

Belief, Knowledge, Revisions, and a Semantics of Non-Monotonic Reasoning

Ján Šefránek

Institute of Informatics, Comenius University
Mlynská dolina, 842 15 Bratislava, Slovakia
e-mail: sefranek@fmph.uniba.sk

Abstract. Przymusiński's Autoepistemic Logic of Knowledge and Belief (AELKB) is a unifying framework for various non-monotonic formalisms. In this paper we present a semantic characterization of AELKB in terms of Dynamic Kripke Structures (DKS). A DKS is composed of two components – a static one (a Kripke structure) and a dynamic one (a set of transformations). Transformations between possible worlds correspond to hypotheses generation and to revisions. Therefore they enable to define a semantics of insertions to and revisions of AELKB-theories. A computation of the transformations (between possible worlds) is based on (an enhanced) model-checking. The transformations may be used as a method of computing static autoepistemic expansions.

Keywords: non-monotonic reasoning, autoepistemic logic of knowledge and belief, dynamic Kripke structure, belief revision, model checking

1 Introduction

The paper is aiming to present Dynamic Kripke Structures (DKS, [10]) as a rather general tool of a semantic characterization of non-monotonic reasoning. The basic idea is as follows: Consider a non-monotonic inference operator Cn_{nmr} and two sets of sentences A, B , such that $A \subset B$ and $Cn_{nmr}(A) \not\subseteq Cn_{nmr}(B)$. It means that there are sentences ϕ, ψ such that $\phi \in B \setminus A$ and $\psi \in Cn_{nmr}(A) \setminus Cn_{nmr}(B)$ (ϕ represents an insertion into A , ψ represents a revision of $Cn_{nmr}(A)$).

The situation described above can be specified semantically by a pair $(\mathcal{F}, \mathcal{K})$. \mathcal{K} is a Kripke structure, a relation \models_{nmr} is defined over \mathcal{K} and \models_{nmr} is a semantic specification of Cn_{nmr} . \mathcal{F} is a set of transformations on possible worlds of \mathcal{K} : $\mathcal{F} = \{f : W \rightarrow W\}$. Consider $w \in W, f(w) \in W$ such that $w \not\models_{nmr} \phi$, but $f(w) \models_{nmr} \phi$. It means that f corresponds to an insertion. Similarly, let us assume that $w \models_{nmr} \psi$ and $f(w) \not\models_{nmr} \psi$. Therefore, f corresponds to a revision.

Let us summarize the basic intuition: a DKS consists of two components – a static one (a Kripke structure) and a dynamic one (a set of transformations between possible worlds). The situations when a new knowledge is acquired and

– as a consequence – a piece of knowledge (accepted before) should be revised are crucial from the non-monotonic reasoning point of view. DKS provide a semantic characterization of these situations. A transformation of one possible world to another represents a change in our knowledge. The transformation is defined on a set of possible worlds and the set of possible worlds produced by the transformation represents the sets of epistemic alternatives after the transformation (after some insertions and some revisions forced by the insertions).

A technical core of the paper is a semantic characterization of Przymusiński’s Autoepistemic Logic of Knowledge and Belief [9] (AELKB) in terms of DKS. Moreover, DKS provides also a semantics of revisions (of knowledge and belief theories). A framework for belief revision of knowledge and belief theories was presented in [1].

Przymusiński augmented Moore’s autoepistemic logic (employing the knowledge operator K) with an additional belief operator B .¹ Przymusiński’s extension of AELK to AELKB reflects an intuition that besides reasoning about statements which are known to be true we also need to reason about those statements that are only believed to be true. The semantics of B operator is determined by minimal entailment (or more general by a non-monotonic entailment). Expressibility is a strong point of AELKB: AELKB is a unifying framework for several major nonmonotonic formalisms [9]. Therefore, a semantics of AELKB in terms of DKS supports the ambition to use DKS as a tool of a general semantic characterization of non-monotonic reasoning (with an incorporated belief revision).

The paper is organized as follows: First we describe the language and the basic concepts of AELKB (Sections 2 and 3). Thereafter in Section 4 we define DKS. In Section 5 are reviewed known results of [8] and [2] concerning characterizations of AELK and AELB in terms of Kripke structures. The results of this paper are presented in Section 6 (a possible-world semantics of AELKB), in Section 7 (insertions into knowledge and belief theories are characterized in terms of DKS and it is outlined how to compute static autoepistemic expansions and how to use model-checking as a method of computing transformations between possible worlds), and in Section 8 (a semantic specification of revisions of AELKB-theories and their computation by enhanced model checking is presented; the computation uses an idea of [3]).

2 Preliminaries

We assume a fixed propositional language \mathcal{L} with standard connectives ($\neg, \Rightarrow, \wedge, \dots$), a countable set of propositional letters $\mathcal{P} = \{p_1, \dots, p_n, \dots\}$ and a special propositional letter \perp denoting *false*. Propositional atom, literal, and formula are defined as usually.

¹ We will use abbreviations *AELKB*, *AELK*, *AELB* for the logics employing both operators, only K , and only B , respectively. There is a little difference between here introduced symbols/abbreviations and the usual usage.

Let \mathcal{L}_A , an extension of \mathcal{L} , be defined as follows: Two (modal) operators K and B are added to the set of symbols. Each atom, literal, and formula of \mathcal{L} is an (objective) atom, literal, and formula of \mathcal{L}_A , respectively. If ϕ is a formula of \mathcal{L}_A , then $B\phi$, $K\phi$ are (subjective) atoms, and $B\phi$, $K\phi$, $\neg B\phi$ and $\neg K\phi$ are (subjective) literals of \mathcal{L}_A . Each (subjective) literal of \mathcal{L}_A is a *formula* of \mathcal{L}_A . If ϕ and ψ are formulae of \mathcal{L}_A , then $\phi \wedge \psi$, $\neg\phi$ are formulae of \mathcal{L}_A (\mathcal{L}_A -formulae). The formulae which contain K or B operators, are called subjective formulae.

Definition 1 (Knowledge and belief theory, [9]) A knowledge and belief theory in \mathcal{L}_A (*AELKB-theory*) is a (possibly infinite) set of formulae of the form

$$\beta_1 \wedge \cdots \wedge \beta_k \wedge B\phi_1 \wedge \cdots \wedge B\phi_l \wedge K\psi_1 \wedge \cdots \wedge K\psi_m \Rightarrow \alpha_1 \vee \cdots \vee \alpha_n \vee B\chi_1 \vee \cdots \vee B\chi_r \vee K\tau_1 \vee \cdots \vee K\tau_s,$$

where α_i s, β_i s are propositional atoms, ϕ_i s, ψ_i s, χ_i s, τ_i s are arbitrary formulae of \mathcal{L}_A . \square

Let us denote by $\mathcal{P}_{\mathcal{L}_A}$ the set of all atoms of \mathcal{L}_A . An *interpretation* of \mathcal{L}_A is a subset of $\mathcal{P}_{\mathcal{L}_A}$.

It is clear that a valuation of an \mathcal{L}_A -formula in an interpretation \mathcal{I} may be defined precisely as the two-valued propositional valuation:

Definition 2 Let \mathcal{I} be an interpretation:

- if ϕ is an atom (objective or subjective) of \mathcal{L}_A , then $\text{val}_{\mathcal{I}}(\phi) = 1$ iff $\phi \in \mathcal{I}$,
- if ϕ is a literal $\neg\psi$, then $\text{val}_{\mathcal{I}}(\phi) = 1$ iff $\psi \notin \mathcal{I}$,
- otherwise ϕ is a boolean combination of literals and $\text{val}_{\mathcal{I}}(\phi)$ is computed according to the rules for boolean combinations.

If X is a set of formulae, then $\text{val}_{\mathcal{I}}(X) = 1$ iff $\text{val}_{\mathcal{I}}(\phi) = 1$ for each $\phi \in X$ and we say that \mathcal{I} is a model of X (X is satisfied in \mathcal{I}). \square

A convention: We will sometimes use an alternative notation for interpretations. If \mathcal{I} is an interpretation of \mathcal{L}_A , it can be denoted by $\mathcal{I} \cup \mathcal{N}$, where $\mathcal{N} = \{\neg\phi : \phi \in \mathcal{P}_{\mathcal{L}_A} \setminus \mathcal{I}\}$.

Definition 3 Let us consider interpretations \mathcal{I}, \mathcal{J} , which coincide on subjective literals. $\mathcal{I} \prec \mathcal{J}$ iff for each objective atom α holds: if $\alpha \in \mathcal{I}$, then $\alpha \in \mathcal{J}$.

Let Σ be a set of interpretations and $\mathcal{I} \in \Sigma$. Then \mathcal{I} is minimal in Σ iff there is no $\mathcal{J} \in \Sigma$ such that $\mathcal{J} \neq \mathcal{I}$ and $\mathcal{J} \prec \mathcal{I}$.

If a formula ϕ is true in all minimal models of a knowledge and belief theory T then we say that ϕ is minimally entailed by T (notation: $T \models_{\text{min}} \phi$). \square

3 Static Autoepistemic Expansions

Truth values of the subjective atoms are independent on the truth values of their arguments. Intuitively, subjective atoms are true only if their arguments are known or believed. An evidence of what is known and/or believed we can represent by a set of subjective atoms (belief set).²

² Later we will use a more general notion of belief set. A decoupling of subjective and objective literals was used in the Definition 3 of minimal interpretations.

Definition 4 Let \mathcal{I} be an interpretation of propositional letters and S be a set of subjective atoms. We define a function val which assigns a value from the set $\{0, 1\}$ to each pair (\mathcal{I}, S) and each \mathcal{L}_A formula:

- if ϕ is an objective atom, then $val_{\mathcal{I}}^S(\phi) = 1$ iff $\phi \in \mathcal{I}$
- if ϕ is a subjective atom, then $val_{\mathcal{I}}^S(\phi) = 1$ iff $\phi \in S$
- if ϕ is $\neg\psi$, then $val_{\mathcal{I}}^S(\phi) = 1$ iff $val_{\mathcal{I}}^S(\psi) = 0$
- if ϕ is $\psi \wedge \tau$, then $val_{\mathcal{I}}^S(\phi) = 1$ iff $val_{\mathcal{I}}^S(\psi) = 1$ and $val_{\mathcal{I}}^S(\tau) = 1$

Let S be fixed. We define $val^S(\tau) = 1$ iff for each \mathcal{I} is $val_{\mathcal{I}}^S(\tau) = 1$.

If X is a set of \mathcal{L}_A -formulae, then $val_{\mathcal{I}}^S(X) = 1$ iff $val_{\mathcal{I}}^S(\phi) = 1$ for each $\phi \in X$ and we say that \mathcal{I} is a model of X (X is satisfied in \mathcal{I}). \square

We will use repeatedly the scheme from the Definition 4 in the following. The only point of difference will be how the set S is specified.

We do not intend to use arbitrary belief sets. It is appropriate to restrict somehow possible belief sets (a belief set should be a reasonable one). There is a variety of possibilities for a decision, some of them are used in the paper.

Definition 5 (Formulae derivable from an AELKB-theory, [9]) Let T be an AELKB-theory. We denote by $Cn_A(T)$ the smallest set of formulae which contains the theory T , and all instances of:

Consistency Axiom $\neg B \perp$

Normality axiom $B(\phi \Rightarrow \psi) \Rightarrow (B\phi \Rightarrow B\psi)$

and is closed under propositional consequence and under Necessitation Inference

Rule $\frac{\phi}{B\phi}$ \square

A consequence operator is a function which assigns a set of formulae to a set of formulae. We will use two consequence operators: Cn_A and Cn_{PL} (the propositional consequence operator). Each set of formulae derivable from an AELKB-theory is – in a sense – a reasonable belief set.

An “introspective content” of an AELKB-theory T can be viewed as an AELKB-theory T^* , called static autoepistemic expansion.

Definition 6 (Static autoepistemic expansion) A theory T^* is called a static autoepistemic expansion (SAE) of a knowledge and belief theory T iff $T^* = Cn_A(T \cup \{K\phi : T^* \models \phi\} \cup \{\neg K\phi : T^* \not\models \phi\} \cup \{B\phi : T^* \models_{min} \phi\})$ \square

Notice that we distinguish three levels of a logical characterization of AELKB-theories:

- two-valued models (and Cn_{PL} -consequence)
- Cn_A -consequence
- static autoepistemic expansion

4 Dynamic Kripke Structures

We can now proceed to the central semantic construction used in this paper.

First a rather general concept of Kripke structures is defined. (Later we will use some of its specializations.)

Definition 7 *Kripke structure is a triple (W, R, m) , where W is a set of possible worlds, $R = \{\rho : \rho \subseteq W \times W\}$ is a set of accessibility relations and m is a (meaning) function assigning to each possible world an interpretation. \square*

Definition 8 *A monoid is a triple (M, \circ, e) , where M is a set, $\circ : M \times M \rightarrow M$ is an associative operation, $e \in M$ and for every $x \in M$ holds $e \circ x = x = x \circ e$. \square*

We are ready to define DKS. The structure consists of a monoid-part and a Kripke-structure-part. The main idea is a transformation of possible worlds to possible worlds. The transformation is specified by monoid elements.

Definition 9 *Dynamic Kripke Structure is a pair $(\mathcal{M}, \mathcal{W})$, where \mathcal{M} is a monoid and \mathcal{W} is a Kripke structure, and for every $x \in M$ there is a function³ $f_x : W \rightarrow W$ such that f_e is an identity mapping and for every $x, y \in M$, for every $w \in W$ holds $f_{x \circ y}(w) = f_x(f_y(w))$ \square*

Dynamic Kripke structures were introduced in [10] together with a demonstration that database updates and Closed World Assumption are expressible in terms of DKS.

A motivation (and an ambition) behind the concept is that it seems that DKS provide a useful tool for a (unifying) semantic characterization of non-monotonic reasoning. The proposed approach is based on a belief that non-monotony is a consequence of some fundamental properties⁴ of hypothetical and context-dependent reasoning, and of belief revision. A close relationship between belief revision and inference is emphasized.

The most significant feature of DKS are transformations between possible worlds. The transformations correspond intuitively to hypotheses generation and to revisions (a hypothesis may be true in the image-world, but not in the source-world and vice versa). Sometimes the accessibility relation is “changed” by a transformation (more precisely – for worlds w_1, w_2 , accessibility relation ρ and transformation f may hold: $(w_1, w_2) \in \rho$, $(f(w_1), f(w_2)) \notin \rho$ or vice versa). Therefore, if a consequence operator Cn is dependent on the accessibility relation, then a transformation results in a non-monotonic Cn . In a sense, DKS is a construction explaining the non-monotony of reasoning. From the DKS point of view: If non-monotony is a symptom, then hypotheses addition and revisions forced by the addition are the essence (of the non-standard, hypothetical, context-dependent reasoning).

³ It is said that there is an *action* of \mathcal{M} on W .

⁴ “... non-monotonic behaviour ... is a *symptom*, rather than the essence of non-standard inference”, see [11].

5 Possible World Semantics

A characterization of AELK in terms of Kripke structures was given by Moore, see [8]. Similarly, Kripke structures were used as a tool of a characterization of AELB in [2]. In this section we summarize the results of [8] and [2], particularly a characterization of SAE of AELK- and AELB-theories in terms of Kripke structures.

Let us restrict the language \mathcal{L}_A in such a way that we do not use belief atoms (knowledge atoms) of the form $B\phi$ ($K\phi$). The language we denote by \mathcal{L}_{AK} (\mathcal{L}_{AB}).⁵ The formulae of both languages we will denote as \mathcal{L}_{AK} - (\mathcal{L}_{AB} -) formulae.

Definition 10 *An AELK (AELB)-theory T_{AK} (T_{AB}) in \mathcal{L}_{AK} (\mathcal{L}_{AB}) is a K - (B -) restriction of an AELKB-theory T iff $T_{AK} = \{\phi \in T : \phi \text{ is a } \mathcal{L}_{AK}\text{-formula}\}$ ($T_{AB} = \{\phi \in T : \phi \text{ is a } \mathcal{L}_{AB}\text{-formula}\}$). \square*

5.1 Possible World Semantics for AELK

Definition 11 *A complete S5-frame is a Kripke structure (W, ρ) such that $\rho = W \times W$.⁶ \square*

Each possible world is accessible from each possible world in a complete S5-frame and a complete S5-frame is uniquely determined by the set of possible worlds W .

Definition 12 *A set S of \mathcal{L}_{AK} -formulae is stable iff*

- $S = Cn_{PL}(S)$
- if $\phi \in S$, then $K\phi \in S$
- if $\phi \notin S$, then $\neg K\phi \in S$

\square

We now introduce a specialization of the Definition 4.

Let M be a complete S5-frame. Let $w \in W$ be an interpretation of propositional letters (as introduced in the Definition 2). We will use a function val_w^S as defined in the Definition 4, but $S = \{K\phi : (\forall w \in W) val_w(\phi) = 1\} \cup \{\neg K\phi : (\exists w \in W) val_w(\phi) = 0\}$. Note that val_w^M we use as a synonym of val_w^S .

Let us recall that a formula ϕ is true in a complete S5-frame M , if for each $w \in W$ is $val_w^S(\phi) = 1$; notation: $val^S(\phi) = 1$ or alternatively $val^M(\phi) = 1$.

Theorem 1 ([8], [7]) *A set of \mathcal{L}_{AK} -formulae S is stable iff S is the set of all \mathcal{L}_{AK} -formulae which are true in some complete S5-frame. \square*

⁵ It means, $\mathcal{L}_{AK} = \{\phi \in \mathcal{L}_A : B \text{ operator does not occur in } \phi\}$. Similarly for \mathcal{L}_{AB} .

⁶ This is our first special case of Kripke structures. For simplicity, we use the symbol ρ instead of $\{\rho\}$ and we identify the set of possible worlds with a set of interpretations – possible worlds are interpretations. (Formally, function m is the identity, but we are omitting an explicit recording of this function.)

We can now define an interpretation consisting of two components – one is an ordinary propositional interpretation, the second is a complete $S5$ -frame (a reasonable belief set is a set of all formulae satisfied in a complete $S5$ -frame).

Definition 13 A possible-world autoepistemic interpretation is a pair $PW = (\mathcal{I}, M)$, where \mathcal{I} is an ordinary interpretation of propositional letters of \mathcal{L} and M is a complete $S5$ -frame. \square

Possible-world model is defined in an obvious way.

Definition 14 Let X be a set of formulae, ϕ a formula. $X \models_{PW} \phi$ iff ϕ is true in every possible-world model of X . \square

We are now able to express a characterization of SAE of AELK-theories in terms of possible-world interpretations.

Theorem 2 ([7]) Let T be an AELK-theory. A set S of \mathcal{L}_{AK} -formulae is a K -restriction of a SAE of T iff $S = \{\phi : (T \cup \{K\psi : \psi \in S_0\} \cup \{\neg K\psi : \psi \in \mathcal{L}_0 \setminus S_0\}) \models_{PW} \phi\}$, where S_0 is the set of all objective formulae from S and \mathcal{L}_0 is the set of all objective formulae from \mathcal{L}_{AK} . \square

5.2 Possible World Semantics for AELB

B -restrictions of SAE can be also characterized in terms of Kripke structures. The result is due to [2].

Let \mathcal{K} be a Kripke structure (W, ρ) , where W is a set of propositional interpretations (a set of sets of objective literals). Functions val_w^S and val^S are defined as above and for each $w \in W$ is $S = \{B\phi : \exists w' ((w, w') \in \rho \wedge val_{w'}(\phi) = 1)\}$. \square We will write also $val^{\mathcal{K}}$ and $val_w^{\mathcal{K}}$ instead of val^S and val_w^S .

Theorem 3 ([2]) Let T be an arbitrary AELB-theory and (W, ρ) be a Kripke structure satisfying

- for every $w \in W$ there is $w' \in W$ such that $w \rho w'$
- each $w \in W$ is a model of T
- for all $w, w' \in W$ such that $w \rho w'$ holds that w' is a minimal model of T

Then $T^* = \{\phi \in \mathcal{L}_{AB} : (\forall w \in W) val_w^{\mathcal{K}}(\phi) = 1\}$ is a B -restriction of a SAE of T . \square

6 AELKB-structures

We are now ready to construct an appropriate Kripke structure which enables a characterization of SAE of (full) AELKB-theories. The possible worlds of our Kripke structures are complete $S5$ -frames and one of the accessibility relations leads to minimal models.

In what follows we assume only a language \mathcal{L}_A with a finite set of propositional letters and *finite* sets of *finite* interpretations.⁷ There are two reasons for the limitation to the finite structures.

First, we are interested in a correspondence between sets of models (possible worlds) and AELKB-theories (sets of all \mathcal{L}_A -formulae true in the given possible world). But there is a countable set M of propositional models⁸ such that there is no set S of \mathcal{L}_A -formulae such that M is the set of all models of S , see [4]. Only for finite sets of propositional models holds: if w is a (finite) set of models, then there is a set S of \mathcal{L}_A -formulae such that w is the set of all models of S .

Second, we propose model checking as a computational method for DKS, therefore the limitation to finite structures is a natural one.

Definition 15 (AELKB-structure) *Let Int be a set of all interpretations of an AELKB-theory in a language \mathcal{L}_A .*

AELKB-structure is a triple (W, R, m) , where $W = \mathcal{P}(Int)$ is the set of all subsets of Int , $R = \{\rho_1, \rho_2\}$, $\rho_1 = \{(w, w') : w \neq w' \wedge (\exists \mathcal{I} \in Int) w = w' \cup \{\mathcal{I}\}\}$, $\rho_2 = \{(w, w') : w' = \{\mathcal{I} : \mathcal{I} \text{ is minimal in } w\}\}$. Finally, m is defined as follows:⁹

- *for an objective formula ϕ : $m_w(\phi) = 1$, if $(\forall \mathcal{I} \in w) \text{val}_{\mathcal{I}}(\phi) = 1$, $m_w(\phi) = 0$ if $(\forall \mathcal{I} \in w) \text{val}_{\mathcal{I}}(\phi) = 0$, otherwise $m_w(\phi) = \frac{1}{2}$*
- *$m_w(K\phi) = 1$ iff $m_w(\phi) = 1$*
- *$m_w(\neg K\phi) = 1$ iff $m_w(\phi) \neq 1$*
- *$m_w(B\phi) = 1$ iff $(w, w') \in \rho_2 \rightarrow m_{w'}(\phi) = 1$, otherwise $m_w(B\phi) = 0$*
- *if ϕ and ψ are \mathcal{L}_A -formulae, then $m_w(\neg\phi) = 1 - m_w(\phi)$ and $m_w(\phi \wedge \psi) = \min\{m_w(\phi), m_w(\psi)\}$.*

If T is a knowledge and belief theory, then $m_w(T) = 1$ iff $(\forall \phi \in T) m_w(\phi) = 1$.
□

Note that the three-valued valuation of objective formulae was defined. We motivate the decision as follows: Each consistent SAE of an arbitrary AELKB-theory T contains exactly one of the complementary literals $K\phi, \neg K\phi$ for each \mathcal{L}_A -formula ϕ . Therefore, we have to define m_w in such a way that for each formula ϕ holds either $K\phi$ or $\neg K\phi$. However, if neither ϕ nor $\neg\phi$ is true in each interpretation of w , then it is natural to accept both $m_w(\neg K\phi) = 1$ and $m_w(\neg K\neg\phi) = 1$. It means that we have to introduce the third truth-value. Two-valued valuations are used for subjective formulae.

Notation: Let T be an AELKB-theory and w be a set of models. We denote by $Mod(T)$ the set of all models of T and by $Th(w)$ the set of all formulae true in each model of w . Obviously, $Th(w) = Cn_A(Th(w))$, $T = Th(Mod(T))$, and $w = Mod(Th(w))$.

⁷ We consider only relevant interpretations.

⁸ Note that we use the concept of two-valued interpretations (models) as defined in the Definition 2.

⁹ m assigns an interpretation $m(w)$ to each possible world w . An application of the interpretation $m(w)$ to a formula ϕ we will denote by $m_w(\phi)$.

Theorem 4 Let T be an AELKB-theory, $\mathcal{K} = (W, \{\rho_1, \rho_2\}, m)$ be an AELKB-structure and $w_T \in W$ be the set of all models of T . Let w_\perp be the empty set of interpretations.

For each possible world $w' \in W$ such that $w_\perp \subseteq w' \subseteq w_T$ holds that the set of formulae $T^* = \{\phi : m_{w'}(\phi) = 1\}$ is a SAE of T .

Proof Sketch : First we prove that $T^* = \{\phi : m_{w'}(\phi) = 1\} \subseteq Cn_A(T \cup \{K\phi : T^* \models \phi\} \cup \{\neg K\phi : T^* \not\models \phi\} \cup \{B\phi : T^* \models_{min} \phi\})$.

Let be $m_{w'}(\phi) = 1$. If ϕ is of the form $K\psi$, then $m_{w'}(\psi) = 1$. It means that each model of T^* is a model of ψ . Therefore $\phi \in \{K\tau : T^* \models \tau\}$. Similarly for $\phi = \neg K\psi$ and $\phi = B\psi$. The closure of T^* under Cn_A is obvious. Finally, each subset of w_T satisfies T , i.e. also w' satisfies T . As a consequence, $T \subseteq T^*$. Obviously, for each objective formula ϕ such that $m_{w'}(\phi) = 1$ holds that $\phi \in Cn_A(T \cup \{K\phi : T^* \models \phi\} \cup \{\neg K\phi : T^* \not\models \phi\} \cup \{B\phi : T^* \models_{min} \phi\})$. It means, T^* is a subset of a SAE.

Conversely, let us assume $\phi \in Cn_A(T \cup \{K\phi : T^* \models \phi\} \cup \{\neg K\phi : T^* \not\models \phi\} \cup \{B\phi : T^* \models_{min} \phi\})$. It is straightforward to show that $m_{w'}(\phi) = 1$. \square

Of course, if T^* is a consistent SAE of T , then $w_\perp \subset w' \subseteq w_T$.

The theorem provides an existential characterization of AELKB-theories (and their SAE) in terms of Kripke structures (AELKB-structures). But the crucial question – which possible worlds determine SAE (which possible worlds are sets of all models of some SAE of an AELKB-theory T) is open.

Our next goal is to present a more constructive method of SAE characterization.

Let w_T be the set of all models of an AELKB-theory T . Consider two sets of formulae: $S = \{\phi : m_{w_T}(\phi) = 1\}$ and $Cn_A(T)$. The next example shows that $S \setminus Cn_A(T) \neq \emptyset$ for some AELKB-theories.

Example 1 ([9]) Let T be $\{B\neg b \wedge B\neg f \Rightarrow r, \neg Kb \wedge \neg Kf \Rightarrow d\}$.

Some of the members of w_T are

$$\nu_1 = \{B\neg b, B\neg f, \neg Kb, \neg Kf, r, d, b, f\}$$

$$\nu_2 = \{\neg B\neg b, \neg B\neg f, Kb, Kf, \neg r, \neg d, \neg b, \neg f\}.$$

Hence, $m_{w_T}(\neg Kb) = 1$ and $m_{w_T}(\neg Kf) = 1$, but $\neg Kb \notin Cn_A(T)$ and $\neg Kf \notin Cn_A(T)$.

Let w'_T be the set of all minimal models of T . If $\nu \in w'_T$, then $\neg b \in \nu$ and $\neg f \in \nu$. Therefore, $m_{w_T}(B\neg b) = 1 = m_{w_T}(B\neg f)$, but $B\neg b, B\neg f \notin Cn_A(T)$. \square

A non-empty $S \setminus Cn_A(T)$ may contain literals of two forms: $\neg K\phi$ or $B\phi$. Intuitively, the function m_w generates two kinds of (defeasible) hypotheses (the sentences which do not belong among Cn_A -consequences of T): belief formulae and introspective formulae stating that something is not known. It remains to show that we can provide a more constructive method of SAE characterization.

Next we define a monotonic mapping of a complete lattice.

Theorem 5 Let be $W = \mathcal{P}(Int)$. Then the mapping $\Phi : W \rightarrow W$ defined as $\Phi(w) = Mod(\{\phi : m_w(\phi) = 1\})$ is monotonic.¹⁰

Proof: If $w \subseteq w'$, then $\{\phi : m_w(\phi) = 1\} \supseteq \{\phi : m_{w'}(\phi) = 1\}$, i.e. $Mod(\{\phi : m_w(\phi) = 1\}) \subseteq Mod(\{\phi : m_{w'}(\phi) = 1\})$.

Remark 1 From the monotony follows that Φ has a least fixpoint and a greatest fixpoint.

We are now able to give a more deep characterization of SAE.

Theorem 6 Let T be an AELKB-theory, $\mathcal{K} = (W, \{\rho_1, \rho_2\}, m)$ be an AELKB-structure and $w_T \in W$ be the set of all models of T .

Then for each possible world $w \in W$, where $w_\perp \subseteq w \subseteq w_T$ holds: if $\Phi(w) = w$, then $Cn_A(Th(w))$ is a SAE of T (we will say that w determines a SAE of T).

There is a naive (and inefficient) method of verifying whether some possible world w determines a SAE of T .

Definition 16 Let w_0, \dots, w_k be a sequence of possible worlds such that for each $i = 0, \dots, k - 1$ holds $(w_i, w_{i+1}) \in \rho_1$. We say that the sequence is a ρ_1 -path.

Obviously, for each pair (w_i, w_j) such that $i < j$ holds that $w_j \subseteq w_i$.

The method consists in searching all ρ_1 -paths and for each w on a ρ_1 -path checking if $\{\phi : m_w(\phi) = 1\}$ is satisfied in w .

A more promising method consists in (non-deterministic) selecting some formulae from the set $S \setminus Cn_A(T)$, inserting them to T and verifying if the insertion leads to a SAE of T .

In simple cases the first attempt is a successful one:

Example 2 Let us return to the Example 1. $\neg Kb, \neg Kf, B\neg b, B\neg f \in S \setminus Cn_A(T)$. If $T' = T \cup \{\neg Kb, \neg Kf, B\neg b, B\neg f\}$ and $w_{T'}$ is the set of all models of T' , then $w_{T'}$ is a fixpoint of Φ , hence $Cn_A(Th(w_{T'}))$ is a SAE of T .

In general, some iteration of insertions is needed. A recursive procedure we outline later.

We have seen that a computation of SAE consists in some insertions to T and checking if a possible world, the set of all models of the extended theory, is a fixed point of Φ .

We are now motivated to study insertions into AELKB-theories. Moreover, a semantic characterization of insertions is interesting in its own right: insertions exhibit the non-monotonic features of autoepistemic theories (or more generally – of each knowledge representation framework).

¹⁰ There is a relation between Φ and the belief closure operator Ψ_T of [9]. A forthcoming paper devoted to a more detailed study of computational aspects will discuss the relation.

7 Dynamic AELKB-structures

In this Section we provide a characterization of insertions into AELKB-theories in terms of dynamic Kripke structures.

Let us begin with a continuation of the example 1:

Example 3 Let T be again $\{B\neg b \wedge B\neg f \Rightarrow r, \neg Kb \wedge \neg Kf \Rightarrow d\}$.

Let us insert into T a formula $b \vee f$, i.e. $T_2 = T \cup \{b \vee f\}$. If w_T is the set of all models of T and $w_{T'}$ is the set of all models of T' , then $w_{T'}$ does not contain the models from w_T with both $\neg b$ and $\neg f$. Therefore also the set of minimal models of T' is changed and corresponding B -consequences, too.

The change may be specified by a transformation. If w is a possible world, then $F_{b \vee f}(w)$ is a possible world $w' = \{m \in w : m(b \vee f) = 1\}$, i.e. $F_{b \vee f}(w)$ is the set of all models from w which satisfy $b \vee f$ (obviously, if there is no such model, then $F_{b \vee f}(w) = w_\perp$). \square

Our characterization of insertions in terms of dynamic Kripke structures is based on some well known relations between sets of models and sets of formulae. The relations provide – in a sense – also a connection between insertions into some theories and corresponding models. They are expressed by the following facts:

Fact 1 Let T be an AELKB-theory and $w = \text{Mod}(T)$. Let w' be a set of models and $w \supseteq w'$.

Then $w' = \text{Mod}(T \cup T')$, where T' is a set of \mathcal{L}_A -formulae.

Therefore, a function f defined on W such that $f(w) \subseteq w$ is a promising candidate of an appropriate transformation of DKS.

Fact 2 Let T, T' be AELKB-theories such that $T \subset T'$. If $w = \text{Mod}(T)$ and $w' = \text{Mod}(T')$, then there is a ρ_1 -path from w to w' .

We can now propose a DKS: a transition from a possible world to another possible world should correspond to insertion of formulae into theories (and vice versa).

Our next goal is to define an appropriate monoid (and corresponding transformations). The basic intuitions: U , a set of insertions, we will represent by a set of formulae. We propose U as a monoid: a concatenation of two insertions is an insertion, the concatenation of insertions is associative, further, an insertion of no proposition plays the role of the unit (of the monoid). To each monoid member is assigned a mapping from possible worlds to possible worlds (see Example 3).

Let w be a set of interpretations and $f_u(w) = w'$ for some u . We need the transformation defined in a unique way: if $u \equiv v$, then should be $f_u(w) = f_v(w)$ for each $w \in W$.

Definition 17 Let u be an \mathcal{L}_A -formula and $[u]_{\equiv} = \{x \in \mathcal{L}_A : x \equiv u\}$. We assume a selection function σ that assigns to each $[u]_{\equiv}$ exactly one representative.

Definition 18 (i-monoid) Let $U = \{u : \exists [u]_{\equiv} \sigma([u]_{\equiv}) = u\}$ be a set of representatives. We define a monoid (called i-monoid) \mathcal{U} over U : For $u, v \in U$ be $u \circ v = \sigma([u \wedge v]_{\equiv})$.

Clearly, the operation \circ is associative and the empty formula plays the role of the monoid unit, $u \circ \epsilon = u = \epsilon \circ u$ for each $u \in U$. By a convention we may consider ϵ as the representative of the class of all propositional tautologies.

Definition 19 (Dynamic AELKB-Structure) Dynamic AELKB-Structure is a pair $(\mathcal{U}, \mathcal{K})$, where $\mathcal{K} = (W, \{\rho_1, \rho_2\}, m)$ is an AELKB-structure and \mathcal{U} is an i-monoid.

An action of the monoid \mathcal{U} on W is defined as follows: for $u \in U$ is $f_u(w) = w' = \text{Mod}(Cn_A(\text{Th}(w) \cup \{u\}))$.

Of course, w' is the (unique) value of $f_u(w)$:

Fact 3 Let T be an AELKB-theory and $\mathcal{K} = (W, \{\rho_1, \rho_2\}, m)$ be an AELKB-structure. Let $w \in W$ be the set of all models of T and $T' = T \cup \{u\}$.

Then there is in W exactly one w' such that $w' = \text{Mod}(Cn_A(T'))$.

Fact 4 Let a dynamic AELKB-structure be given. It holds:

- $f_{\epsilon}(w) = w$
- $f_{u \circ v}(w) = f_u(f_v(w))$
- if $f_{u \circ v}(w) = w'$, then $\text{Th}(w') = Cn_A(\text{Th}(w) \cup \{u \wedge v\}) = Cn_A(\text{Th}(w) \cup \{u\} \cup \{v\})$

We are ready to outline an insertion-based procedure for computing SAE. Let an AELKB-theory T and a corresponding dynamic AELKB-structure \mathcal{K} are given. Let $w_T \in W$ be the set of all models of T .

- select a hypothesis h from $S \setminus Cn_A(T)$
- compute $f_h(w_T) = w'$
- if w' is a fixpoint of Φ , then return the computed SAE (and search for another SAE), else select a hypothesis h' from $S' \setminus Cn_A(T')$, where $S' = \{\phi : m_{w'}(\phi) = 1\}$ and $T' = Cn_A(T \cup \{h\})$,¹¹ continue the (recursive) computation

Remark 2 A backtracking is assumed – it may be useful to revise the initial selection. For example, a premature selection of formulae of the form $\neg K\phi$ leads sometimes to a direct construction of an inconsistent SAE.

It remains to show that the computation of f_h may be based on model checking.

Let an AELKB-theory T and a possible world $w_T = \text{Mod}(T)$ be given. We can use (an adaptation of) model checking algorithm of [5]¹² in order to compute the value of $f_h(w_T)$.

¹¹ If $S' \setminus Cn_A(T') = \emptyset$, then w' is a fixpoint of Φ .

¹² Symbolic model checking may be used in real applications.

We search through all ρ_1 -paths (breadth-first search is necessary) until we find a possible world w such that for each $\mathcal{I} \in w$ is $\mathcal{I}(h) = 1$.¹³ Therefore, $f_h(w_T) = w$ and $Cn_A(Th(w)) = Cn_A(T \cup \{h\})$.

8 Revisions

Finally, we give a characterization of revisions in terms of DKS.

The power of AELKB (more precisely, of AELB) is demonstrated also by a belief revision framework presented in [1]. We try to use also DKS as a tool of revisions specification and computation. We also compare the reached results with the results of [1]. In what follows we assume only AELB-theories.

Example 4 ([1]) *Let be $T = \{B \neg broken \Rightarrow runs\}$. The set of all models of T is $w = \{\{B \neg b, r, b\}, \{B \neg b, r, \neg b\}, \{\neg B \neg b, r, b\}, \{\neg B \neg b, r, \neg b\}, \{\neg B \neg b, \neg r, b\}, \{\neg B \neg b, \neg r, \neg b\}\}$.*

Let $u = \{\neg runs\}$. It holds that $f_u(w) = w'$, where $w' = \{\{\neg B \neg b, \neg r, b\}, \{\neg B \neg b, \neg r, \neg b\}\}$ is the set of all models of $T' = T \cup \{\neg runs\}$. The set of minimal models is $w_{min} = \{\{\neg B \neg b, \neg r, \neg b\}\}$. Hence, T'^ , the only SAE of T' is inconsistent: $T' \models_{min} \neg broken, B \neg broken \in T'^*, T'^* \models \neg runs \wedge runs$.*

We have seen that the semantics of minimal models has some undesirable consequences in a context of incomplete information. In our example the inconsistency was caused by the hypothesis $B \neg broken$. The hypothesis is a member of a SAE (defined in the standard way). It seems that we need a modified – as compared with SAE and minimal entailment – idea of reasonable hypotheses. Our proposal consists in a reconstruction of the given dynamic AELKB-structure. The SAE of the reconstructed structure satisfies our intuitive requirements.

Example 5 *Consider a reason of inconsistency observed in the Example 4. The minimal model $\{\neg B \neg b, \neg r, \neg b\}$ is in a sense a pathological one. It contains the pair $(\neg B \neg b, \neg b)$ – let us call it a gang (according to [6]) – with a potential conflict between claiming $\neg b$ and disbelieving $\neg b$. We repair the pathology using a technique of [3]. The essence of the technique is a modification of the accessibility relation ρ_2 . The modification consists in a removal of the pair (w', w_{min}) from ρ_2 and an insertion of an improvement of the pair to ρ_2 . The goal of the improvement is a minimization of undesirable consequences.*

The basic idea of the improvement is to replace the gang by a more rational choice. For example, the more rational choice may be $w_{rat} = \{\neg B \neg b, \neg r, b\}$ (the interpretation $\{B \neg b, \neg r, \neg b\}$ is not a model of T').

Therefore, we may insert (w', w_{rat}) into ρ_2 . After the revision of ρ_2 – the new ρ_2 is $(\rho_2 \setminus (w', w_{min})) \cup (w', w_{rat})$ – holds $T' \models_{min} b$, therefore $Bb \in T'^$ and $T'^* \not\models \neg r \wedge r$.*

¹³ The relation ρ_1 allows to define a semantics of branching time. From this point of view, the application of the algorithm consist in checking the formula $EF h$. The formula means that there is some ρ_1 -path from w_T to some w such that h holds at w .

Definition 20 Let ϕ be a literal and \mathcal{I} be an interpretation. \mathcal{I} is called rational iff each of the following rationality conditions is satisfied:

$$\begin{aligned} K\phi \in \mathcal{I} &\Rightarrow \phi \in \mathcal{I} \\ B\phi \in \mathcal{I} &\Rightarrow \phi \in \mathcal{I} \\ \neg K\phi \in \mathcal{I} &\Rightarrow \phi \notin \mathcal{I} \\ \neg B\phi \in \mathcal{I} &\Rightarrow \phi \notin \mathcal{I} \end{aligned}$$

Definition 21 Let ϕ be an objective and ψ a subjective literal. A gang is a pair of literals (ϕ, ψ) such that it does not satisfy a rationality condition.

A rational modification of a gang (ϕ, ψ) is a pair of literals (ϕ', ψ') or (ϕ, ψ') , where ϕ' is a complementary literal to ϕ and ψ' to ψ .

If an interpretation \mathcal{I} of an AELKB-theory T contains a gang, then a repair of \mathcal{I} is a set S of interpretations \mathcal{J} such that some¹⁴ gangs of \mathcal{I} are in \mathcal{J} replaced by their rational modifications and each \mathcal{J} is a model of T .

Let an AELKB-structure $\mathcal{K} = (W, \{\rho_1, \rho_2\}, m)$ be given. Another AELKB-structure $\mathcal{K}' = (W, \{\rho_1, \rho_2'\}, m)$ ¹⁵ is called a reconstruction of \mathcal{K} , if there is at least one pair $(w_1, w_2) \in \rho_2 \setminus \rho_2'$ and a possible world w_2 containing a repair of an interpretation $\mathcal{I} \in w_2$ such that $(w_1, w_2) \in \rho_2' \setminus \rho_2$.

A computation of a reconstruction (of an AELKB-structure): If the model checking algorithm gives w_\perp for some u and w , we can proceed as follows. Let w_{rat} contains a repair of an interpretation $\mathcal{I} \in w_{min}$. Repeat: put $\rho_2 := (\rho_2 \setminus (w, w_{min})) \cup (w, w_{rat})$ and compute SAE again (until a consistent SAE is gained).

A summary: We do not change the concept of SAE, but the underlying semantic structure is changed. The modified AELKB-structure determines a modification of (minimal) entailment. Therefore, the set of derivable hypotheses of the form $B\phi$ is changed. The reasoning specified by the semantics can be called dynamic preferential entailment. (If some facts from the knowledge base contradict derivable beliefs, then we modify the given semantic specification of the entailment.)

We now compare our results concerning the revisions of AELB-theories with the results of [1].

A concept of careful SAE is introduced in [1]: First we define $Y \triangleleft X$ as Z , if Z is a maximal subset of X such that $Y \cup Z$ is consistent. Otherwise $Y \triangleleft X$ is \emptyset .

Definition 22 A careful static autoepistemic expansion of an AELB-theory T is $T^* = Cn_A(T \cup (T^* \triangleleft \{B\phi : T^* \models_{min} \phi\}))$.

A set $\mathcal{R}(T^*) = \{\phi : (T^* \models_{min} \phi) \wedge (B\phi \notin T^*)\}$ is called a revision set.

The next theorem corresponds to the Fundamental Theorem of Belief Revision, [1].

Theorem 7 Let $\mathcal{K} = (W, \{\rho_1, \rho_2\}, m)$ be an AELKB-structure. Let T be a consistent AELB-theory, $w_T = Mod(T)$.

¹⁴ There is a freedom in improving the impact of a gang. Our goal is not to use only rational interpretations (in order to avoid some non-intuitive consequences).

¹⁵ The only difference between \mathcal{K} and \mathcal{K}' is in accessibility relation: $\rho_2 \neq \rho_2'$.

Then holds: if $\{\phi : m_{w_T}(\phi) = 1\}$, where m_{w_T} is computed according to \mathcal{K} , is an inconsistent SAE of T , then there is an AELKB-structure \mathcal{K}' , a reconstruction of \mathcal{K} , such that the set $\{\phi : m_{w_T}(\phi) = 1\}$ is a careful SAE of T (for m_{w_T} computed according to \mathcal{K}').

Conversely, if a careful SAE of T is given, we can compute it as a SAE specified by a reconstruction of the corresponding AELKB-structure.

Theorem 8 *Let T , \mathcal{K} and w_T be as in the Theorem 7. Let T^* be a careful SAE of T .*

Then there is a $\mathcal{K}' = (W, \{\rho_1, \rho_2'\}, m)$, a reconstruction of \mathcal{K} , such that $T^ = \{\phi : m_{w_T}(\phi) = 1\}$, where m_{w_T} is computed according to \mathcal{K}' .*

Proof Sketch: Let T be an AELKB-theory, and $(w_T, w') \in \rho_2$. Select a literal $\phi \in \mathcal{R}(T^*)$ and make a repair S of a model \mathcal{I} from w' such that $m_{w_{rat}}(\phi) \neq 1$, where $w_{rat} = (w' \setminus \{\mathcal{I}\}) \cup S$. Reconstruct the underlying AELKB-structure. Repeat until $T^* = \{\phi : m_{w_T}(\phi) = 1\}$. \square

Finally, a remark concerning a comparison of the presented approach to the other results of [1]: both the belief revision by theory change and the belief completion of [1] may be simulated by modifying w_T -component of pairs $(w_T, w_{min}) \in \rho_2$ (by transforming w_T to $f(w_T)$, to the set of all models of the changed theory).

9 Conclusions

Summary of the results: We have introduced AELKB-structures and provided a characterization of static autoepistemic expansions of AELKB-theories in terms of AELKB-structures was given. A method of computing SAE of AELKB-theories was outlined. Further, a DKS-characterization of insertions into AELKB-theories (together with a corresponding computation using model checking) was presented. Finally, a characterization of revisions of AELKB-theories (and a computation using an enhanced model checking) was described.

The approach of the Section 8 motivates a generalization of the DKS. DKS may be extended by a set of mappings from accessibility relations to accessibility relations. Moreover, other transformations may be added – transformations extending the sets of possible worlds or transformations extending the vocabularies associated to possible worlds.

Some of the other goals of the future research – a detailed study of dynamic preferential entailment (modifications of minimal entailment in the presence of incomplete knowledge), computation of static autoepistemic expansions of AELKB theories, a characterization of deletions (from full AELKB-theories) in terms of DKS, default reasoning in DKS, a semantic characterization of reasoning about action in terms of DKS.

References

1. Alferes, J., Pereira, L., Przymusiński, T., *Belief Revision in Non-Monotonic Reasoning and Logic Programming*. Fundamenta Informaticae 1(1996)
2. Brass, S., Dix, J., Przymusiński, T., *Super Logic Programs*. Technical Report, Universität Hannover, Institut für Informatik, 1997
3. Buccafurri, F., Eiter, T., Gottlob, G., Leone, N., *Enhancing Symbolic Model Checking by AI Techniques*. FIG Research Report 9701, September 1997.
4. Chang, C., Keisler, H., *Model Theory*. North-Holland Publ. Comp., 1973
5. Clarke, E.M., Emerson, E.A., Sistla, A.P., *Automatic Verification of Finite-State Concurrent Systems Using Temporal Logic Specifications*. ACM Transactions on Programming Languages and Systems, vol. 8, No. 2, April 1986, 244-263
6. Kifer, M., Lozinskii, E., *RI: A Logic for Reasoning with Inconsistency*. Proc. of LICS 1989, IEEE Computer Society Press.
7. Lukasiewicz, W., *Non-Monotonic Reasoning. Formalization of Commonsense Reasoning*. Ellis Horwood Ltd. 1990
8. Moore, R., *Possible-World Semantics for Autoepistemic Logic*. Proc. AAAI Workshop on Non-Monotonic Reasoning, New Paltz, NY, 344-354
9. Przymusiński, T., *Autoepistemic logic of knowledge and beliefs*. Artificial Intelligence 95 (1997), 115-154
10. Šeřránek, J., *Dynamic Kripke structures*. CAEPIA'97. Actas de la VII Conferencia de la Asociación Española para la Inteligencia Artificial, Malaga 1997, 271-283
11. J. Van Benthem, *Semantic Parallels in Natural Language and Computation*, in: Logic Colloquium '87, eds. Ebbinghaus H.-D. et al., 1989, pp. 331-375, North Holland, Amsterdam