Computation of the semantics of autoepistemic belief theories

Stefan Brass a, Jürgen Dix b,*, Teodor C. Przymusinski c,1

a School of Information Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA
b Department of Computer Science, University of Maryland, A.V. Williams Build., College Park,
MD 20752-3255, USA
c Department of Computer Science, University of California, Riverside, CA 92521, USA

Received 24 February 1999; received in revised form 31 May 1999

Abstract

Recently, one of the authors introduced a simple and yet powerful non-monotonic knowledge representation framework, called the Autoepistemic Logic of Beliefs, AEB. Theories in AEB are called autoepistemic belief theories. Every belief theory $T$ has been shown to have the least static expansion $\overline{T}$ which is computed by iterating a natural monotonic belief closure operator $\Psi_T$ starting from $T$. This way, the least static expansion $\overline{T}$ of any belief theory provides its natural non-monotonic semantics which is called the static semantics.

It is easy to see that if a belief theory $T$ is finite then the construction of its least static expansion $\overline{T}$ stops after countably many iterations. However, a somewhat surprising result obtained in this paper shows that the least static expansion of any finite belief theory $T$ is in fact obtained by means of a single iteration of the belief closure operator $\Psi_T$ (although this requires $T$ to be of a special form, we also show that $T$ can be always put in this form). This result eliminates the need for multiple iterations in the computation of static semantics and allows us to replace the fixed-point definition of static semantics by the equivalent explicit and straightforward definition given by $\overline{T} = \Psi_T(T)$.

The second, closely related result establishes an intriguing relationship between static semantics $\overline{T}$ and Clark’s completions $\text{comp}(T)$ of finite belief theories. Here we use a slightly generalized version of $\text{comp}(T)$ (see Definition 3.2). It shows that the static semantics $\overline{T}$ of $T$ is obtained by augmenting $T$ with the set $B_{\text{comp}(T)} = \{BF: F \in \text{comp}(T)\}$ thus ensuring that all formulae that belong to Clark’s completion $\text{comp}(T)$ of $T$ are believed to be true.


* Corresponding author. Email: dix@cs.umd.edu.

1 Partially supported by the National Science Foundation grant #IRI-9313061.
Both results open the way for a more efficient implementation of static semantics: the first, because only one iteration is needed, and the second because reasoning in a non-standard logic (belief theories under static semantics) can be reduced to classical theorem proving involving Clark’s completion.

© 1999 Elsevier Science B.V. All rights reserved.

Keywords: Non-monotonic reasoning; Logics of knowledge and beliefs; Semantics of logic program; Deductive databases

1. Introduction

The problem of finding inference mechanisms, capable to model human commonsense reasoning and to reason in the absence of complete information, is one of the major research and implementation problems in AI. Accordingly, over the last 20–30 years a large number of formalizations of commonsense (or non-monotonic) reasoning, and their inference engines, have been proposed. Among these, the most prominent and most actively studied turned out to be Reiter’s CW A [26], McCarthy’s circumscription [18], Moore’s autoepistemic logic [20], Reiter’s default logic [25] and Kowalski’s logic programming [17].

While logic programming is often just viewed as a formal specification for declarative programming languages, such as PROLOG, in the authors’ view and in the view of many researchers actively studying this area, logic programming plays a much more important role as a non-monotonic knowledge representation language. Its importance stems from several factors. First of all, logic programming is a relatively simple non-monotonic formalism (compared to other major formalisms): it allows only atomic heads of clauses and implements only one non-monotonic feature, namely, default negation (or negation as failure) and it is restricted to atomic formulae. On the other hand, as shown by the extensive research conducted during the last decade or two, logic programming is sufficiently expressive to allow formalization of many important problems in commonsense reasoning. As a result, logic programming offers fertile testing grounds for studying, relating and comparing of various formalizations of commonsense reasoning in a relatively simple and yet quite expressive framework.

The results of such extensive studies have been quite impressive. A number of different semantics for logic programming have been proposed and studied in detail. Two of them turned out to be most promising: the stable semantics [15] and the well founded semantics [31]. The relationship of the stable semantics to other major non-monotonic formalisms has been well established. It turns out that, under a suitable embedding, static semantics can be simply viewed as a special case of Moore’s autoepistemic logic AEL [20].

Recently, these semantics have also been used in various areas of traditional AI, like planning [9], agent reasoning [10,13,14,28,29] or in verification (see references in [30]). In particular it has been shown that systems based on these semantics (like smodels [21] or XSB [24]) can successfully compete with dedicated systems (the recent workshop on Logic based AI organized by Jack Minker, June 1999, gave impressive evidence for this).

Consequently, a further analysis of how these semantics can be extended, unified and computed in a feasible way will most probably lead to new applications or better algorithms in the above mentioned or similar areas of AI. Our approach can be viewed as extending
traditional semantics to semantics dealing with beliefs. Beliefs play an important role whenever available information is incomplete and can be described using statements expressing beliefs.

Our investigations show that the fixpoint of a certain sequence defining the static semantics (which itself contains most of the traditional non-monotonic semantics) is:

1. reached after finitely many steps, and
2. it is closely related to a version of Clark’s completion for finite belief theories.

Thus our results may also have an impact on those semantics.

We now explain our results in more detail. In [22,23], one of the authors introduced a simple and yet powerful non-monotonic knowledge representation framework which isomorphically contains all of the above mentioned non-monotonic formalisms and semantics as special cases and yet is significantly more expressive than each one of these formalisms considered individually. The new logic, called the Autoepistemic Logic of Beliefs, AEB, has been shown to constitute a powerful new formalism which can serve as a unifying framework for several major non-monotonic formalisms. It allows us to better understand mutual relationships existing between different formalisms and semantics and enables us to provide them with simpler and more natural definitions. It also naturally leads to new, even more expressive, flexible and modular formalizations and semantics [23].

Theories in the Autoepistemic Logic of Beliefs are called autoepistemic belief theories. As it is the case in Moore’s autoepistemic logic, AEL, semantics of belief theories is defined by means of introducing a class of their so called static expansions. Static expansions are fixed points of a natural monotonic belief closure operator \( \Psi_T \). However, as opposed to autoepistemic logic and its stable expansions, every belief theory \( T \) has the least static expansion \( T \) which can be computed by iterating the belief operator \( \Psi_T \). This way, the least static expansion of any belief theory provides its natural non-monotonic semantics which is called the static semantics, namely the set of all formulae contained in the least static expansion \( T \).

It is easy to see that if a belief theory \( T \) is finite then the construction of its least static expansion \( T \) will stop after countably many iterations. However, a powerful and somewhat surprising result obtained in this paper shows that the least static expansion of any finite belief theory \( T \) is in fact obtained by means of a single iteration of the belief closure operator \( \Psi_T \). We will later show that \( T \) can be always put in this form. This result eliminates the need for multiple iterations in the computation of static semantics and allows us to replace the fixed-point definition of static semantics by the equivalent explicit and straightforward definition given by \( T = \Psi_T(T) \). Needless to say, the existence of an equivalent non-fixed-point definition of static semantics significantly simplifies this notion and the underlying theory. It also provides the foundation for the interesting results and applications obtained in [5].

The second, closely related result establishes a very interesting and somewhat intriguing relationship between static semantics \( T \) and Clark’s completions \( \text{comp}(T) \) of finite belief theories. Note that we use here a slightly generalized version of Clark’s completion introduced in [1,4] (see Definition 3.2). Namely we first transform the theory \( T \) in a special normal form, called the residual theory. This transformed theory has exactly the

---

\(^2\) To be precise, this requires \( T \) to be in a special normal form, called a residual program.
It shows that the static semantics $T$ of $T$ is obtained by augmenting $T$ with the set $Bcomp(T) = \{BF: F \in comp(T)\}$ thus ensuring that all formulae that belong to Clark’s completion $comp(T)$ of $T$ are believed to be true. It reduces reasoning under the static semantics $T$ to the easily accomplished computation of Clark’s completion $comp(T)$ together with theorem proving in the underlying modal logic $AEB$, which can be done either by hand or using an automated theorem prover.

Both results open the way for a more efficient implementation of static semantics: the first, because only one iteration is needed, and the second because reasoning in a non-standard logic (belief theories under static semantics) can be reduced to classical theorem proving involving Clark’s completion.

2. Autoepistemic belief theories

We begin by recalling the definition and basic properties of autoepistemic belief theories, $AEB$, originally introduced in [22,23]. Consider a fixed propositional language $L$ with standard connectives ($\vee$, $\wedge$, $\neg$, $\rightarrow$, $\leftrightarrow$, $\equiv$) and the propositional letters true and false. We denote the set of its propositions by $(At)$. Extend the language $L$ to a propositional modal language $L_{AEB}$ by augmenting it with a modal operator $B$, called the belief operator. The formulae of the form $BF$, where $F$ is an arbitrary formula of $L_{AEB}$, are called belief atoms and are considered to be atomic formulae in the extended propositional modal language $L_{AEB}$. The formulae of the original language $L$ are called objective. Any propositional theory in the modal language $L_{AEB}$ will be called an autoepistemic belief theory (or just a “belief theory” for short).

**Definition 2.1** (Autoepistemic belief theories—Przymusinski [23]). By a belief theory we mean an arbitrary theory in the propositional language $L_{AEB}$, i.e., a (possibly infinite) set of clauses of the form:

$$B_1 \wedge \cdots \wedge B_m \wedge BG_1 \wedge \cdots \wedge BG_k \rightarrow A_1 \vee \cdots \vee A_l \vee BF_1 \vee \cdots \vee BF_n,$$

where $m, n, k, l \geq 0$, $A_i$’s and $B_i$’s are objective atoms and $F_i$’s and $G_i$’s are arbitrary formulae of $L_{AEB}$.

We assume the following two simple axiom schemata and one inference rule describing the arguably obvious properties of belief atoms:

**CA** Consistency Axiom. $B(\text{true})$ and $\neg B(\text{false})$.

**DA** Distributive Axiom. $B(F \land G) \leftrightarrow BF \land BG$.

**IR** Invariance Inference Rule. \[
\frac{F \leftrightarrow G}{BF \leftrightarrow BG}.
\]
Remark 2.1. Although the set of axioms used in the above definition is slightly different from the one used in the original definition of autoepistemic belief theories given in [23], the two axiom systems are completely equivalent (see Appendix A.4).

As usual, a mapping $I : At_L \cup \{AF : F \in L_{AEB}\} \rightarrow \{true, false\}$ denotes a propositional interpretation of $L_{AEB}$, i.e., we simply treat the formulae $AF$ as new propositions. Therefore, the notion of a model carries over from propositional logic. A formula $F \in L_{AEB}$ is a propositional consequence of $T \subseteq L_{AEB}$ iff for every interpretation $I : I \models T \Rightarrow I \models F$. In the examples, we will represent models by sets of literals showing the truth values of only those objective and belief atoms which are relevant to our considerations.

Definition 2.2 (Derivable formulae (Przymusinski [23])). For any belief theory $T$ we denote by $C_{AEB}(T)$ the smallest set of formulae of the language $L_{AEB}$ which contains $T$, all (substitution instances of) the axioms CA and DA and is closed under both propositional consequence and the invariance rule IR.

We say that a formula $F$ is derivable from the belief theory $T$ if $F$ belongs to $C_{AEB}(T)$. We denote this fact by $T \vdash_{AEB} F$. A belief theory $T$ is consistent if $C_{AEB}(T)$ is consistent, i.e., if $C_{AEB}(T) \not\vdash_{AEB} false$.

From the results established in [23] it immediately follows that the following lemma, which will be needed in the sequel, holds.

Proposition 2.3. For any belief theory $T$ and formulae $F, G \in L_{AEB}$:

- $T \vdash_{AEB} (AF \rightarrow \neg B \neg F)$,
- $T \vdash_{AEB} (B (F \rightarrow G) \land BF) \rightarrow BG$.
- If $T \vdash_{AEB} F$ then $T \vdash_{AEB} BF$.

The first statement says that if a formula $F$ is believed to be true then its negation $\neg F$ is not. The second statement says that if the formulae $F \rightarrow G$ and $F$ are believed then so is $G$. The last inference rule says that if a formula $F$ is derivable then it is also believed.

2.1. Semantics of autoepistemic belief theories

We now need to define a non-monotonic semantics for autoepistemic belief theories. Intuitively, we need to provide a meaning to the belief atoms $AF$. We want the intended meaning of belief atoms $BF$ to be based on the principle of predicate minimization (see [16,18,19]): $BF$ holds if $F$ is minimally entailed, or, equivalently: $BF$ holds if $F$ is true in all minimal models. In order to make this intended meaning precise we first have to define what we mean by a minimal model of a belief theory.

Definition 2.4 (Minimal models—Przymusinski [23]). A model $M$ is smaller than a model $N$ if it contains the same belief atoms but has fewer objective atoms, i.e.,

$\{p \in At_L : M \models p\} \subset \{p \in At_L : N \models p\}$.
By a minimal model of a belief theory $T$ we mean a model $M$ of $T$ with the property that there is no smaller model $N$ of $T$. If a formula $F$ is true in all minimal models of $T$ then we write $T \models_{\min} F$ and say that $F$ is minimally entailed by $T$.

Example 2.5. Consider the following simple belief theory $T$:

\[ \neg Car \rightarrow Car. \]

Let us prove that $T$ minimally entails $\neg Broken$, i.e., $T \models_{\min} \neg Broken$. Indeed, in order to find minimal models of $T$ we need to assign an arbitrary truth value to the only belief atom $B: \neg Broken$, and then minimize the objective atoms $\neg Broken$, $Car$ and $Runs$. We easily see that $T$ has the following two minimal models (truth values of the remaining belief atoms are irrelevant and are therefore omitted): $M_1 = [B: \neg Broken, Car, Runs, \neg Broken]$ and $M_2 = [\neg B: \neg Broken, Car, \neg Runs, \neg Broken]$. Since in both of them $Car$ is true, and $Broken$ is false, we deduce that $T \models_{\min} Car$ and $T \models_{\min} \neg Broken$.

As in Moore’s Autoepistemic Logic, the intended meaning of belief atoms in autoepistemic belief theories is enforced by defining suitable expansions of such theories.

Definition 2.6 (Static expansions of belief theories—Przymusinski [22,23]). A belief theory $T$ is called a static expansion of a belief theory $T$ if it satisfies the following fixed-point equation

\[ T = Cn_{AEB} \left( T \cup \{ BF: T \models_{\min} F \} \right), \]

where $F$ ranges over all formulae in $L_{AEB}$.

Thus a static expansion $T^\circ$ of $T$ is obtained by augmenting $T$ with all belief atoms $BF$ with the property that $F$ is minimally entailed by the fixed point $T^\circ$.

Theorem 2.7 (Least static expansion (Przymusinski [22,23])). Every belief theory $T$ has the least static expansion, namely, the least fixed point $\overline{T}$ of the operator $\Psi_T(X)Cn_{AEB}(T \cup \{ BF: X \models_{\min} F \})$. Moreover this belief closure operator $\Psi_T$ is monotonic.

More precisely, the least static expansion $\overline{T}$ of $T$ can be constructed as follows. Let $T^0 = T$ and suppose that $T^\alpha$ has already been defined for any ordinal number $\alpha < \beta$. If $\beta = \alpha + 1$ is a successor ordinal then define:

\[ T^{\alpha + 1} = \Psi_T \left( T^\alpha \right) \overset{\text{def}}{=} Cn_{AEB}(T \cup \{ BF: T^\alpha \models_{\min} F \}), \]

where $F$ ranges over all formulae in $L_{AEB}$. Else, if $\beta$ is a limit ordinal then define $T^\beta = \bigcup_{\alpha < \beta} T^\alpha$. The sequence $\{ T^\alpha \}$ is monotonically increasing and has a unique fixed point $\overline{T} = T^\lambda = \Psi_T(T^\lambda)$, for some ordinal $\lambda$.

The above result allows us to establish the following useful characterization of the least static completion of a belief theory.
Theorem 2.8 (Characterization of least static expansions). The least static expansion of a belief theory $T$ coincides with the smallest theory $\tilde{T}$ satisfying the conditions:

1. $T \subseteq \tilde{T}$;
2. $\tilde{T} = \text{Cn}_{AEB}(\tilde{T})$;
3. if $\tilde{T} \models_{\min} F$ then $BF \in \tilde{T}$.

Proof. Clearly the least static expansion $\tilde{T}$ of $T$ satisfies the above conditions (1)–(3) so $\tilde{T} \subseteq T$. Moreover, from the conditions (1)–(3) and the very definition of $\Psi_T$ it follows immediately that $\Psi_T(\tilde{T}) \subseteq \tilde{T}$. Since the operator $\Psi_T$ is monotonic on such theories (see [23]), we conclude that $T^1 = \Psi_T(T) \subseteq \Psi_T(\tilde{T}) \subseteq \tilde{T}$, and, more generally, $T^\alpha \subseteq \tilde{T}$, for any ordinal $\alpha$. We conclude therefore that $T \subseteq \tilde{T}$. □

Observe that the least static expansion $\tilde{T}$ of $T$ contains those and only those formulae which are true in all static expansions of $T$. It constitutes the so called static completion of a belief theory $T$.

Example 2.9. Consider a slightly more complex belief theory $T$:

$B \sim \text{Broken} \rightarrow \text{Runs}$.

$B \sim \text{Fixed} \rightarrow \text{Broken}$.

In order to iteratively compute its static completion $\tilde{T}$ we let $T^0 = T$. As in Example 2.5, one easily checks that $T^0 \models_{\min} \neg \text{Fixed}$. Since

$$T^1 = \Psi_T(T^0) = \text{Cn}_{AEB}(T \cup \{BF : T^0 \models_{\min} F\})$$

it follows that $B \sim \text{Fixed} \in T^1$ and therefore $\text{Broken} \in T^1$. Since,

$$T^2 = \Psi_T(T^1) = \text{Cn}_{AEB}(T \cup \{BF : T^1 \models_{\min} F\})$$

it follows that $B \sim \text{Broken} \in T^2$. From Proposition 2.3 we conclude that $\neg B \sim \text{Broken} \in T^2$ and therefore $T^2 \models_{\min} \neg \text{Runs}$. Accordingly, since:

$$T^3 = \Psi_T(T^2) = \text{Cn}_{AEB}(T \cup \{BF : T^2 \models_{\min} F\})$$

we infer that $B \sim \text{Runs} \in T^3$. As expected, the static completion $\tilde{T}$ of $T$, which contains $T^3$, asserts that the car is considered not to be fixed and therefore broken and thus is not in a running condition.

Definition 2.10 (Static completion and static semantics). The least static expansion $\tilde{T}$ of a belief theory $T$ is called the static completion of $T$. It describes the static semantics of a belief theory $T$.

Consequently, like the predicate completion semantics of a logic program $P$ is completely determined by its Clark’s completion $\text{comp}(P)$, the static semantics of a belief theory $T$ is fully determined by its static completion $\tilde{T}$. It is easy to verify that a belief theory $T$ either has a consistent static completion $\tilde{T}$ or it does not have any consistent static expansions at all.
3. Static completions versus Clark’s completions

It is easy to see that if a belief theory is finite then the construction of its static completion (or the least static expansion) will stop after countably many steps. However, the surprising result obtained in this article shows that static completions of finite belief theories \( T \) are in fact obtained by means of a single iteration of the belief closure operator \( \Psi_T \). This result eliminates the need for multiple iterations in the computation of static completions and allows us to replace the fixed-point definition of static completions by the equivalent explicit definition given by \( \overline{T} = \Psi_T(T) \).

The second, closely related result establishes a very interesting and somewhat intriguing relationship between static completions \( \overline{T} \) and Clark’s completions \( \text{comp}(T) \) of finite belief theories. It shows that the static completion \( \overline{T} \) of a belief theory \( T \) coincides with \( C_{AEB}(T \cup \{ BF: F \in \text{comp}(T) \}) \), i.e., with the set of formulae derivable from the belief theory \( T \) augmented with the set

\[
\text{Bcomp}(T) = \{ BF: F \in \text{comp}(T) \}.
\]

The latter set represents the set of beliefs in formulae that belong to Clark’s completion. Even though, strictly speaking, the last result only applies to the so called residual theories (see below), we do not lose generality, as we will show that every finite belief theory can be transformed into a finite residual belief theory with an equivalent set of minimal models. Since Clark’s completion \( \text{comp}(T) \) is easily computable, this result reduces reasoning under the static semantics to theorem proving in the underlying modal logic \( AEB \) (this can be done either by hand or by using an automated theorem prover).

The proof of these two powerful results is based on the idea of adding to a belief theory \( T \) the set of formulae which ensure that models “seen” through the belief operator \( B \) are in fact minimal. As we will soon see, this task can be accomplished by a suitable generalization of Clark’s completion but it only works for a restricted class of belief theories, namely those whose clauses do not have any objective premises. Such residual theories were previously introduced and investigated in the class of logic programs [1,4,6,7,11,12]. Here we give a slightly more general definition.

**Definition 3.1 (Residual belief theories—Brass and Dix [1])**. Arbitrary belief theory whose clauses do not contain any objective (positive) premises, i.e., a (possibly infinite) set of arbitrary clauses

\[
BG_1 \land \cdots \land BG_k \Rightarrow A_1 \lor \cdots \lor A_l \lor BF_1 \lor \cdots \lor BF_n,
\]

where \( n, k, l \geq 0 \), \( A_i \)'s are objective atoms and \( F_i \)'s and \( G_i \)'s are arbitrary formulae of \( L_{AEB} \).

Whenever convenient, clauses

\[
BG_1 \land \cdots \land BG_k \Rightarrow BF_1 \lor \cdots \lor BF_n,
\]

i.e., clauses without any objective atoms in their heads, will be considered as clauses with a single objective atom \( \text{false} \) in their head, i.e., clauses of the form:

\[
BG_1 \land \cdots \land BG_k \Rightarrow \text{false} \lor BF_1 \lor \cdots \lor BF_n.
\]
Now we describe a natural extension of the notion of Clark’s completion $\text{comp}(T)$. Clark’s completion was initially introduced in [8] for the class of normal logic programs. Subsequently, its generalizations to disjunctive programs without positive premises were studied in [1]. Here we extend it to the class of all residual belief theories. As usual, the goal is to ensure that an objective atom $A$ is true only if it really has to be true, i.e., if there is a rule with $A$ in the head, in which the body is true and all the other head literals are false.

**Definition 3.2 (Extended Clark’s completion).** Given a finite residual belief theory $T$ and an objective atom $A$ from $\mathcal{L}$ (including the falsity atom $\text{false}$) we define the extended Clark’s completion $\text{comp}(A, T)$ of the atom $A$ in $T$ to be the formula:

$$A \iff \bigvee_{1 \leq m \leq s} (B G_{1,m} \wedge \cdots \wedge B G_{k,m} \wedge \neg A_{1,m} \wedge \cdots \wedge \neg A_{l,m} \wedge \neg B F_{1,m} \wedge \cdots \wedge \neg B F_{n,m}),$$

where

$$B G_{1,m} \wedge \cdots \wedge B G_{k,m} \implies A \vee A_{1,m} \vee \cdots \vee A_{l,m} \vee B F_{1,m} \vee \cdots \vee B F_{n,m},$$

for $1 \leq m \leq s$, are all clauses in $T$ containing the atom $A$ in their heads.

By the extended Clark’s completion $\text{comp}(T)$ of a belief theory $T$ we mean the union of all completions $\text{comp}(A, T)$ of $A$ in $T$, for all objective atoms $A$ in the language $\mathcal{L}$ (including the falsity atom $\text{false}$).

The following interesting result shows that if $T$ is a residual belief theory, then the models of Clark’s completion $\text{comp}(T)$ are precisely the minimal models of $T$ itself. In other words, Clark’s completion precisely describes the minimal model semantics of a residual belief theory $T$.

**Proposition 3.3.** Let $T$ be a finite residual belief theory. An interpretation $M$ of the language $\mathcal{L}_{AEB}$ is a minimal model of $T$ if and only if $M$ is a model of Clark’s completion $\text{comp}(T)$ of $T$.

**Proof.** Contained in Appendix A.1. $\square$

It is worth pointing out that one of the weaknesses of the original version of Clark’s completion proposed in [8] was the fact that it was applied not just to residual but rather to arbitrary normal programs. As a result, Clark’s original completion did not enforce minimal model semantics, e.g., a tautology like $p \iff p$ made the completion axiom for $p$ useless. Another problem with the original definition of Clark’s completion was that it did not distinguish between logical negation $\neg p$ and belief $\mathcal{B}\neg p$. As a result, Clark’s completion could be inconsistent when rules like $p \iff \mathcal{B}\neg p$ were present.

Although the above result, as well as the definition of Clark’s completion, applies only to the class of residual belief theories, it turns out that every finite belief theory $T$ can be transformed into a finite residual belief theory $T_{\text{res}}$ so that their sets of minimal models coincide.
Proposition 3.4. Every finite belief theory $T$ can be transformed into a finite residual belief theory $T_{\text{res}}$, called the residuum of $T$, so that an interpretation $M$ of the language $\mathcal{L}_{\text{AEB}}$ is a minimal model of $T$ if and only if $M$ is a minimal model of $T_{\text{res}}$.

Proof. Contained in Appendix A.2. \hfill \square

In fact, just two elementary theory transformations, namely, “unfolding” (or partial evaluation, GPPE [4,27]) and “elimination of tautologies”, are sufficient to obtain such a residual program. For a detailed definition of these transformations the reader is referred to [1–4]. Here we describe them very briefly.

The “elimination of tautologies” allows us to remove a rule if it contains the same atom in both, head and body. Such rules are always trivially satisfied and therefore useless. The transformation of “unfolding”—also called GPPE—is defined as follows. Suppose that $B$ is an objective atom and let $B \lor A_i \leftarrow R_i$, $i = 1, \ldots, n$, be all rules about $B$, i.e., rules containing $B$ in their head. The application of GPPE to the rule $A \leftarrow B \land R$, that contains $B$ in its body, results in a new belief theory obtained by deleting the rule $A \leftarrow B \land R$ and adding to the belief theory the new rules $A \lor A_i \leftarrow R \land R_i$.

We now give a simple example. Whenever convenient, we use the traditional inverse notation for clauses. Consider the following belief theory $T$. The first two clauses already have the required form. In the last clause, we unfold the body literal $q$ and obtain the residual program $T_{\text{res}}$.

\begin{align*}
T: & p \lor q \leftarrow B \neg r. & T_{\text{res}}: & p \lor q \leftarrow B \neg r. \\
& q \leftarrow B \neg q. & & q \leftarrow B \neg q. \\
& r \leftarrow q. & & p \lor r \leftarrow B \neg r. \\
& & & r \leftarrow B \neg q.
\end{align*}

One can easily see that this procedure represents a form of hyper-resolution.

As an immediate consequence of Propositions 3.3 and 3.4 we obtain the following important theorem which says that for any finite belief theory $T$, Clark’s completion of its residuum $T_{\text{res}}$ precisely describes the minimal model semantics of $T$.

Theorem 3.5. Let $T$ be any finite belief theory. An interpretation $M$ of the language $\mathcal{L}_{\text{AEB}}$ is a minimal model of $T$ if and only if $M$ is a model of Clark’s completion $\text{comp}(T_{\text{res}})$ of the residuum $T_{\text{res}}$ of $T$.

Finally, let us denote by $\mathcal{B}_{\text{comp}}(T_{\text{res}})$ the set \{BF: $F \in \text{comp}(T_{\text{res}})$\}. Intuitively, augmenting a given belief theory $S$ with $\mathcal{B}_{\text{comp}}(T_{\text{res}})$ ensures that all formulae that belong to Clark’s completion $\text{comp}(T_{\text{res}})$ are believed. We will need the following lemma.

Lemma 3.6. Let $T$ be an arbitrary finite belief theory and let

\[ \hat{T} = C_{\text{naeb}}(T \cup \mathcal{B}_{\text{comp}}(T_{\text{res}})). \]

If a formula $F$ is true in all minimal models of $\hat{T}$ then $\mathcal{B}F$ is already contained in $\hat{T}$. In other words, $\Psi_T(\hat{T}) \subseteq \hat{T}$.
Proof. Contained in Appendix A.3. □

Now we are ready to state the first of the two fundamental results obtained in this article.

**Theorem 3.7** (Static completion versus Clark’s completion). The static completion $T$ of a finite belief theory $T$ can be equivalently defined as:

$$T = C_{NaE}(T \cup \mathcal{B} \mathcal{C}(T_{res}))$$

**Proof.** Let $\hat{T} = C_{NaE}(T \cup \mathcal{B} \mathcal{C}(T_{res}))$. We first show that $\hat{T} = \Psi_T(T)$, i.e., $\hat{T} = C_{NaE}(T \cup \{BF: T \models_{\min} F\})$.

1. Let $BG \in \mathcal{B} \mathcal{C}(T_{res})$, i.e., $G \in \mathcal{C}(T_{res})$. By Theorem 3.5, $T \models_{\min} G$, and thus $\mathcal{B}(G)$ is contained in the right hand side.

2. Let $T \models_{\min} F$. By Theorem 3.5, $\mathcal{C}(T_{res}) \models F$. Then there are $G_1, \ldots, G_n \in \mathcal{C}(T_{res})$ such that

$$G_1 \rightarrow (G_2 \rightarrow (\ldots (G_n \rightarrow F) \ldots))$$

is a propositional tautology. Thus $\mathcal{B}(G_1 \rightarrow (G_2 \rightarrow (\ldots (G_n \rightarrow F) \ldots)))$ is contained in $\hat{T}$ (see Proposition 2.3). Since $\mathcal{B}(G_1) \in \hat{T}$ by construction, we get $BF \in \hat{T}$ (again by Proposition 2.3).

From Lemma 3.6 it follows that $\Psi_T(\hat{T}) \subseteq \hat{T}$. Since $\hat{T} = \Psi_T(T)$ we conclude that:

$$\Psi_T(\Psi_T(T)) \subseteq \Psi_T(T)$$

which means that $\Psi_T(\Psi_T(T)) = \Psi_T(T)$ (because $\Psi_T$ is monotonically increasing by Theorem 2.7) and therefore $\Psi_T(\hat{T}) = \hat{T}$. This proves that $\hat{T}$ is a fixed point of the belief closure operator $\Psi_T$ and thus is a static expansion of $T$. Since $\hat{T} = \Psi_T(T)$ we conclude that $\hat{T}$ is the least such fixed point and thus coincides with the static completion $\mathcal{T}$ of $T$. □

The above result states that the static completion $\mathcal{T}$ of $T$ is obtained by augmenting $T$ with the set $\mathcal{B} \mathcal{C}(T_{res})$ thus ensuring that all formulae that belong to Clark’s completion $\mathcal{C}(T_{res})$ of $T_{res}$ are believed. It establishes an interesting and somewhat intriguing relationship between static completions and Clark’s completions of finite belief theories. It also reduces reasoning under the static semantics (i.e., the computation of static completions $\mathcal{T}$) to the easily accomplished computation of Clark’s completion $\mathcal{C}(T_{res})$ together with theorem proving in the underlying modal logic.

As a consequence of the proof of the above result we immediately obtain the second of the two fundamental theorems obtained in this paper, namely a result stating that static completion $\mathcal{T}$ of an arbitrary finite belief theory $T$ is always obtained by a single iteration of the belief closure operator $\Psi_T(T)$.

**Theorem 3.8** (Explicit characterization of static completions). The static completion $\mathcal{T}$ of finite belief theory $T$ is always obtained by a single iteration of the belief closure operator $\Psi_T(T)$: $\mathcal{T} = \Psi_T(T)$. In other words,

$$\mathcal{T} = C_{NaE}(T \cup \{BF: T \models_{\min} F\})$$
Proof. We have to compare the definition of the operator $\Psi_T$ with the statement in Theorem 3.7. In particular we have to compare the set $\{BF: T \models_{\min} F\}$ with the set $\mathcal{B} \operatorname{comp}(T_{\text{res}})$. By Proposition 3.3 they are identical. □

The above theorem states that the static completion $\overline{T}$ of $T$ is obtained by augmenting $T$ with beliefs $BF$ in all formulae $F$ which are true in all minimal models of the theory $T$ itself. It eliminates therefore the need for multiple iterations in the computation of static completions and allows us to replace the fixed-point definition of static completions by the equivalent explicit definition. In addition to Theorems 2.7, 2.8 and 3.7, this is the fourth characterization of static completions presented in this paper. Needless to say, the existence of an equivalent non-fixed-point definition of static completions significantly simplifies this notion and the underlying theory. It also provides the foundation for the results obtained in [5].

Example 3.9. Consider the belief theory $T$ from Example 2.9. We already noted that $T \models_{\min} \neg\text{Fixed}$ and thus

$$\mathcal{B} \models \neg\text{Fixed} \in \overline{T} = \text{Ch}_{\text{AE}}(T \cup \{BF: T \models_{\min} F\}).$$

This implies that $\overline{T}$ contains also $\text{Broken}$. By Proposition 2.3, $\mathcal{B} \text{Broken}$ belongs to the static completion and so does $\neg\mathcal{B} \text{Broken}$. From the same proposition it follows that $\mathcal{B} \text{Broken}$. Moreover, it is easy to verify that $T \models_{\min} \mathcal{B} \text{Broken} \leftrightarrow \text{Runs}$ (in fact, the formula $\mathcal{B} \text{Broken} \leftrightarrow \text{Runs}$ belongs to Clark’s completion of $T$) and therefore

$$T \models_{\min} \neg\mathcal{B} \text{Broken} \leftrightarrow \neg\text{Runs}.$$

It follows that $\overline{T}$ contains $\mathcal{B}(\neg\mathcal{B} \text{Broken} \leftrightarrow \neg\text{Runs})$, which can also be written in the form

$$\mathcal{B}(\neg\mathcal{B} \text{Broken} \rightarrow \neg\text{Runs}) \land (\neg\mathcal{B} \text{Broken} \leftarrow \neg\text{Runs}),$$

and thus the distributive axiom DA can be applied:

$$\mathcal{B}(\neg\mathcal{B} \text{Broken} \rightarrow \neg\text{Runs}) \land \mathcal{B}(\neg\mathcal{B} \text{Broken} \leftarrow \neg\text{Runs}).$$

Now we can apply the second formula of Proposition 2.3 to the first conjunct (we already know that $\mathcal{B} \text{Broken}$ belongs to the completion). Therefore $\overline{T}$ also contains $\mathcal{B} \text{Runs}$.

We again conclude that the static semantics of $T$ derives $\mathcal{B} \text{Fixed}$, $\text{Broken}$ and $\mathcal{B} \text{Runs}$. Observe that while the above reasoning utilizes only one iteration of the belief closure operator, it is slightly more complex than the one used in Example 2.9 where several iterations were performed. Needless to say, both methods lead to identical results.

Appendix A. Proofs of theorems from Section 3

A.1. Proof of Proposition 3.3

$(\Rightarrow)$ Suppose $\mathcal{I}$ is not a model of $\text{comp}(T)$, so there is a completion axiom $\text{comp}(A, T)$ of an atom $A$ in $T$:
\[ A \leftrightarrow \bigvee_{1 \leq m \leq s} (BG_{1,m} \land \cdots \land BG_{km,m} \land \neg A_{1,m} \land \cdots \land \neg A_{lm,m} \land \neg BF_{1,m} \land \cdots \land \neg BF_{nm,m}), \]

which is not satisfied in \( \mathcal{I} \). If the direction \( \leftarrow \) is not satisfied then \( \mathcal{I} \) obviously is not even a model of \( T \). So assume that this direction is satisfied for all completion axioms, i.e., \( \mathcal{I} \) is a model of \( T \). If the direction \( \rightarrow \) is not satisfied then \( \mathcal{I} \models A \) (consequently, \( A \neq false \)) but for every rule

\[ BG_{1,m} \land \cdots \land BG_{km,m} \rightarrow A \lor A_{1,m} \lor \cdots \lor A_{lm,m} \lor BF_{1,m} \lor \cdots \lor BF_{nm,m}, \]

with \( A \) in the head, one of the \( A_{ij} \) is true or one of the \( BF_{ij} \) is true or one of the \( BG_{ij} \) is false. But then the interpretation \( T' \), which differs from \( \mathcal{I} \) only in \( T' \not\models A \), is also a model of \( T \) and therefore \( \mathcal{I} \) is not minimal.

\((\Leftarrow)\) Suppose conversely that \( \mathcal{I} \) is not a minimal model of \( T \). If \( \mathcal{I} \) is not even a model, i.e., does not satisfy some rule

\[ BG_{1,m} \land \cdots \land BG_{km,m} \rightarrow A \lor A_{1,m} \lor \cdots \lor A_{lm,m} \lor BF_{1,m} \lor \cdots \lor BF_{nm,m}, \]

then obviously the direction \( \leftarrow \) of the completion axioms for every \( A_i \) is also violated. So let us assume that \( \mathcal{I} \) is a model, but not minimal, so there is a smaller model \( \mathcal{I} \). We can choose \( \mathcal{I} \) such a way that it is minimal. Let \( A \) be an objective atom with \( \mathcal{I} \models A \) and \( \mathcal{I} \not\models A \) (consequently, \( A \neq false \)). Let us consider the completion axiom \( comp(A, T) \) for \( A \):

\[ A \leftrightarrow \bigvee_{1 \leq m \leq s} (BG_{1,m} \land \cdots \land BG_{km,m} \land \neg A_{1,m} \land \cdots \land \neg A_{lm,m} \land \neg BF_{1,m} \land \cdots \land \neg BF_{nm,m}), \]

By the contraposition of the already proven opposite direction, the minimal model \( T' \) is a model of \( \text{comp}(T) \), so it satisfies the completion axiom. Since \( T' \not\models A \), for every \( i \) there is a \( j \) such that \( T' \models A_{ij} \) or \( T' \models BF_{ij} \) or \( T' \not\models BG_{ij} \). But since \( T' \) is smaller than \( T \), \( T' \models A_{ij} \Rightarrow \mathcal{I} \models A_{ij} \) and \( T' \models BF_{ij} \Rightarrow \mathcal{I} \models BF_{ij} \) and \( T' \not\models BG_{ij} \Rightarrow \mathcal{I} \not\models BG_{ij} \). Thus, the right hand side is also false in \( \mathcal{I} \), but \( A \) is true in \( \mathcal{I} \), and therefore we conclude that \( \mathcal{I} \not\models \text{comp}(T) \).

A.2. Proof of Proposition 3.4

We will show that every belief theory \( T \) can be transformed into a residual belief theory \( T_{\text{res}} \) by applying just two transformations, namely, “elimination of tautologies” and “unfolding”.

It is easy to see that each application of \( \text{GPPE} \), perhaps combined with tautology elimination, removes a given objective atom \( B \) from the body of one clause without introducing any new clauses containing \( B \) in their bodies. This means that a finite number of such transformations will lead to a belief theory that does not contain \( B \)'s in bodies of its clauses. Applying this procedure to all objective atoms appearing in a (finite) belief theory will therefore produce a residual belief theory.
Accordingly it suffices to establish the following lemma.

**Lemma A.1.** Let $T_1$ and $T_2$ be two belief theories such that $T_2$ results from $T_1$ by an application of GPPE or deletion of tautologies. Then $T_1$ and $T_2$ have the same minimal models.

The deletion of tautologies is an equivalence transformation, so it does not change the set of models, and thus also not the set of minimal models. Now suppose that we apply GPPE to a rule $A \leftarrow B \land R$, where $B$ is an objective atom, and suppose that $B \lor A_i \leftarrow R_i$, $i = 1, \ldots, n$, are all the rules about $B$, i.e., rules containing $B$ in their heads.

Let $I$ be a minimal model of $T$. We first show that it is a model of $T$ (not necessarily minimal). Suppose this is not the case, so the rule $A \leftarrow B \land R$ is violated in $I$ (this is the only rule in $T_1 \setminus T_2$). This means that $I \models B$, $I \models R$, and $I \not\models A$. Since $I$ is a minimal model of $T_2$, the interpretation $I'$ with $I' \not\models B$ (and otherwise equal to $I$) cannot be a model of $T_2$. Since objective atoms do not occur negated in rule bodies, this means that there is a rule with $B$ in the head, i.e., one of the rules $B \lor A_i \leftarrow R_i$, such that $I' \models R_i$ and $I' \not\models A_i$. Since $I' \models R_i$, the atom $B$ cannot be contained in $R_i$, and since we exclude multiple occurrences in the head, it is also not contained in $A_i$. Thus, $I \models R_i$ and $I \not\models A_i$. But now we infer that the new rule $A \lor A_i \leftarrow R \land R_i$ contained in $T_2$ is violated by $I$, which is a contradiction.

**A.3. Proof of Lemma 3.6**

First we prove the following lemma which characterizes minimal models of special belief theories.

**Lemma A.2.** Let $T$ be an arbitrary belief theory and $T_B$ be a set of formulae which contain only belief atoms. Minimal models of $T \cup T_B$ are precisely those minimal models of $T$ which satisfy $T_B$.

**Proof.**

- Suppose that $I$ is a minimal model of $T \cup T_B$. Of course, $I$ is a model of $T$ and of $T_B$. Now suppose that there is a smaller model $I'$ of $T$. Since $I$ and $I'$ do not differ in the

---

3 It is not a new rule $A \lor A_i \leftarrow R \land R_i$, because otherwise $A$ would contain $B$, and the unfolded rule $A \leftarrow B \land R$ would be a tautology and could never be violated.
interpretation of belief atoms, $T'$ is also a model of $T_B$, and thus a model of $T \cup T_B$. But this contradicts the assumed minimality of $\mathcal{I}$.

- Let $\mathcal{I}$ be a minimal model of $T$ which also satisfies $T_B$. Clearly, $\mathcal{I}$ is a model of $T \cup T_B$. Since $\mathcal{I}'$ is also a model of $T$, the existence of a smaller model $\mathcal{I}'$ would contradict the minimality of $\mathcal{I}$. \hfill $\Box$

Now we continue the proof of Lemma 3.6.

**Lemma 3.6.** Let $T$ be an arbitrary finite belief theory and let

$$\mathcal{T} = \text{Cn}_{AEB}(T \cup B\text{comp}(T_{res})),$$

If a formula $F$ is true in all minimal models of $\mathcal{T}$ then $BF$ is already contained in $\mathcal{T}$. In other words, $\Psi_T(\mathcal{T}) \subseteq \mathcal{T}$.

Let $T_B$ be the subset of $\mathcal{T}$ which consists entirely of belief atoms, i.e., $B(\text{comp}(T_{res}))$ together with the axiom CA, all instances of the axiom DA and all formulae forced in by the derivation rule IR. Then the models of $\mathcal{T}$ coincide with the models of $T \cup T_B$ because the required closure under logical consequences does not change the set of models.

By Lemma A.2, minimal models of $T \cup T_B$ are precisely those minimal models of $T$ which satisfy $T_B$. Suppose that $F$ holds in all minimal models of $\mathcal{T}$.

By Proposition 3.3, the minimal models of $T$ coincide with the models of $\text{comp}(T_{res})$ and therefore $F$ holds in all models of $\text{comp}(T_{res}) \cup T_B$, i.e., it is a propositional consequence of $\text{comp}(T_{res}) \cup T_B$. By the compactness of propositional logic, there is a finite subset

$$\{F_1, \ldots, F_n\} \subseteq \text{comp}(T_{res}) \cup T_B$$

such that $\{F_1, \ldots, F_n\} \models F$. But then $F_1 \land \cdots \land F_n \rightarrow F$ is a tautology and therefore, by Proposition 2.3, $B(F_1) \land \cdots \land B(F_n) \rightarrow B(F)$ is contained in $\mathcal{T}$.

If $F_i$ is from $\text{comp}(T_{res})$, then $B(F_i)$ is obviously contained in $\mathcal{T}$ and otherwise we get $B(F_i)$ by necessitation (see Proposition 2.3). Thus, by the closure under classical consequences, we conclude $B(F) \in \mathcal{T}$.

A.4. Proof of the equivalence of the axioms used in the definition of AEB to the original axioms

We show now that the axioms used in Definition 2.1 are completely equivalent to the following ones used in the original definition of autoepistemic belief theories given in [23]:

**(D) Consistency Axiom.** $\neg B\text{false}$.

**(K) Normality Axiom.** $B(F \rightarrow G) \rightarrow (BF \rightarrow BG)$.

**(N) Necessitation Inference Rule.** $\frac{F}{BF}$.
We recall that, in the presence of the axiom K, the axiom D is equivalent to the axiom $BF \rightarrow \neg BF$, stating that if we believe in a formula $F$ then we do not believe in $F$ (see [23] for details).

**Proposition A.3.** The axioms D and K and the necessitation rule N are equivalent to the axioms CA and DA and the inference rule IR.

More precisely, a theory in the language $\mathcal{L}_{\text{AEB}}$ contains all (substitution instances of) the axioms D and K and is closed under both standard propositional consequence and the necessitation rule N if and only if it contains all (substitution instances of) the axioms CA and DA and is closed under both standard propositional consequence and the inference rule IR.

**Proof.** ($\Rightarrow$) Suppose that a theory $T$ contains all (substitution instances of) the axioms D and K and is closed under both standard propositional consequence and the Necessitation Rule N. Since true belongs to $T$, from the Necessitation Rule N it follows that so does $BF$ and thus the axiom CA holds.

To show that the axiom DA holds we first show that for any formulae $F$ and $G : B(F \land G) \rightarrow BF \land BG$ belongs to $T$. Clearly $(F \land G) \rightarrow F \in T$ and therefore, by the Necessitation Rule N, $B((F \land G) \rightarrow F) \in T$. From the Normality Axiom K we infer that $B(F \land G) \rightarrow BF \in T$. Similarly, $B(F \land G) \rightarrow BG \in T$. It follows that $B(F \land G) \rightarrow BF \land BG \in T$.

To show that the axiom DA holds it now suffices to prove that for any formulae $F$ and $G : BF \land BG \rightarrow B(F \land G)$ belongs to $T$. Clearly, $F \rightarrow (G \rightarrow F \land G) \in T$ and therefore, by the Necessitation Rule N,

$B(F \rightarrow (G \rightarrow F \land G)) \in T$.

From the Normality Axiom K we infer that $BF \rightarrow B(G \rightarrow F \land G) \in T$. Applying the Normality Axiom K again we conclude that $BF \rightarrow (BG \rightarrow B(F \land G)) \in T$. This shows that

$BF \land BG \rightarrow B(F \land G) \in T$.

To demonstrate that the inference rule IR holds suppose that $F \leftrightarrow G$ belongs to $T$. By necessitation, also $B(F \leftrightarrow G) \rightarrow T$. From the Normality Axiom it follows immediately that so does $BF \leftrightarrow BG$.

($\Leftarrow$) Suppose that a theory $T$ contains all (substitution instances of) the axioms CA and DA and is closed under both standard propositional consequence and the inference rule IR. Clearly, the axiom D also holds. Since $(F \rightarrow G) \land F \leftrightarrow F \land G$ belongs to $T$, from the rule IR it follows that $B((F \rightarrow G) \land F) \leftrightarrow B(F \land G)$ belongs to $T$. From the axiom DA we conclude that also $B(F \rightarrow G) \land B(F) \leftrightarrow BF \land BG$ belongs to $T$. This proves that $B((F \rightarrow G)) \land B(F) \rightarrow B(G) \in T$ and thus the Normality Axiom K holds.

It remains to show that $T$ is closed under necessitation. If $F$ belongs to $T$ then so does $F \leftrightarrow \text{true}$. By the rule IR, the formula $BF \leftrightarrow B\text{true}$ belongs to $T$ which implies that $BF$ also belongs to $T$. \qed
References


250

S. Brass et al. / Artificial Intelligence 112 (1999) 233–250


