta$ for every $\phi \in \mathcal{L}_{\mathrm{BD}}$. Thus every model $\mathcal{M}$ of a separated stable expansion Δ is a model of $\mathbf{B}\phi$ or a model of $\mathbf{D}\phi$. As an example, we consider $\mathbf{D}^{\mathrm{C}}$. Assuming $\mathbf{C}\phi \rightarrow (\mathbf{B}\phi \rightarrow \mathbf{D}\neg\phi) \notin \Delta$ implies $\mathcal{M} \models \mathbf{C}\phi$, $\mathcal{M} \models \mathbf{B}\phi$ and $\mathcal{M} \not\models \mathbf{D}\neg\phi$ for a model $\mathcal{M}$ of Δ. Thus $\mathcal{M}$ is a model of $\neg(\mathbf{B}\phi \wedge \mathbf{D}\phi)$ as well as $\mathbf{B}\neg\phi \vee \mathbf{D}\neg\phi$ implying that $\mathcal{M} \not\models \mathbf{D}\phi$ and $\mathcal{M} \models \mathbf{B}\neg\phi$. Since $\mathcal{M}$ is a model of $\mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta}$, it follows that $\phi \in \Delta$ and $\neg\phi \in \Delta$. But this contradicts the propositional consistency of Δ. $\square$

To develop a further characterization, we restrict to a fragment of separated autoepistemic logic where the sentences $\{\mathbf{C}\phi \mid \phi \in \mathcal{L}_{\mathrm{BD}}\}$ are taken as axioms. In the sequel, we assume that these axioms are included in the premises Σ directly so that we need not redefine separated stable expansions. It is obvious that Theorem 21 extends for this kind of separated autoepistemic theories. In Theorem 23, we characterize separated stable expansions in terms of the schema $\mathbf{5}'$ which replaces positive introspection. This is an analog of the result by Marek, Schwarz, and Truszczyński [12] stating that propositionally consistent stable expansions Δ of a conventional autoepistemic theory $\Sigma \subseteq \mathcal{L}_{\mathrm{B}}$ satisfy a similar characterization $\Delta = \mathrm{Cn}_{\mathbf{B5}}(\Sigma \cup \neg\mathbf{B}\overline{\Delta})$ where $\mathbf{B}$ denotes the necessitation rule and $\mathbf{5}$ is the standard schema $\neg\mathbf{B}\neg\mathbf{B}\phi \to \mathbf{B}\phi$.

Theorem 23. *Let $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ such that $\{\mathbf{C}\phi \mid \phi \in \mathcal{L}_{\mathrm{BD}}\} \subseteq \Sigma$. Then $\Delta = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$ if and only if $\Delta = \mathrm{Cn}_{\mathbf{B5}'}(\Sigma \cup \mathbf{D}\overline{\Delta})$ for propositionally consistent theories $\Delta \subseteq \mathcal{L}_{\mathrm{BD}}$.*

Proof sketch. ($\Rightarrow$) Now $\mathbf{B}\phi \in \mathbf{B}\Delta$ implies $\mathbf{D}\phi \notin \Delta$ and $\mathbf{DD}\phi \in \Delta$, since $\mathbf{C}\phi \in \Sigma$ and Δ is propositionally consistent. It follows that $\mathbf{B}\phi \in \mathrm{Cn}_{\mathbf{B5}'}(\Sigma \cup \mathbf{D}\overline{\Delta})$ which contains the instance $\mathbf{DD}\phi \to \mathbf{B}\phi$ of $\mathbf{5}'$. Hence $\mathbf{B}\Delta$ and Δ are subsets of $\mathrm{Cn}_{\mathbf{B5}'}(\Sigma \cup \mathbf{D}\overline{\Delta})$. Note that Δ contains all instances of $\mathbf{5}'$ by Theorem 22, since $\mathbf{CD}\phi \in \Sigma \subseteq \Delta$. Then $\mathrm{Cn}_{\mathbf{B5}'}(\Sigma \cup \mathbf{D}\overline{\Delta}) \subseteq \Delta$ can be shown by induction on the lengths of $\mathbf{B}$-proofs. ($\Leftarrow$) Now $\Delta = \mathrm{Cn}_{\mathbf{B5}'}(\Sigma \cup \mathbf{D}\overline{\Delta})$ implies $\Delta = \mathrm{Cn}_{\mathbf{5}'}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$, since Δ is closed with respect to the necessitation rule. It can be shown that all instances of $\mathbf{5}'$ are contained in $\mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$. It follows that $\Delta = \mathrm{Cn}_{\mathbf{5}'}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta}) = \mathrm{Cn}(\Sigma \cup \mathbf{B}\Delta \cup \mathbf{D}\overline{\Delta})$. $\square$

7 Related Work

In Section 4, we placed a lot of emphasis on analyzing the relationship between separated autoepistemic logic and default logic and discussed several approaches to embed default logic into conventional autoepistemic logic. These approaches are not further analyzed, since relevant comparisons were already made. In the sequel, we discuss related work under three further topics. In Section 7.1, we look at other autoepistemic characterizations of stationary extensions and compare them with our approach. The other approaches are only concerned with the well-founded models [29] of normal logic programs. These models correspond to stationary extensions of the default theory obtained by translating program rules $p \leftarrow q_1, ..., q_n, {\sim}r_1, ..., {\sim}r_m$ into defaults $\frac{q_1 \wedge ... \wedge q_m : \neg r_1, ..., \neg r_m}{p}$ [22]. In this sense, autoepistemic characterizations of well-founded models are also autoepistemic characterizations of stationary extensions. Note also that various semantics of normal logic programs can be captured by translating the program rules into $\mathbf{B}q_1 \wedge ... \wedge \mathbf{B}q_n \wedge \mathbf{D}r_1 \wedge ... \wedge \mathbf{D}r_m \to p$ directly. In Section 7.2, we explore non-monotonic modal logics based on two modalities and observe that stationary semantics does not easily extend for them. Alternative notions of extensions are considered in Section 7.3.

7.1 Autoepistemic Characterizations of Stationary Semantics

Przymusinski [25] has studied a variant of autoepistemic logic which is based on three-valued propositional logic. He introduces a notion of three-valued autoepistemic expansions which capture the well-founded models of normal logic programs [29] under a suitable translation. Przymusinski's characterization resorts to a three-valued implication which is closer to two-valued implication than the three-valued implication based on Kleene's strong tables. In contrast, our approach is based on standard two-valued propositional logic where implication has a unique, well-understood interpretation. The three-valued character of expansions is achieved by defining separated stationary expansions in analogy to stationary extensions of default theories.

Przymusinski has proposed also further derivatives of autoepistemic logic [23, 24] by replacing negative introspection by a new primitive corresponding to Lifschitz's parallel circumscription [9]. This leads to the notion of the *static semantics* of disjunctive logic programs. For normal logic programs, this semantics coincides with the well-founded model. In this way, Przymusinski's approach provides a further autoepistemic characterization of the well-founded model though parallel circumscription is used for negative introspection. By our results, it is not necessary to change the negative introspection principle of autoepistemic logic in order to capture well-founded models. The same effect is obtained by changing the notion of expansions.

Bonatti [3] has studied the semantics of logic programs in terms of a three-valued autoepistemic logic. His system captures various models of logic programs including stable models [5], well-founded models [29] and supported models [1]. To achieve this, Bonatti employs *different* translations of the program rules. Marek and Truszczyński [14] have shown that Reiter's extensions and weak extensions capture the stable and supported models of a normal logic programs, respectively. The same is achieved with separated stable/iterative expansions by Theorem 13. Thus our system captures various models of normal logic programs (including stable, well-founded and supported) using the *same translation function* but different kinds of separated expansions (iterative, stationary and stable). This speaks for a uniform view that $\mathrm{Tr}_{\mathrm{K}}^{\mathrm{D}}$ *preserves* the mentioned models of normal logic programs.

Yuan [31] proposes an autoepistemic logic of first order. Our comparison will restrict to the propositional case. In Yuan's approach, the scope of the belief operator **B** is restricted to propositional atoms whereas separated autoepistemic logic is based on a full modal language. To embed default logic in his system, Yuan translates a default rule to a sentence $\mathbf{B}\mathrm{p}_\alpha \wedge \neg\mathbf{B}\mathrm{p}_{\bar{\beta}_1} \wedge \ldots \wedge \neg\mathbf{B}\mathrm{p}_{\bar{\beta}_n} \rightarrow \gamma$ where p_α and $p_{\bar{\beta}_1}, \ldots, p_{\bar{\beta}_n}$ are new atoms. Moreover, he needs additional equivalences $p_\alpha \leftrightarrow \alpha$ and $p_{\bar{\beta}_i} \leftrightarrow \neg\beta_i$ for $0 < i \leq n$. Compared to this, our approach is much simpler: we can do without extra atoms because of the full modal language and the separate modality **D** for disbeliefs. In addition, Yuan's results cover only Reiter's extensions.

7.2 Non-monotonic Logics with Two Modalities

Lin and Shoham [11] propose the *logic of knowledge and justified assumptions* GK. As suggested by the name of the logic, modalities **K** for knowledge and **A** for assumption are used. The semantics of GK is based on Kripke-style models for the two modalities. In particular, Lin and Shoham impose a preference order on models and the semantics of a theory in GK is determined by its *preferred models*. The system GK captures propositionally consistent extensions of default theories by a translation $\mathbf{K}\alpha \wedge \neg\mathbf{A}\neg\beta_1 \wedge ... \wedge \neg\mathbf{A}\neg\beta_n \rightarrow \mathbf{K}\gamma$. Another system with two modalities is the *logic of minimal knowledge with negation as failure* MKNF by Lifschitz [10]. This logic has a strong logic programming flavor, since the second modality is directly the "negation as failure" operator $\sim$. MKNF is also able to capture extensions of default theories trough a translation $\mathbf{K}\alpha \wedge {\sim}\neg\beta_1 \wedge ... \wedge {\sim}\neg\beta_n \rightarrow \mathbf{K}\gamma$ of default rules.

Let us now consider the default theory $\langle\{\frac{:\neg \mathrm{p}}{p}\}, \emptyset\rangle$ to demonstrate that GK and MKNF are incapable of capturing stationary extensions. Lin and Shoham translate the above default into $\mathbf{K}\top \wedge \neg\mathbf{A}\neg\neg\mathrm{p} \rightarrow \mathbf{K}\mathrm{p}$. In GK, this is equivalent to $\neg\mathbf{A}\mathrm{p} \rightarrow \mathbf{K}\mathrm{p}$. This sentence does not have *preferred* models [11, Example 3.4]. Lifschitz translates the same default into $\mathbf{K}\top \wedge {\sim}\neg\neg\mathrm{p} \rightarrow \mathbf{K}\mathrm{p}$. In MKNF, this is equivalent to ${\sim}\mathrm{p} \rightarrow \mathbf{K}\mathrm{p}$ which does not have models [10, Theorem 1, Part B]. This suggests that the Kripke-style models of GK and MKNF are too strict and have to changed in order to capture stationary extensions. To the contrary, the translation $\mathrm{Tr}^{\mathrm{D}}_{\mathrm{K}}(\langle\{\frac{:\neg \mathrm{p}}{p}\}, \emptyset\rangle) = \{\mathbf{B}\top \wedge \mathbf{D}\neg\neg\mathrm{p} \rightarrow \mathrm{p}\}$ has a propositionally consistent separated stationary expansion.

7.3 Alternative Notions of Extensions

So far we have considered three kinds of extensions for default theories including Reiter's extensions, weak extensions and stationary extensions. There are also other possibilities. You and Yuan [30] have proposed *regular models* for normal logic programs that are maximal well-founded models with respect to set inclusion. The same idea can be applied to stationary extensions of default theories and separated stationary expansions of autoepistemic theories. The translation functions $\mathrm{Tr}^{\mathrm{D}}_{\mathrm{K}}$ and $\mathrm{Tr}_{\mathrm{DL}}$ preserve this semantics, since stationary extensions and separated stationary expansions are preserved.

Extension classes proposed by Baral and Subrahmanian [2] form another interesting approach. Such classes are closed under the operator $\Gamma(E) = \mathrm{Cn}^{D_E}(W)$ for a default theory $\langle D, E\rangle$, i.e. if E belongs to a class, so does $\Gamma(E)$. For a finite default theory $\langle D, W\rangle$, a non-empty and *strict* extension class can be represented as a sequence $E_0, ..., E_n$ of extensions where $n \geq 1$, $E_0 = E_n$ and $E_i = \Gamma(E_i)$ for $0 < i \leq n$. Non-empty and strict classes of separated iterative expansions $\Delta_0, ..., \Delta_n$ of $\Sigma \subseteq \mathcal{L}_{\mathrm{BD}}$ can be similarly defined using an operator $\Gamma'(\Delta) = \mathrm{Cn}_{\mathbf{B}}(\Sigma \cup \mathbf{D}\overline{\Delta})$ for a finite autoepistemic theory Σ. It can be shown that even classes of this kind are preserved under the translation functions $\mathrm{Tr}^{\mathrm{D}}_{\mathrm{K}}$ and $\mathrm{Tr}_{\mathrm{DL}}$. This is yet further evidence on the close interconnection of separated autoepistemic logic and default logic.

8 Conclusions

We have proposed a variant of Moore's autoepistemic logic based on separate modalities for belief and disbelief. This kind of separation has several advantages over the conventional approach: (i) there is a counterpart of stationary default extensions in the novel variant and (ii) there are natural ways of translating default theories into autoepistemic theories and back by *linear and modular* translation functions $\mathrm{Tr}_{\mathrm{K}}^{\mathrm{D}}$ and $\mathrm{Tr}_{\mathrm{DL}}$. These preserve various kinds of extensions proposed for default theories – including original extensions by Reiter, weak extensions by Marek and Truszczyński, stationary extensions by Przymusinska and Przymusinski, regular extensions by You and Yuan and extension classes by Baral and Subrahmanian. Moreover, (iii) conventional autoepistemic reasoning can be also captured and (iv) it is possible to handle consistently expansions that are inconsistent at the level of beliefs.

We take the modularity of our translation functions as a strong indication that separated autoepistemic theories and default theories are of equal expressive power under the notions of extensions and separated expansions considered in this paper. Consequently, separated autoepistemic logic and default logic can be applied to very similar knowledge representation tasks. One of the advantages of separated autoepistemic logic is that it provides a uniform query language which allows referring to disbeliefs conveniently. Default logic lacks this possibility and queries expressing that some sentence does not belong to an extension have to be represented somehow else. From the logic programming point of view, separated autoepistemic logic provides a versatile framework where the modality $\mathbf{D}$ gives a declarative meaning to the procedural notion of "negation as failure". Our framework captures various semantics of normal logic programs. We expect that also other classes of logic programs can be handled, since our results hold for arbitrary default theories and not just for the fragment corresponding to normal logic programs.

Despite some fundamental changes with respect to conventional autoepistemic logic, many properties remain untouched. In Section 3, we provided a finitary characterization for separated stable and iterative expansions. This characterization can be extended e.g. for separated stationary expansions and it serves as the basis for developing decision procedures for separated autoepistemic logic. We expect that the computational complexity of major reasoning tasks is not affected. This view is supported the presented translation functions that lead to linear transformations between decision problems of default logic and separated autoepistemic logic. It seems also that efficient implementation techniques of conventional autoepistemic logic and default logic [20, 21] carry over to separated autoepistemic logic. Even improvements are in sight, since separated autoepistemic logic has a reasonable notion of stationary expansions providing approximations when separated iterative expansions are computed. Finally, the concept of separated stationary expansions gives rise to a new class of tractable autoepistemic theories.

9 Acknowledgments

The author thanks Dr. Ilkka Niemelä and anonymous referees for their comments and suggestions to improve the paper. The financial support from Emil Aaltonen Foundation, Foundation of Technology, Alfred Kordelin Foundation as well as Jenny and Antti Wihuri Foundation is gratefully acknowledged.

References

1. K.R. Apt, H.A. Blair, and A. Walker. Towards a theory of declarative knowledge. In J. Minker, editor, *Foundations of Deductive Databases and Logic Programming*, pages 89–148. Morgan Kaufmann Publishers, Los Altos, 1988.
2. C. R. Baral and V.S. Subrahmanian. Stable and extension class theory for logic programs and default logics. *Journal of Automated Reasoning*, 8:345–366, 1992.
3. P. Bonatti. Autoepistemic logics as a unifying framework for the semantics of logic programs. *Journal of Logic Programming*, 22(2):91–149, 1995.
4. D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors. *Handbook of Logic and Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*. Oxford University Press, Oxford, 1994.
5. M. Gelfond and V. Lifschitz. 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.
6. G. Gottlob. Complexity results for nonmonotonic logics. *Journal of Logic and Computation*, 2(3):397–425, June 1992.
7. T. Janhunen. Representing autoepistemic introspection in terms of default rules. In *Proceedings of the European Conference on Artificial Intelligence*, pages 70–74, Budapest, Hungary, August 1996. John Wiley.
8. K. Konolige. On the relation between default and autoepistemic logic. *Artificial Intelligence*, 35:343–382, 1988.
9. V. Lifschitz. Computing circumscription. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pages 121–127, Los Angeles, California, USA, August 1985. Morgan Kaufmann Publishers.
10. V. Lifschitz. Nonmonotonic databases and epistemic queries. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 381–386, Sydney, Australia, August 1991. Morgan Kaufmann Publishers.
11. F. Lin and Y. Shoham. A logic of knowledge and justified assumptions. *Artificial Intelligence*, 57:271–289, 1992.
12. W. Marek, G.F. Schwarz, and M. Truszczyński. Modal nonmonotonic logics: Ranges, characterization, computation. *Journal of the ACM*, 40(4):963–990, 1993.
13. W. Marek and M. Truszczyński. Modal logic for default reasoning. *Annals of Mathematics and Artificial Intelligence*, 1:275–302, 1990.
14. W. Marek and M. Truszczyński. *Nonmonotonic Logic: Context-Dependent Reasoning*. Springer-Verlag, Berlin, 1993.
15. D. McDermott. Non-monotonic logic II. *Journal of the ACM*, 29:33–57, 1982.
16. D. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13:41–72, 1980.
17. R.C. Moore. Autoepistemic logic. In P. Smets, E.H. Mamdani, and D. Dubois, editors, *Non-Standard Logics for Automated Reasoning*, pages 105–136. Academic Press, London, 1988.

18. I. Niemelä. Towards automatic autoepistemic reasoning. In *Proceedings of the European Workshop on Logics in Artificial Intelligence—JELIA'90*, pages 428–443, Amsterdam, The Netherlands, September 1990. Springer-Verlag.
19. I. Niemelä. A unifying framework for nonmonotonic reasoning. In *Proceedings of the 10th European Conference on Artificial Intelligence*, pages 334–338, Vienna, Austria, August 1992. John Wiley.
20. I. Niemelä. A decision method for nonmonotonic reasoning based on autoepistemic reasoning. *Journal of Automated Reasoning*, 14:3–42, 1995.
21. I. Niemelä. Towards efficient default reasoning. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 312–318, Montreal, Canada, August 1995. Morgan Kaufmann Publishers.
22. H. Przymusinska and T.C. Przymusinski. Stationary default extensions. In *Working Notes of the 4th International Workshop on on Nonmonotonic Reasoning*, pages 179–193, Plymouth, Vermont, USA, May 1992.
23. T. C. Przymusinski. Static semantics for normal and disjunctive logic programs. *Annals of Mathematics and Artificial Intelligence*, 14:323–357, 1995.
24. T.C. Przymusinski. 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.
25. T.C. Przymusinski. Three-valued nonmonotonic formalisms and semantics of logic programming. *Artificial Intelligence*, 49:309–343, 1991.
26. R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.
27. R. Stalnaker. A note on non-monotonic modal logic. *Artificial Intelligence*, 64:183–196, 1993.
28. M. Truszczyński. Embedding default logic into modal nonmonotonic logics. In *Proceedings of the 1st International Workshop on Logic Programming and Nonmonotonic Reasoning*, pages 151–165, Washington, D.C., USA, July 1991. The MIT Press.
29. A. Van Gelder, K.A. Ross, and J.S. Schlipf. The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):620–650, July 1991.
30. J. You and L. Yuan. A three-valued semantics for deductive databases and logic programs. *Journal of Computer and System Sciences*, 49:334–361, 1994.
31. L. Yuan. Autoepistemic logic of first order and its expressive power. *Journal of Automated Reasoning*, 13(1):69–82, 1994.

This article was processed using the LaTeX macro package with LLNCS style