

Predicate Introduction for Logics with Fixpoint Semantics. Part II: Autoepistemic Logic. *

Joost Vennekens

joost.vennekens@cs.kuleuven.be
*Department of Computer Science, K.U. Leuven,
Celestijnlaan 200A
B-3001 Leuven, Belgium*

Maarten Mariën

maartenm@cs.kuleuven.be
*Department of Computer Science, K.U. Leuven,
Celestijnlaan 200A
B-3001 Leuven, Belgium*

Johan Wittocx[†]

johan@cs.kuleuven.be
*Department of Computer Science, K.U. Leuven,
Celestijnlaan 200A
B-3001 Leuven, Belgium*

Marc Denecker

marcd@cs.kuleuven.be
*Department of Computer Science, K.U. Leuven,
Celestijnlaan 200A
B-3001 Leuven, Belgium*

Abstract. We study the transformation of “predicate introduction” in non-monotonic logics. By this, we mean the act of replacing a complex formula by a newly defined predicate. From a knowledge representation perspective, such transformations can be used to eliminate redundancy or to simplify a theory. From a more practical point of view, they can also be used to transform a theory into a normal form imposed by certain inference programs or theorems. In a companion paper, we developed an algebraic theory that considers predicate introduction within the framework of “approximation theory,” a fixpoint theory for non-monotone operators that generalizes all main semantics of various non-monotonic logics, including logic programming, default logic and autoepistemic logic. We then used these results to show that certain logic programming transformations are equivalence preserving under, among others, both the stable and well-founded semantics. In this paper, we now apply the same algebraic results to autoepistemic logic and prove that a transformation to reduce the nesting depth of modal operators is equivalence preserving under a family of semantics for this logic. This not only provides useful theorems for autoepistemic logic, but also demonstrates that our algebraic theory does indeed capture the essence of predicate introduction in a generally applicable way.

*Works supported by FWO-Vlaanderen, IWT-Vlaanderen, and by GOA/2003/08.

[†]Research Assistant of the Fund for Scientific Research-Flanders (Belgium) (FWO-Vlaanderen).

1. Introduction

In a companion paper [10], an abstract theory of predicate introduction was developed within the algebraic framework of approximation theory. We then applied these abstract results to logic programming, showing that under certain circumstances, it is possible to replace a complex formula in the body of a rule by a new predicate symbol. A new feature of our work was that this new predicate could itself be defined by a—possibly recursive—set of rules. This allowed us to, among others, develop a general method of eliminating universal quantifiers in rule bodies. Due to the general nature of our algebraic results, they can equally be applied to other logics whose semantics admit a fixpoint characterization in the setting of approximation theory. In this paper, we demonstrate this by applying said results to autoepistemic logic.

Concretely, we study transformations that introduce new propositions to reduce the nesting level of the modal operator K . For instance, in the formula

$$\neg K(r \vee \neg Ks)$$

the K operator is nested to depth 2. By introducing a new proposition p to replace Ks , we can transform this formula to $\neg K(r \vee \neg p)$, with nesting depth 1. The new proposition p can then be ‘defined’ by the formula $Ks \Rightarrow p$. We will show that, on an algebraic level, what happens here is precisely the same as what happens with the predicate introduction transformation for logic programming, that we studied in our companion paper. As such, we will be able to use our algebraic results to prove that this transformation is equivalence preserving under a number of different semantics for autoepistemic logic.

This paper is structured as follows. Section 2 briefly summarizes the essential results of our algebraic treatment of predicate introduction in the context of approximation theory. In Section 3, we discuss how autoepistemic logic fits into the framework of approximation theory. Section 4 then contains the core result of this papers, namely, the application of our algebraic theorems to autoepistemic logic.

2. Preliminaries

For reasons of self-containedness, this section summarizes some essential concepts and results of approximation theory and of our study of predicate introduction in this setting.

2.1. Approximation theory

We use the following notations. Let $\langle L, \leq \rangle$ be a complete lattice. A fixpoint of an operator $O : L \rightarrow L$ on L is an element $x \in L$ for which $x = O(x)$; a prefixpoint of O is an x such that $x \geq O(x)$. We denote the set of all fixpoints of O as $fp(O)$. If O is monotone, then it has a unique least fixpoint x , which is also its unique least prefixpoint. We denote this x by $lfp(O)$.

Our presentation of approximation theory is based on [3, 4]. We consider the square L^2 of the domain of some lattice L . We will denote an element of L^2 as $(x \ y)$. We introduce the following projection functions: for a tuple $(x \ y)$, we denote by $[(x \ y)]$ the first element x of this pair and by $|(x \ y)|$ the second element y . The obvious point-wise extension of \leq to L^2 is called the *product order* on L^2 , which we also denote by \leq : i.e., for all $(x \ y), (x' \ y') \in L^2$, $(x \ y) \leq (x' \ y')$ iff $x \leq x'$ and $y \leq y'$. An element $(x \ y)$ of L^2 can be seen as approximating certain elements of L , namely those in the (possibly empty) interval $[x, y] = \{z \in L \mid x \leq z \text{ and } z \leq y\}$. Using this intuition, we can derive a second order,

the *precision order* \leq_p , on L^2 : for each $(x y), (x' y') \in L^2$, $(x y) \leq_p (x' y')$ iff $x \leq x'$ and $y' \leq y$. Indeed, if $(x y) \leq_p (x' y')$, then $[x, y] \supseteq [x', y']$, i.e., $(x' y')$ approximates fewer elements than $(x y)$. It can easily be seen that $\langle L^2, \leq_p \rangle$ is also a lattice. The structure $\langle L^2, \leq, \leq_p \rangle$ is the *bilattice* corresponding to L . If $\langle L, \leq \rangle$ is complete, then so are $\langle L^2, \leq \rangle$ and $\langle L^2, \leq_p \rangle$. Elements $(x x)$ of L^2 are called *exact*. The set of exact elements forms a natural embedding of L in L^2 . We denote the set of all exact elements of a lattice L^2 as $Diag(L^2)$.

Approximation theory is based on the study of operators which are monotone w.r.t. \leq_p . Such operators are called *approximations*. An approximation A approximates an operator O on L if for each $x \in L$, $A(x x)$ contains $O(x)$, i.e. $[A(x x)] \leq O(x) \leq |A(x x)|$. An exact approximation is one which maps exact elements to exact elements, i.e., for all $x \in L$, $[A(x x)] = |A(x x)|$. Each exact approximation A approximates a unique operator O on L , namely the one that maps each $x \in L$ to $[A(x x)] = |A(x x)|$. An approximation A is *symmetric* if $\forall (x y) \in L^2$, if $A(x y) = (x' y')$ then $A(y x) = (y' x')$. A symmetric approximation is exact.

For an approximation A on L^2 , we define the operator $[A(\cdot y)]$ on L that maps an element $x \in L$ to $[A(x y)]$, i.e. $[A(\cdot y)] = \lambda x. [A(x y)]$, and $|A(x \cdot)|$ that maps an element $y \in L$ to $|A(x y)|$. These operators are monotone. We define an operator C_A^\downarrow on L , called the *lower stable operator* of A , as $C_A^\downarrow(y) = lfp([A(\cdot y)])$. We also define the *upper stable operator* C_A^\uparrow of A as $C_A^\uparrow(x) = lfp(|A(x \cdot)|)$. Note that if A is symmetric, both operators are identical. We define the *stable operator* $\mathcal{C}_A : L^2 \rightarrow L^2$ of A by $\mathcal{C}_A(x y) = (C_A^\downarrow(y) C_A^\uparrow(x))$. Because both C_A^\downarrow and C_A^\uparrow are anti-monotone, \mathcal{C}_A is \leq_p -monotone.

An approximation A defines a number of different fixpoints: the least fixpoint of A is called its *Kripke-Kleene fixpoint*, fixpoints of its stable operator \mathcal{C}_A are *stable fixpoints* and the least fixpoint of \mathcal{C}_A is called the *well-founded fixpoint* of A . In [3, 4], it was shown that all main semantics of logic programming, autoepistemic logic and default logic can be characterized in terms of these fixpoints. In Section 3, we will recall how autoepistemic logic fits into this framework.

2.2. Fixpoint extension

In this section, we summarize our results from [10] on the concept of a *fixpoint extension*, which is our main algebraic tool for analyzing predicate introduction transformations.

We assume two complete lattices $\langle L_1, \leq_1 \rangle$ and $\langle L_2, \leq_2 \rangle$ and consider the square $(L_1 \times L_2)^2$ of the Cartesian product $L_1 \times L_2$, which is isomorphic to the Cartesian product $L_1^2 \times L_2^2$ of the squares of these lattices. (For instance, the function that maps each $((x u) (y v)) \in (L_1 \times L_2)^2$ to $((x y) (u v)) \in L_1^2 \times L_2^2$ is an isomorphism.) We denote pairs $P = ((x y) (u v))$ of this latter Cartesian product $L_1^2 \times L_2^2$ as $\binom{x y}{u v}$, where $(x y) \in L_1^2$ and $(u v) \in L_2^2$. We define the following projection functions: by $[P]$ we denote the pair $\binom{x y}{u}$, by $|P|$ the pair $\binom{y}{v}$, by $\lceil P \rceil$ the pair $(x y)$, by $\lfloor P \rfloor$ the pair $(u v)$, by $\lfloor P \rfloor$ the element u , by $\lceil P \rceil$ the element x , by $|P|$ the element y , and by $|P|$ the element v .

For an operator B on $L_1^2 \times L_2^2$, we now define the following three monotonicity properties.

Definition 2.1. Let B be an approximation on $L_1^2 \times L_2^2$.

- B is *part-to-part monotone* iff for each $(x y) \in L_1^2$ and $(u v) \leq (u' v') \in L_2^2$,

$$\lfloor B\left(\binom{x y}{u v}\right) \rfloor \leq \lfloor B\left(\binom{x y}{u' v'}\right) \rfloor;$$

- B is *part-to-whole monotone* iff for each $(x \ y) \in L_1^2$ and $(u \ v) \leq (u' \ v') \in L_2^2$,

$$B\left(\begin{smallmatrix} x & y \\ u & v \end{smallmatrix}\right) \leq B\left(\begin{smallmatrix} x & y \\ u' & v' \end{smallmatrix}\right);$$

- B is *whole-to-part monotone* iff for all $\left(\begin{smallmatrix} x & y \\ u & v \end{smallmatrix}\right) \leq \left(\begin{smallmatrix} x' & y' \\ u' & v' \end{smallmatrix}\right) \in L_1^2 \times L_2^2$,

$$\lfloor B\left(\begin{smallmatrix} x & y \\ u & v \end{smallmatrix}\right) \rfloor \leq \lfloor B\left(\begin{smallmatrix} x' & y' \\ u' & v' \end{smallmatrix}\right) \rfloor.$$

Given an operator B on $L_1^2 \times L_2^2$ and a pair $(x \ y) \in L_1^2$, we define the operator $B^{(x \ y)}$ on L_2^2 as $\lambda(u \ v).\lfloor B\left(\begin{smallmatrix} x & y \\ u & v \end{smallmatrix}\right) \rfloor$. Conversely, given a pair $(u \ v) \in L_2^2$, we define the operator $B_{(u \ v)}$ on L_1^2 as $\lambda(x \ y).\lfloor B\left(\begin{smallmatrix} x & y \\ u & v \end{smallmatrix}\right) \rfloor$. Now, if B is part-to-part monotone, then every $B^{(x \ y)}$ is a \leq -monotone operator, which must therefore have a \leq -least fixpoint $\text{lfp}(B^{(x \ y)})$.

Our results concern the relation between the fixpoints of an approximation B on $(L_1 \times L_2)^2$ and those of an approximation A on L_1^2 , where A and B satisfy the following property.

Definition 2.2. (Fixpoint extension)

Let B be an approximation on $L_1^2 \times L_2^2$ and A an approximation on L_1^2 . B is a *fixpoint extension* of A iff it is part-to-part monotone and, for all $x, y \in L_1$, $B_{\text{lfp}(B^{(x \ y)})}(x \ y) = A(x \ y)$.

In [10], the following results were proven.

Theorem 2.1. Let B be a fixpoint extension of A . A pair $(x \ y)$ is a fixpoint of A iff $\left(\begin{smallmatrix} x \\ \text{lfp}(B^{(x \ y)}) \end{smallmatrix}\right)$ is a fixpoint of B .

If the operator B happens to be such that, for all of its fixpoints $\left(\begin{smallmatrix} x & y \\ u & v \end{smallmatrix}\right)$, $(u \ v) = \text{lfp}(B^{(x \ y)})$, then the above theorem implies a stronger property: in this case, the set $\text{fp}(A)$ of fixpoints of A coincides with the set $\lfloor \text{fp}(B) \rfloor$ of restrictions to L_1^2 of fixpoints of B and, therefore, the \leq_p -least fixpoint of A must also be equal to the restriction $\lfloor \text{lfp}(B) \rfloor$ of the \leq_p -least fixpoint of B . An interesting special case of this is when the operator B is *part-to-part constant*, meaning that for every $(x \ y)$, the operator $B^{(x \ y)}$ is constant.

If the fixpoint extension B satisfies the additional property of being either whole-to-part or part-to-whole monotone, then its stable and well-founded fixpoints are related to those of A in the following way.

Theorem 2.2. Let B be a fixpoint extension of A , such that B is either part-to-whole or whole-to-part monotone. B has a stable fixpoint $\left(\begin{smallmatrix} x & y \\ u & v \end{smallmatrix}\right)$ iff $(x \ y)$ is a stable fixpoint of A . Moreover, $(x \ y)$ is the well-founded fixpoint of A iff for some $(u \ v)$, $\left(\begin{smallmatrix} x & y \\ u & v \end{smallmatrix}\right)$ is the well-founded fixpoint of B .

3. Autoepistemic logic and approximation theory

In this section, we describe the syntax of autoepistemic logic and give a brief overview, based on [4], of how a number of different semantics for this logic can be defined using concepts of approximation theory.

Let \mathcal{L} be the language of propositional logic based on a set of atoms Σ . Extending this language with a modal operator K , gives a language \mathcal{L}_K of modal propositional logic. An autoepistemic theory is a set of formulas in this language \mathcal{L}_K . For such a formula φ , the subset of Σ containing all atoms which appear in φ , is denoted by $At(\varphi)$; atoms which appear in φ at least once outside the scope of the modal operator K are called *objective* atoms of φ and the set of all objective atoms of φ is denoted by $At_O(\varphi)$. A *modal literal* is a formula of the form $K(\psi)$, with ψ a formula. If φ is a subformula of ψ and φ appears negatively in ψ , we write $\varphi \in^- \psi$; if φ appears positively in ψ , we write $\varphi \in^+ \psi$. By the K -rank of an occurrence of a subformula φ in a formula ψ , we mean the number of modal operators in ψ , in whose scope φ occurs. As such, the objective atoms of ψ are precisely those atoms that have an occurrence of K -rank zero in ψ .

To illustrate, consider the following example:

$$T = \{\varphi_1 = p \vee \neg Kp; \varphi_2 = K(p \vee Kq) \vee q\}$$

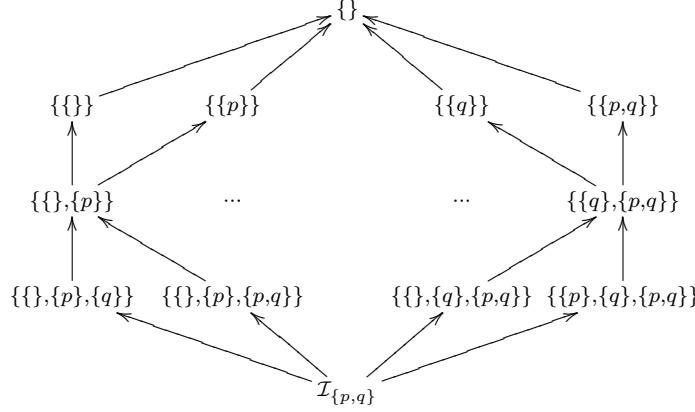
The objective atoms $At_O(\varphi_2)$ of φ_2 are $\{q\}$, while the atoms $At(\varphi_2)$ are $\{p, q\}$. The formula $K(p \vee q)$ is a modal literal of φ_2 . We also have that $Kp \in^- \varphi_1$ and $p \in^+ \varphi_2$. The K -rank of q in $K(p \vee Kq)$ is 2, whereas the K -rank of the second occurrence of p in φ_1 is 1.

An *interpretation* or *world* is a subset of the alphabet Σ . The set of all interpretations of Σ is denoted by \mathcal{I}_Σ , i.e. $\mathcal{I}_\Sigma = 2^\Sigma$. A *possible world structure* is a set of interpretations, i.e. the set of all possible world structures \mathcal{W}_Σ is defined as $2^{\mathcal{I}_\Sigma}$. Intuitively, a possible world structure sums up all “situations” which are possible. It therefore makes sense to order these according to inverse set inclusion to get a *knowledge order* \leq , i.e. for two possible world structures Q, Q' , $Q \leq Q'$ iff $Q \supseteq Q'$. Indeed, if a possible world structure contains *more* possibilities, it actually contains *less* knowledge. Figure 1 shows part of the lattice $\mathcal{W}_{At(T)}$ for the above example T .

Following [4], we will define the semantics of an autoepistemic theory by an operator on the bilattice $\mathcal{B}_\Sigma = \mathcal{W}_\Sigma^2$. An element $(P S)$ of \mathcal{B}_Σ is known as a *belief pair* and is called *consistent* iff $P \leq S$. In a consistent belief pair $(P S)$, P can be viewed as describing what must *certainly* be known, i.e., as giving an *underestimate* of what is known, while S can be viewed as denoting what might *possibly* be known, i.e. as giving an *overestimate*. Based on this intuition, there are two ways of estimating the truth of modal formulas according to $(P S)$. To conservatively estimate the truth of a formula φ in a world I and a consistent belief pair (P, S) , we evaluate all positively occurring modal literals $K\psi \in^+ \varphi$ in the possible world set with the least knowledge, i.e., in P , and all negatively occurring modal literals in the possible world set with the most knowledge, i.e., in S . Vice versa, to liberally estimate the truth of a formula φ in I and (P, S) , we evaluate positive modal literals in S and negative modal literals in P . To formalize these intuitions, we define the following truth assignment:

Definition 3.1. For each $(P S) \in \mathcal{B}$, $I \in \mathcal{I}_\Sigma$, $a \in \Sigma$ and formulas φ, φ_1 and φ_2 , we inductively define $\mathcal{H}_{I, (P S)}$ as:

- For each atom p , $\mathcal{H}_{I, (P S)}(p) = \mathbf{t}$ iff $p \in I$;

Figure 1. Part of the lattice $\mathcal{W}_{\{p,q\}}$.

- $\mathcal{H}_{I,(P\ S)}(\varphi_1 \wedge \varphi_2) = \mathbf{t}$, iff $\mathcal{H}_{I,(P\ S)}(\varphi_1) = \mathbf{t}$ and $\mathcal{H}_{I,(P\ S)}(\varphi_2) = \mathbf{t}$;
- $\mathcal{H}_{I,(P\ S)}(\varphi_1 \vee \varphi_2) = \mathbf{t}$, iff $\mathcal{H}_{I,(P\ S)}(\varphi_1) = \mathbf{t}$ or $\mathcal{H}_{I,(P\ S)}(\varphi_2) = \mathbf{t}$;
- $\mathcal{H}_{I,(P\ S)}(\neg\varphi) = \neg\mathcal{H}_{I,(S\ P)}(\varphi)$;
- $\mathcal{H}_{I,(P\ S)}(K\varphi) = \mathbf{t}$ iff $\mathcal{H}_{J,(P\ S)}(\varphi) = \mathbf{t}$ for all $J \in P$;

This evaluation function has two important properties. Firstly, if we consider an exact belief pair, i.e., one of the form $(Q\ Q)$, then $\mathcal{H}_{I,(Q\ Q)}(\varphi)$ corresponds to the standard S_5 evaluation [8] of φ in the possible world structure Q and world I . Secondly, there is an exact sense in which this function can be used to conservatively or liberally estimate the truth of a formula φ . A conservative estimate can be achieved by considering $\mathcal{H}_{I,(P\ S)}(\varphi)$. It can then be shown that for any possible world structure Q , with $P \leq Q \leq S$, it is indeed the case that $\mathcal{H}_{I,(P\ S)}(\varphi) \leq \mathcal{H}_{I,(Q\ Q)}(\varphi)$. Conversely, a liberal estimate consists of $\mathcal{H}_{I,(S\ P)}(\varphi)$ and, indeed, for any possible world structure Q , with $P \leq Q \leq S$, it is indeed the case that $\mathcal{H}_{I,(S\ P)}(\varphi) \geq \mathcal{H}_{I,(Q\ Q)}(\varphi)$.

We remark that the evaluation $\mathcal{H}_{I,(P\ S)}(K\varphi)$ of a modal literal $K\varphi$ depends only on $(P\ S)$ and not on I . We will sometimes emphasize such properties by replacing the irrelevant symbol by a dot, e.g., by writing $\mathcal{H}_{\cdot,(P\ S)}(K\varphi)$. Similarly, $\mathcal{H}_{I,(P\ S)}(\varphi)$ of an objective formula φ depends only on I , i.e., we can also write $\mathcal{H}_{I,(\cdot)}(\varphi)$.

The conservative and liberal way of estimating the truth of a theory can now be used to derive a new, more precise belief pair $(P'\ S')$ from an original pair $(P\ S)$. First, we will focus on constructing the new overestimate S' . As S' needs to overestimate knowledge, it needs to contain as few interpretations as possible. This means that S' should consist of only those interpretations, which manage to satisfy the theory even if the truth of its modal literals is conservatively estimated. So, S' should contain those interpretations I for which, for all φ in T , $\mathcal{H}_{I,(P\ S)}(\varphi) = \mathbf{t}$. Conversely, to construct the new underestimate P' , we need as many interpretations as possible. This means that P' should contain those interpretations I which satisfy the theory, when liberally evaluating its modal literals, i.e., for which, for all φ in T , $\mathcal{H}_{I,(S\ P)}(\varphi) = \mathbf{t}$.

These intuitions motivate the following definition of the operator \mathcal{D}_T on \mathcal{B} :

$$\mathcal{D}_T(P S) = (\mathcal{D}_T^u(S P) \mathcal{D}_T^u(P S))$$

with $\mathcal{D}_T^u(P S) = \{I \in \mathcal{I}_\Sigma \mid \forall \varphi \in T : \mathcal{H}_{I, (P S)}(\varphi) = \mathbf{t}\}$.

It can be illuminating to reformulate this definition using some more standard concepts and notation. Given a pair $(P S)$ and a formula φ , it is obviously the case that, in any evaluation $\mathcal{H}_{I, (P S)}(\varphi)$, all *positively* occurring modal literals $K\psi$ will be interpreted as $\mathcal{H}_{\cdot, (P S)}(K\psi)$, while all *negatively* occurring modal literals $K\psi$ will be interpreted as $\mathcal{H}_{\cdot, (S P)}(K\psi)$. Let us denote by $\varphi\langle P S \rangle$ the formula φ' that is the result of filling in these truth values, i.e., of replacing each top-level modal literal by \mathbf{t} or \mathbf{f} in the appropriate way. Every such $\varphi\langle P S \rangle$ is of course simply a propositional formula. What the function \mathcal{D}_T^u now actually does is simply map a pair $(P S)$ to the set $Mod(T\langle P S \rangle)$ of all classical models of the propositional theory $T\langle P S \rangle = \{\varphi\langle P S \rangle \mid \varphi \in T\}$. So, the operator \mathcal{D}_T can be equivalently defined as:

$$\mathcal{D}_T(P S) = (Mod(T\langle S P \rangle) Mod(T\langle P S \rangle)).$$

It can be shown that every operator \mathcal{D}_T is an approximation [4]. Moreover, since $|\mathcal{D}_T(P S)| = \mathcal{D}_T^u(S P) = |\mathcal{D}_T(S P)|$ and $|\mathcal{D}_T(P S)| = \mathcal{D}_T^u(P S) = |\mathcal{D}_T(S P)|$, it is symmetric and therefore approximates a unique operator on \mathcal{W}_Σ , namely the operator D_T , which maps each Q to $\mathcal{D}_T^u(Q Q)$. This operator D_T is precisely the operator considered in [9]. As shown in [4], these operators define a family of semantics for a theory T :

- fixpoints of D_T are *expansions* of T [9],
- fixpoints of \mathcal{D}_T are *partial expansions* of T [2],
- the least fixpoint of \mathcal{D}_T is the *Kripke-Kleene fixpoint* of T [2],
- fixpoints of $\mathcal{C}_{\mathcal{D}_T}^\downarrow$ are *extensions* of T [4],
- fixpoints of $\mathcal{C}_{\mathcal{D}_T}$ are *partial extensions* of T [4]
- the least fixpoint of $\mathcal{C}_{\mathcal{D}_T}$ is the *well-founded model* of T [4].

These various dialects of autoepistemic logic differ in their treatment of “ungrounded” expansions [5], i.e., expansions which arise from cyclicities such as $Kp \Rightarrow p$.

Example 3.1. To illustrate these definitions, we will compute the Kripke-Kleene model of our example theory $T = \{p \vee \neg Kp; K(p \vee Kq) \vee q\}$. This computation starts at the least precise element $(\mathcal{I}_{\{p,q\}} \{\})$ of $\mathcal{B}_{\{p,q\}}$. We first construct the new underestimate $\mathcal{D}_T^u(\{\} \mathcal{I}_{\{p,q\}}) = Mod(T\langle \{\} \mathcal{I}_{\{p,q\}} \rangle)$. It is easy to see that, for the negatively occurring modal literal Kp , $\mathcal{H}_{\cdot, (\mathcal{I}_{\{p,q\}} \{\})}(Kp) = \mathbf{t}$, and for the positively occurring modal literal $K(p \vee Kq)$, $\mathcal{H}_{\cdot, (\{\} \mathcal{I}_{\{p,q\}})}(K(p \vee Kq)) = \mathbf{t}$. Therefore, $T\langle \{\} \mathcal{I}_{\{p,q\}} \rangle = \{p \vee \neg \mathbf{f}; q \vee \mathbf{t}\}$ and $\mathcal{D}_T^u(\{\} \mathcal{I}_{\{p,q\}}) = \mathcal{I}_{\{p,q\}}$. Now, to compute the new overestimate $\mathcal{D}_T^u(\mathcal{I}_{\{p,q\}} \{\}) = Mod(T\langle \mathcal{I}_{\{p,q\}} \{\} \rangle)$, we note that

$$\mathcal{H}_{\cdot, (\{\} \mathcal{I}_{\{p,q\}})}(Kp) = \mathbf{t},$$

and

$$\mathcal{H}_{\cdot, (\mathcal{I}_{\{p,q\}} \cdot)}(K(p \vee Kq)) = \mathbf{f}.$$

Therefore, $T(\mathcal{I}_{\{p,q\}} \{\}) = \{p \vee \neg \mathbf{f}; q \vee \mathbf{f}\}$ and $\mathcal{D}_T^u(\mathcal{I}_{\{p,q\}} \{\}) = \{\{p, q\}\}$. So, $\mathcal{D}_T(\mathcal{I}_{\{p,q\}} \{\}) = (\mathcal{I}_{\{p,q\}} \{\{p, q\}\})$.

To compute $\mathcal{D}_T^u(\{p, q\} \mathcal{I}_{\{p,q\}})$, we note that it is still the case that:

$$\mathcal{H}_{\cdot, (\mathcal{I}_{\{p,q\}} \cdot)}(Kp) = \mathbf{t} \text{ and } \mathcal{H}_{\cdot, (\{\{p,q\}\} \cdot)}(K(p \vee Kq)) = \mathbf{t}.$$

So, $\mathcal{D}_T^u(\{\{p, q\}\} \mathcal{I}_{\{p,q\}}) = \mathcal{I}_{\{p,q\}}$. Similarly, still both $\mathcal{H}_{\cdot, (\{\{p,q\}\} \cdot)}(Kp) = \mathbf{t}$ and $\mathcal{H}_{\cdot, (\mathcal{I}_{\{p,q\}} \cdot)}(K(p \vee Kq)) = \mathbf{f}$. So, $\mathcal{D}_T^u(\mathcal{I}_{\{p,q\}} \{\{p, q\}\}) = \{\{p, q\}\}$. Therefore, $(\mathcal{I}_{\{p,q\}} \{\{p, q\}\})$ is the least fixpoint of \mathcal{D}_T , i.e., the Kripke-Kleene model of T .

4. Predicate introduction in AEL

In this section, we use our algebraic results to study the problem of predicate introduction in autoepistemic logic. Concretely, the goal of the transformations we consider is to eliminate nested modal operators by the introduction of a new propositional symbol. We begin with an informal analysis of such transformations.

4.1. Introduction to the problem

As an example, let us consider the following formula:

$$t \Rightarrow K(Kr \Rightarrow Ks).$$

This formula states that, under some condition t , the reasoner knows that knowledge about r implies knowledge about s .

Now, suppose we introduce some proposition p , not belonging to the original alphabet $\Sigma = \{t, r, s\}$ of this formula and that we would like p to mean “ s is known”. In general, we have a formula F of a theory T and want to replace a subformula $K\varphi$ (here, $\varphi = s$), that appears inside the scope of some other modal operator. We can assume without loss of generality that $T = \{F\}$, because every theory is equivalent to the singleton theory consisting of the conjunction of its formulas. Let ψ be the smallest modal literal of F that contains $K\varphi$. In our example, $\psi = K(Kr \Rightarrow Ks)$. Let F' be the result of replacing $K\varphi$ by p , i.e., in this case $F' = (t \Rightarrow K(Kr \Rightarrow p))$. Intuitively, what we want to do now is construct a formula F_p that defines the new atom p in such a way that the models (under some semantics) of the original theory T coincide with the restrictions to the original alphabet Σ of the models of the new theory $T' = \{F', F_p\}$. We will now give some intuitions on how we should construct such an F_p for the above example.

Perhaps the most obvious candidate formula would be $Ks \Leftrightarrow p$, which abbreviates $(Ks \Rightarrow p) \wedge (Ks \Leftarrow p)$. However, it turns out that this will not work. If we abbreviate the set of all interpretations $\mathcal{I}_{\{t,r,s,p\}}$ for the alphabet of T' as \mathcal{I} , we see that T' has a partial expansion $(\mathcal{I} \{\})$. Indeed, on the one hand, $T' \langle \{\} \mathcal{I} \rangle = \{t \Rightarrow \mathbf{t}; \mathbf{f} \Rightarrow p; \mathbf{t} \Leftarrow p\}$, whose set of models is \mathcal{I} , while, on the other hand, $T' \langle \mathcal{I} \{\} \rangle$ contains both the formula $\mathbf{t} \Rightarrow p$ and $\mathbf{f} \Leftarrow p$ and, as such, has no models. However, the corresponding pair $(\mathcal{I}_{\{r,t,s\}} \{\})$ is not a partial expansion of the original formula F , because $F \langle \mathcal{I}_{\{r,t,s\}} \{\} \rangle = (t \Rightarrow \mathbf{f})$, of which every interpretation in which t is false is a model, so $Mod(t \Rightarrow \mathbf{f}) \neq \{\}$. We therefore need to look for a different formula F_p .

It is clear that, in order to ensure that we do get the result we want, it would suffice to get $\psi = K(Kr \Rightarrow Ks)$ to be equivalent to $\psi' = K(Kr \Rightarrow p)$. Now, ψ holds iff it is the case that either there is a possible world in which r is false or in all possible worlds, s is true. The formula ψ' , on the other hand, holds iff it is the case that either there is a possible world in which r is false or in all possible worlds, p is true. As such, the formula F_p should force p to be known iff originally s was known. This suggests the formula $Ks \Rightarrow p$. Indeed, if it was originally the case that Ks , then the only possible world for p will be $\{p\}$, but otherwise both $\{\}$ and $\{p\}$ will be possible and p will not be known. A similar line of reasoning applies whenever we want to replace a formula $K\varphi$ that appears *positively* within the scope of the smallest modal literal ψ that contains it.

To illustrate the other case, let us now try to replace Kr in the above formula F by a new atom q . Once again, it suffices to ensure that ψ is equivalent to the formula $\psi'' = K(q \Rightarrow Ks)$. Now, ψ'' holds iff either in all possible worlds q is false or in all possible worlds s is true. As such, our formula F_q should, in this case, make sure that q is known to be false whenever, originally, r was not known to be true. That is to say, if there exists a world in which r is false, then q should be false in all worlds. This suggests the formula $\neg Kr \Rightarrow \neg q$, which is of course simply a rewriting of $Kr \Leftarrow q$. Once again, a similar line of reasoning applies whenever we are trying to replace a formula $K\varphi$ that appears *negatively* within the scope of the smallest modal literal ψ that contains it.

Let us now formally define the problem that we want to consider.

Definition 4.1. Let $T = \{F\}$ be an autoepistemic theory. We consider an occurrence of a modal literal $K\varphi$ with K -rank at least 1. Let p be a proposition that does not belong to the alphabet of T . The result of *introducing p to replace (this occurrence of) $K\varphi$* is the theory $T' = \{F', F_p\}$, where F' is the result of replacing the selected occurrence of $K\varphi$ in F by p and F_p is defined as follows, depending on how $K\varphi$ appears inside the scope of the smallest modal literal ψ that contains it:

- If $K\varphi \in^+ \psi$, then $F_p = (K\varphi \Rightarrow p)$;
- If $K\varphi \in^- \psi$, then $F_p = (K\varphi \Leftarrow p)$.

The question we we will study is when (and for which of the previously mentioned semantics) the result of introducing p to replace $K\varphi$ will be equivalent to the original theory. The rest of this section will continue to use the notations introduced in the above definition. We will also use Σ to denote the alphabet of the original theory T and Σ' to denote the alphabet $\Sigma \cup \{p\}$ of T' . By ψ' we will denote the result of replacing $K\varphi$ by p in ψ .

4.2. Application of the algebraic results

In this section, we now apply our algebraic theorems to the problem at hand. Recall that these relate an approximation A on a lattice L_1^2 to a fixpoint extension B of A , which is an operator on a lattice $L_1^2 \times L_2^2$. In the current case, L_1^2 will be the lattice \mathcal{B}_Σ of pairs of possible worlds structures for the original alphabet Σ and L_2^2 will be the lattice $\mathcal{B}_{\{p\}}$ of pairs of possible world structures for the new alphabet $\{p\}$. Therefore, to play the role of the fixpoint extension B , we need an operator on the square of the product lattice $\mathcal{W}_\Sigma \times \mathcal{W}_{\{p\}}$. Let us denote this product lattice by $\tilde{\mathcal{W}}_{\Sigma'}$ and its square $\tilde{\mathcal{W}}_{\Sigma'}^2$ by $\tilde{\mathcal{B}}_{\Sigma'}$.

4.2.1. An intermediate operator

The most obvious candidate for such a fixpoint extension B is of course the operator $\mathcal{D}_{T'}$. However, it turns out that we cannot use $\mathcal{D}_{T'}$, because the lattice $\tilde{\mathcal{B}}_{\Sigma'}$ on which it operates is not isomorphic to the lattice $\tilde{\mathcal{B}}_{\Sigma}$. Let us examine this problem more closely. A pair $\binom{X}{U} \in \tilde{\mathcal{W}}_{\Sigma'}$ can be mapped to the possible world structure P that consists of precisely all interpretations $I \cup J$ for which $I \in X$ and $J \in U$. We denote this mapping by $\kappa : \tilde{\mathcal{W}}_{\Sigma'} \rightarrow \mathcal{W}_{\Sigma}$ and also define a similar function $\bar{\kappa} : \tilde{\mathcal{B}}_{\Sigma'} \rightarrow \mathcal{B}_{\Sigma}$ as mapping each $\binom{X}{U} \in \tilde{\mathcal{B}}_{\Sigma'}$ to the belief pair $(\kappa(\binom{X}{U}) \ \kappa(\binom{Y}{V}))$.

Now, this κ is a homomorphism, preserving the knowledge order \leq , which implies that $\bar{\kappa}$ is also a homomorphism, preserving both precision and product order. However, κ is not an isomorphism. Firstly, it is easy to see that κ is not surjective. For instance, with $\Sigma = \{q\}$, the possible world structure $Q = \{\{p, q\}, \{\}\}$ does not belong to $\kappa(\tilde{\mathcal{W}}_{\Sigma'})$. Indeed, for Q to be equal to some $\kappa(\binom{X}{U})$, it would have to be the case that X contains both $\{p\}$ and $\{\}$, while U would have to contain both $\{q\}$ and $\{\}$, but then $\{p\} \cup \{q\}$ would also have to belong to Q . However, none of the possible world sets outside of $\kappa(\tilde{\mathcal{W}}_{\Sigma'})$ are relevant for the operator $\mathcal{D}_{T'}$. Indeed, for any belief pair $(P' \ S')$ in the image of this operator, both P' and S' will be of the form $Mod(T' \langle P \ S \rangle)$ for some belief pair $(P \ S)$. Now, every $T' \langle P \ S \rangle$ is a—classical—propositional theory, which consists of a formula $F' \langle P \ S \rangle$ in alphabet Σ , and a formula $F_p \langle P \ S \rangle$ in alphabet $\{p\}$. Because these two alphabets are disjoint, it is an obvious property of propositional logic that $Mod(T' \langle P \ S \rangle)$ consists of all $I \cup J$ for which I is a model of $F' \langle P \ S \rangle$ and J is a model of $F_p \langle P \ S \rangle$. In other words, for any $(P \ S)$, $Mod(T' \langle P \ S \rangle) = \kappa(\binom{Mod(F' \langle P \ S \rangle)}{Mod(F_p \langle P \ S \rangle)})$. We therefore conclude that $\mathcal{D}_{T'}(\tilde{\mathcal{B}}_{\Sigma'}) \subseteq \bar{\kappa}(\tilde{\mathcal{B}}_{\Sigma'})$.

Besides not being surjective, κ is also not injective. Indeed, any pair of the form $\binom{\{\}}{U}$ or $\binom{X}{\{\}}$ is mapped to $\{\}$. To overcome this problem, we will often restrict our attention to certain subsets of $\tilde{\mathcal{W}}_{\Sigma'}$. By $\tilde{\mathcal{W}}_{\Sigma'}^c$, we denote the set of all consistent elements of $\tilde{\mathcal{W}}_{\Sigma'}$, that is, all $\binom{X}{U}$ for which both $X \neq \{\}$ and $U \neq \{\}$ or, equivalently, $\kappa(\binom{X}{U}) \neq \{\}$. By $\tilde{\mathcal{W}}_{\Sigma'}^{\downarrow c}$ we denote the set of all $\binom{X}{U}$ for which $U \neq \{\}$ and $\tilde{\mathcal{W}}_{\Sigma'}^{\uparrow c}$ denotes the set of all $\binom{X}{U}$ for which $X \neq \{\}$. We use similar notations $\tilde{\mathcal{B}}_{\Sigma'}^c$, $\tilde{\mathcal{B}}_{\Sigma'}^{\downarrow c}$ and $\tilde{\mathcal{B}}_{\Sigma'}^{\uparrow c}$ for the squares of these lattices.

We now summarize some relevant properties of κ that follow directly from the above discussion.

Lemma 4.1. The function κ has the following properties:

1. κ is injective on the subset $\tilde{\mathcal{W}}_{\Sigma'}^c$ of its domain;
2. For all $\binom{X}{U} \in \tilde{\mathcal{W}}_{\Sigma'}^{\downarrow c}$, $\kappa(\binom{X}{U})|_{\Sigma} = X$ and for all $\binom{X}{U} \in \tilde{\mathcal{W}}_{\Sigma'}^{\uparrow c}$, $\kappa(\binom{X}{U})|_{\{p\}} = U$;
3. $\mathcal{D}_{T'}(\tilde{\mathcal{B}}_{\Sigma'}) \subseteq \bar{\kappa}(\tilde{\mathcal{B}}_{\Sigma'})$.

Because $\tilde{\mathcal{B}}_{\Sigma'}$ and \mathcal{B}_{Σ} are not isomorphic, we cannot directly apply our algebraic results to the operators $\mathcal{D}_{T'}$ and \mathcal{D}_T . Instead, we will define an intermediate operator $\tilde{\mathcal{D}}_{T'}$ on $\tilde{\mathcal{B}}_{\Sigma'}$, such that, on the one hand, this $\tilde{\mathcal{D}}_{T'}$ is a fixpoint extension of the operator \mathcal{D}_T and, on the other hand, the fixpoints and stable fixpoints of $\tilde{\mathcal{D}}_{T'}$ correspond to those of $\mathcal{D}_{T'}$. Together, these two results will then of course provide us with the wanted correspondence between fixpoints and stable fixpoints of $\mathcal{D}_{T'}$ and \mathcal{D}_T .

Definition 4.2. We define the function $\tilde{\mathcal{D}}_T^u$ from $\tilde{\mathcal{B}}_{\Sigma'}$ to $\tilde{\mathcal{W}}_{\Sigma'}$ as mapping every $\binom{X}{U} \in \tilde{\mathcal{B}}_{\Sigma'}$ to the pair $\binom{X'}{U'}$ for which:

- $X' = Mod(F' \langle \bar{\kappa}(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix}) \rangle)$;
- $U' = Mod(F_p \langle X Y \rangle)$

We also define $\tilde{\mathcal{D}}_T(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})$ as $(\tilde{\mathcal{D}}_T^u(\begin{smallmatrix} Y & X \\ V & U \end{smallmatrix}) \tilde{\mathcal{D}}_T^u(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix}))$ and $\tilde{D}_T(\begin{smallmatrix} X \\ U \end{smallmatrix})$ as $\tilde{\mathcal{D}}_{T'}^u(\begin{smallmatrix} X & X \\ U & U \end{smallmatrix})$.

This operator $\tilde{\mathcal{D}}_{T'}$ differs from $\mathcal{D}_{T'}$ in two respects. Firstly, in the construction of a new belief pair for alphabet Σ , it considers only the formula F , whereas, in the construction of a new belief pair for $\{p\}$, it only considers F_p . Secondly, a new belief pair for $\{p\}$ is constructed using only the original belief pair $(X Y)$ for Σ , instead of the entire belief pair $\bar{\kappa}(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})$. Because of this, every operator $\tilde{\mathcal{D}}_{T'}^{(X Y)} = [\tilde{\mathcal{D}}_{T'}(\begin{smallmatrix} X & Y \\ \cdot & \cdot \end{smallmatrix})]$ is actually constant. We now investigate how we can characterize the pair $(U V)$ that is the unique element in the image of some $\tilde{\mathcal{D}}_{T'}^{(X Y)}$.

First, let us observe that if $K\varphi \in^+ \psi$, we can determine $Mod(F_p \langle X Y \rangle)$ as follows:

$K\varphi \in^+ \psi$	$K\varphi \langle Y X \rangle = \mathbf{t}$	$K\varphi \langle Y X \rangle = \mathbf{f}$
F_p	$K\varphi \Rightarrow p$	$K\varphi \Rightarrow p$
$F_p \langle X Y \rangle$	$\mathbf{t} \Rightarrow p$	$\mathbf{f} \Rightarrow p$
$Mod(F_p \langle X Y \rangle)$	$\{\{p\}\}$	$\{\{\}, \{p\}\}$

For $K\varphi \in^- \psi$, the analogous table is as follows:

$K\varphi \in^- \psi$	$K\varphi \langle X Y \rangle = \mathbf{t}$	$K\varphi \langle X Y \rangle = \mathbf{f}$
F_p	$K\varphi \Leftarrow p$	$K\varphi \Leftarrow p$
$F_p \langle X Y \rangle$	$\mathbf{t} \Leftarrow p$	$\mathbf{f} \Leftarrow p$
$Mod(F_p \langle X Y \rangle)$	$\{\{\}, \{p\}\}$	$\{\{\}\}$

By definition, if $(U V)$ is the unique element in the image of some $\tilde{\mathcal{D}}_{T'}^{(X Y)}$, then U is $Mod(F_p \langle Y X \rangle)$ and V is $Mod(F_p \langle X Y \rangle)$, so we can find the precise values of these two possible world structures by means of the above tables. We remark that, in particular, it is always the case that both $U \neq \{\}$ and $V \neq \{\}$, i.e., $\tilde{\mathcal{D}}_{T'}(\tilde{\mathcal{B}}_{\Sigma'}) \subseteq \tilde{\mathcal{B}}_{\Sigma'}^{\downarrow c}$.

We are of course mainly interested in fixpoints of $\tilde{\mathcal{D}}_{T'}$. Clearly, for every such fixpoint $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})$, it has to be the case that $(U V)$ is the unique element in the image of $\tilde{\mathcal{D}}_{T'}^{(X Y)}$. Let us denote by $\tilde{\mathcal{B}}_{\Sigma'}^*$ the set of all $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})$ for which this last property holds. It can easily be seen from the above discussion that, in every such element of $\tilde{\mathcal{B}}_{\Sigma'}^*$, the values of U and V are as follows.

Lemma 4.2. For all $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix}) \in \tilde{\mathcal{B}}_{\Sigma'}^*$, the values of U and V depend on the values of X and Y , and on how $K\varphi$ appears in ψ , as given by the following table:

	$K\varphi\langle X Y \rangle$		$K\varphi\langle Y X \rangle$	
	t	f	t	f
$K\varphi \in^+ \psi$	$U = \{\{p\}\}$	$U = \{\{\}, \{p\}\}$	$V = \{\{p\}\}$	$V = \{\{\}, \{p\}\}$
$K\varphi \in^- \psi$	$V = \{\{\}, \{p\}\}$	$V = \{\{\}\}$	$U = \{\{\}, \{p\}\}$	$U = \{\{\}\}$

4.2.2. Relating fixpoints of $\tilde{\mathcal{D}}_{T'}$ and $\mathcal{D}_{T'}$

We now show that $\tilde{\mathcal{D}}_{T'}$ and $\mathcal{D}_{T'}$ have the same fixpoints. Because every fixpoint of $\tilde{\mathcal{D}}_{T'}$ must obviously belong to its image, it suffices to consider only $\tilde{\mathcal{B}}_{\Sigma'}^{\downarrow c}$, i.e., the set of all $\begin{pmatrix} X & Y \\ U & V \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}$ for which $U, V \neq \{\}$. On this particular part of its domain, the operator $\tilde{\mathcal{D}}_{T'}$ is related to $\mathcal{D}_{T'}$ in the following way.

Lemma 4.3. For all $\begin{pmatrix} X & Y \\ U & V \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}^{\downarrow c}$, $\bar{\kappa}(\tilde{\mathcal{D}}_{T'}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix})) = \mathcal{D}_{T'}(\bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix}))$.

Proof:

Let $\begin{pmatrix} X & Y \\ U & V \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}^{\downarrow c}$ and let $(P S)$ be $\bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix})$. As we already showed in the discussion leading up to Lemma 4.1, $Mod(T'\langle P S \rangle) = \kappa_{Mod(F_p\langle P S \rangle)}^{Mod(F'\langle P S \rangle)}$. Therefore, it suffices to show that $F_p\langle P S \rangle = F_p\langle X Y \rangle$. Because $U, V \neq \{\}$, we have that $(P S)|_{\Sigma} = (X Y)$, which proves the result. \square

From this lemma, the correspondence between fixpoints now follows.

Theorem 4.1. For all $(P S) \in \mathcal{B}_{\Sigma}$, $(P S)$ is a fixpoint (respectively, the Kripke-Kleene fixpoint) of $\mathcal{D}_{T'}$ iff there exists a $\begin{pmatrix} X & Y \\ U & V \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}$ such that $\bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix}) = (P S)$ and $\begin{pmatrix} X & Y \\ U & V \end{pmatrix}$ is a fixpoint (the Kripke-Kleene fixpoint) of $\tilde{\mathcal{D}}_{T'}$.

Proof:

Every fixpoint $\begin{pmatrix} X & Y \\ U & V \end{pmatrix}$ of $\tilde{\mathcal{D}}_{T'}$ must belong to $\tilde{\mathcal{B}}_{\Sigma'}^{\downarrow c}$ and therefore it follows directly from Lemma 4.3 that $\bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix})$ is a fixpoint of $\mathcal{D}_{T'}$. To prove the other direction, we must show that for every fixpoint $(P S)$ of $\mathcal{D}_{T'}$, there exists a fixpoint $\begin{pmatrix} X & Y \\ U & V \end{pmatrix}$ of $\tilde{\mathcal{D}}_{T'}$ such that $\bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix}) = (P S)$. Let $(X Y)$ be $(P S)|_{\Sigma}$. We now define U as follows: if $P = \{\}$, then also $X = \{\}$ and we define $U = Mod(F_p\langle Y \{\} \rangle)$; otherwise, we define $U = P|_{\{p\}}$. Similarly, we define V as: if $S = \{\}$, then $V = Mod(F_p\langle X \{\} \rangle)$; otherwise $V = S|_{\{p\}}$. We now show that this $\begin{pmatrix} X & Y \\ U & V \end{pmatrix}$ satisfies the desired properties.

Firstly, we show that $\bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix}) = (P S)$, i.e., that $\kappa(\begin{pmatrix} X \\ U \end{pmatrix}) = P$ and $\kappa(\begin{pmatrix} Y \\ V \end{pmatrix}) = S$. If $P = \{\}$, then $\kappa(\begin{pmatrix} X \\ U \end{pmatrix}) = \kappa(\begin{pmatrix} \{\} \\ \{\} \end{pmatrix}) = \{\}$. If $P \neq \{\}$, then $X = P|_{\Sigma}$ and $U = P|_{\{p\}}$. Because $(P S)$ belongs to $\mathcal{D}_{T'}(\tilde{\mathcal{B}}_{\Sigma'}) \subseteq \bar{\kappa}(\tilde{\mathcal{B}}_{\Sigma'})$ (Lemma 4.1), we have that $P \in \kappa(\tilde{\mathcal{W}}_{\Sigma'})$. It follows that $P = \kappa_{P|_{\{p\}}}^{P|_{\Sigma}}$ and, therefore, $P = \kappa(\begin{pmatrix} X \\ U \end{pmatrix})$.

A similar argument shows that also $S = \kappa(\begin{pmatrix} Y \\ V \end{pmatrix})$. Secondly, we show that $\begin{pmatrix} X & Y \\ U & V \end{pmatrix}$ is indeed a fixpoint of $\tilde{\mathcal{D}}_{T'}$. Let $\begin{pmatrix} X' & Y' \\ U' & V' \end{pmatrix}$ be $\tilde{\mathcal{D}}_{T'}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix})$. Because by construction $U, V \neq \{\}$, Lemma 4.3 implies that

$\bar{\kappa}(\begin{smallmatrix} X' & Y' \\ U' & V' \end{smallmatrix}) = \mathcal{D}_T(\bar{\kappa}(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})) = \bar{\kappa}(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})$, that is, $\kappa(\begin{smallmatrix} X' \\ U' \end{smallmatrix}) = \kappa(\begin{smallmatrix} X \\ U \end{smallmatrix})$ and $\kappa(\begin{smallmatrix} Y' \\ V' \end{smallmatrix}) = \kappa(\begin{smallmatrix} Y \\ V \end{smallmatrix})$. We now show that this implies $(\begin{smallmatrix} X \\ U \end{smallmatrix}) = (\begin{smallmatrix} X' \\ U' \end{smallmatrix})$. Because both $U, U' \neq \{\}$, the equality $\kappa(\begin{smallmatrix} X' \\ U' \end{smallmatrix}) = \kappa(\begin{smallmatrix} X \\ U \end{smallmatrix})$ implies that $X = X'$. Now, if $X, X' \neq \{\}$, then this equality also implies $U = U'$. If, on the other hand, $X = X' = \{\}$, then by construction, $U = \text{Mod}(F_p\langle Y \{\}\rangle) = U'$. We conclude that in both cases $(\begin{smallmatrix} X \\ U \end{smallmatrix}) = (\begin{smallmatrix} X' \\ U' \end{smallmatrix})$. By a similar argument it follows that also $(\begin{smallmatrix} Y \\ V \end{smallmatrix}) = (\begin{smallmatrix} Y' \\ V' \end{smallmatrix})$, so $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})$ is indeed a fixpoint of $\tilde{\mathcal{D}}_{T'}$. \square

We now also examine the relation between stable fixpoints of $\mathcal{D}_{T'}$ and $\tilde{\mathcal{D}}_{T'}$. To this end, we compare the lower stable operators $C_{\mathcal{D}_{T'}}^\downarrow$ and $C_{\tilde{\mathcal{D}}_{T'}}^\downarrow$. It suffices to compare only these two operators, because, due to the symmetry of $\mathcal{D}_{T'}$ and $\tilde{\mathcal{D}}_{T'}$, we have that $C_{\mathcal{D}_{T'}}^\uparrow = C_{\tilde{\mathcal{D}}_{T'}}^\downarrow$ and $C_{\tilde{\mathcal{D}}_{T'}}^\uparrow = C_{\mathcal{D}_{T'}}^\downarrow$. Recall that $C_{\mathcal{D}_{T'}}^\downarrow$ is defined as mapping each S to $\text{lfp}([\mathcal{D}_T(\cdot, S)])$ and, similarly, $C_{\tilde{\mathcal{D}}_{T'}}^\downarrow$ maps each $(\begin{smallmatrix} Y \\ V \end{smallmatrix})$ to $\text{lfp}([\tilde{\mathcal{D}}_{T'}(\cdot, \begin{smallmatrix} Y \\ V \end{smallmatrix})])$. By a straightforward induction over the construction of these least fixpoints, Lemma 4.3 now implies the following result.

Lemma 4.4. For all $(\begin{smallmatrix} X \\ U \end{smallmatrix}) \in \mathcal{W}_\Sigma^{\downarrow c}$, $\kappa(C_{\tilde{\mathcal{D}}_{T'}}^\downarrow(\begin{smallmatrix} X \\ U \end{smallmatrix})) = C_{\mathcal{D}_{T'}}^\downarrow(\kappa(\begin{smallmatrix} X \\ U \end{smallmatrix}))$.

This lemma now implies the following correspondence between stable fixpoints of $\mathcal{D}_{T'}$ and $\tilde{\mathcal{D}}_{T'}$, in precisely the same way as Theorem 4.1 follows from Lemma 4.3.

Theorem 4.2. For all $(P \ S) \in \mathcal{B}_\Sigma$, $(P \ S)$ is a stable fixpoint (respectively, the well-founded fixpoint) of $\mathcal{D}_{T'}$ iff there exists a $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix}) \in \tilde{\mathcal{B}}_{\Sigma'}$ such that $\bar{\kappa}(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix}) = (P \ S)$ and $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})$ is a stable fixpoint (the well-founded fixpoint) of $\tilde{\mathcal{D}}_{T'}$.

Having shown that $\tilde{\mathcal{D}}_{T'}$ and $\mathcal{D}_{T'}$ have the same (stable) fixpoints, we can now proceed to relate the models of T' (under the various semantics we consider) to those of T , by using our algebraic theory of fixpoint extension to establish a correspondence between (stable) fixpoints of $\tilde{\mathcal{D}}_{T'}$ and \mathcal{D}_T .

4.2.3. $\tilde{\mathcal{D}}_{T'}$ is a fixpoint extension of \mathcal{D}_T

We now show that $\tilde{\mathcal{D}}_{T'}$ is indeed a fixpoint extension of \mathcal{D}_T . We first observe that, because $\mathcal{D}_{T'}$ is part-to-part constant, it is also part-to-part monotone. Therefore, all that remains to be shown is that, for all $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})$ with $(U \ V) = \text{lfp}(\tilde{\mathcal{D}}_{T'}^{(X \ Y)})$, $[\tilde{\mathcal{D}}_{T'}(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})] = \mathcal{D}_T(X \ Y)$. Of course, in this case, the condition that $(U \ V) = \text{lfp}(\tilde{\mathcal{D}}_{T'}^{(X \ Y)})$ is simply equivalent to $(U \ V)$ being the unique element in the image of $\tilde{\mathcal{D}}_{T'}^{(X \ Y)}$, i.e., to $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})$ belonging to $\tilde{\mathcal{B}}_{\Sigma'}^*$.

Lemma 4.5. $\tilde{\mathcal{D}}_{T'}$ is a fixpoint extension of \mathcal{D}_T .

Proof:

We need to prove that for all $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix}) \in \tilde{\mathcal{B}}_{\Sigma'}^*$, $[\tilde{\mathcal{D}}_{T'}(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})] = \mathcal{D}_T(X \ Y)$. By definition of these two operators, it suffices to show that, for all $(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix}) \in \tilde{\mathcal{B}}_{\Sigma'}^*$, $F'\langle\bar{\kappa}(\begin{smallmatrix} X & Y \\ U & V \end{smallmatrix})\rangle = F\langle X \ Y \rangle$ and $F'\langle\bar{\kappa}(\begin{smallmatrix} Y & X \\ V & U \end{smallmatrix})\rangle =$

$F\langle Y X \rangle$. Due to the symmetry of $\tilde{\mathcal{D}}_{T'}$, we have that $\begin{pmatrix} X & Y \\ U & V \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}^*$ iff $\begin{pmatrix} Y & X \\ V & U \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}^*$. Therefore, it suffices to show that for all $\begin{pmatrix} X & Y \\ U & V \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}^*$, $F'\langle \bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix}) \rangle = F\langle X Y \rangle$. Because the only difference between F' and F lies in the modal literals ψ' and ψ , it suffices to show that the way in which ψ' is evaluated in $F'\langle \bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix}) \rangle$ coincides with the way in which ψ is evaluated in $F\langle X Y \rangle$. The precise property that needs to be proven now depends on whether $\psi \in^+ F$ or $\psi \in^- F$, but by the same symmetry argument as above, we can cover both cases by showing that for all $\begin{pmatrix} X & Y \\ U & V \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}^*$, $\psi'\langle \bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix}) \rangle = \psi\langle X Y \rangle$.

Let us first consider the modal literal ψ' and let ρ' be the formula for which $\psi' = K\rho'$. For any belief pair $(P S)$, $K\rho'\langle P S \rangle$ is equal to the minimum of all truth values $\mathcal{H}_{I,(P S)}(\rho')$ for which $I \in P$. In the case of $(P S)$ being equal to $\bar{\kappa}(\begin{pmatrix} X & Y \\ U & V \end{pmatrix})$ for some $\begin{pmatrix} X & Y \\ U & V \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}^*$, this is equal to the minimum m of all $\mathcal{H}_{I \cup J,(P S)}(\rho')$ for which $I \in X$ and $J \in U$. Moreover, because there is only one occurrence of the atom p in the entire formula ρ' , there must exist a single “worst” truth value \mathbf{v} for p which gives rise to this minimum; more formally put, for some \mathbf{v} , m is equal to the minimum of all $\mathcal{H}_{I \cup J,(P S)}(\rho'[p/\mathbf{v}])$ for which $I \in X$ and $J \in U$. Indeed, on the one hand, if $p \in^+ \rho'$, then \mathbf{v} is the minimum of all $\mathcal{H}_{J,(\cdot)}(p)$ with $J \in U$. On the other hand, if $p \in^- \rho'$, then \mathbf{v} is the maximum of all $\mathcal{H}_{J,(\cdot)}(p)$ with $J \in U$. Because $U, V \neq \{\}$ and the formula $\rho'[p/\mathbf{v}]$ no longer contains p , we now have that m is equal to the minimum of all $\mathcal{H}_{I,(X Y)}(\rho'[p/\mathbf{v}])$ for which $I \in X$. This is of course by definition equal to $\mathcal{H}_{\cdot,(X Y)}(K\rho'[p/\mathbf{v}]) = (\psi'[p/\mathbf{v}])\langle X Y \rangle$.

It now suffices to show that this \mathbf{v} is equal to the way in which the modal literal $K\varphi$ is evaluated during the construction of $\psi\langle X Y \rangle$, i.e., that also $\psi\langle X Y \rangle = (\psi[K\varphi/\mathbf{v}])\langle X Y \rangle$. If $K\varphi \in^+ \psi$, then $K\varphi$ is evaluated in the belief pair $(X Y)$, i.e., it then suffices to show that $K\varphi\langle X Y \rangle = \mathbf{v} = \min_{J \in U}(\mathcal{H}_{J,(\cdot)}(p))$. On the other hand, if $K\varphi \in^- \psi$, then $K\varphi$ is evaluated in the belief pair $(Y X)$, i.e., it then suffices to show that $K\varphi\langle Y X \rangle = \mathbf{v} = \max_{J \in U}(\mathcal{H}_{J,(\cdot)}(p))$. It can now easily be checked from Lemma 4.2 that in both cases the needed equality holds. \square

We now have that, on the one hand, the (stable) fixpoints of $\mathcal{D}_{T'}$ coincide with those of $\tilde{\mathcal{D}}_{T'}$, while, on the other hand, the above lemma implies the correspondences between (stable) fixpoints of $\tilde{\mathcal{D}}_{T'}$ and \mathcal{D}_T , that were summarized in Section 2.2. This now allows us to relate the models of T' and T under various semantics.

4.2.4. Expansion, partial expansions and the Kripke-Kleene model

Because $\tilde{\mathcal{D}}_{T'}$ is part-to-part constant, Theorem 2.1 implies a correspondence between fixpoints of $\tilde{\mathcal{D}}_{T'}$ and \mathcal{D}_T . This gives the following result.

Theorem 4.3. A belief pair $(P S) \in \mathcal{B}_{\Sigma}$ is a partial expansion (respectively, the Kripke-Kleene model) of T iff there exists a belief pair $(P' S') \in \mathcal{B}_{\Sigma'}$ such that $(P' S')|_{\Sigma} = (P S)$ and $(P' S')$ is a partial expansion (the Kripke-Kleene model) of T' .

It is easy to see that for all $\begin{pmatrix} X & Y \\ U & V \end{pmatrix} \in \tilde{\mathcal{B}}_{\Sigma'}^*$, $\begin{pmatrix} X & Y \\ U & V \end{pmatrix}$ is exact iff $(X Y)$ is exact. Therefore, this correspondence between partial expansion also implies a correspondence between expansions.

$\psi \in F$	$K\varphi \in \psi$	F_p	part-to-part	part-to-whole	whole-to-part
+	+	$K\varphi \Rightarrow p$	✓	×	✓*
-	+	$K\varphi \Rightarrow p$	✓	✓	✓*
+	-	$K\varphi \Leftarrow p$	✓	×	×
-	-	$K\varphi \Leftarrow p$	✓	✓	×

(*): Holds only if φ is objective.

Figure 2. Monotonicity properties of $\tilde{\mathcal{D}}_{T'}$.

Theorem 4.4. A possible world structure $P \in \mathcal{W}_\Sigma$ is an expansion of T iff there exists a possible world structure $P' \in \mathcal{W}_{\Sigma'}$ such that $P'|_\Sigma = P$ and P' is an expansion of T' .

4.2.5. Extensions, partial extensions and the well-founded model

As we recall from Section 2.2, to get a correspondence between the stable and well-founded fixpoints of our operators, we need an additional monotonicity property. Concretely, $\tilde{\mathcal{D}}_{T'}$ needs to be either part-to-whole or whole-to-part monotone. We now investigate when this is the case.

Lemma 4.6. If $\psi \in^- F$, then $\tilde{\mathcal{D}}_{T'}$ is part-to-whole monotone.

Proof:

By symmetry of the operator $\tilde{\mathcal{D}}_{T'}$, it suffices to show that for all (X, Y) and $(U \ V) \leq (U' \ V')$, $\tilde{\mathcal{D}}_{T'}^u(\frac{X \ Y}{U \ V}) \leq \tilde{\mathcal{D}}_{T'}^u(\frac{X \ Y}{U' \ V'})$. Furthermore, because $\tilde{\mathcal{D}}_{T'}$ is already known to be part-to-part monotone, it suffices to show that $Mod(F'(\bar{\kappa}(\frac{X \ Y}{U \ V}))) \supseteq Mod(F'(\bar{\kappa}(\frac{X \ Y}{U' \ V'})))$. Because $\psi \in^- F$, this will be the case if $\psi' \langle \frac{Y \ X}{V \cdot} \rangle \leq \psi' \langle \frac{Y \ X}{V' \cdot} \rangle$. This now follows from $V \leq V'$. \square

Lemma 4.7. If $\psi \in^+ F$, $K\varphi \in^+ \psi$ and φ is an objective formula, then $\tilde{\mathcal{D}}_{T'}$ is whole-to-part monotone.

Proof:

By symmetry of $\tilde{\mathcal{D}}_{T'}$, it suffices to show that for all $(\frac{X \ Y}{U \ V}) \leq (\frac{X' \ Y'}{U' \ V'})$, $[\tilde{\mathcal{D}}_{T'}^u(\frac{X \ Y}{U \ V})] \leq [\tilde{\mathcal{D}}_{T'}^u(\frac{X' \ Y'}{U' \ V'})]$. This is the case if $Mod(F_p \langle X \ Y \rangle) \supseteq Mod(F_p \langle X' \ Y' \rangle)$. Because $F_p = (K\varphi \Rightarrow p)$, this will be the case if $(K\varphi) \langle Y \ X \rangle \leq (K\varphi) \langle Y' \ X' \rangle$. Because φ is objective, $(K\varphi) \langle Y \ X \rangle$ depends only on Y and $(K\varphi) \langle Y' \ X' \rangle$ depends only on Y' . The fact that $Y \leq Y'$ now implies that $(K\varphi) \langle Y \cdot \rangle \leq (K\varphi) \langle Y' \cdot \rangle$. \square

A summary of the monotonicity properties of $\tilde{\mathcal{D}}_{T'}$ can be found in Figure 2. By Theorem 4.2, we now obtain the following result.

Theorem 4.5. If at least one of these conditions is satisfied:

- $\psi \in^- F$ or
- $K\varphi \in^+ \psi$ and φ is objective,

then a belief $(P \ S) \in \mathcal{B}_\Sigma$ is a partial extension (respectively, the well-founded model) of T iff there exists a belief pair $(P' \ S') \in \mathcal{B}_{\Sigma'}$ such that $(P' \ S')|_\Sigma = (P \ S)$ and $(P' \ S')$ is a partial extension (the well-founded model) of T' . Moreover, under the same condition, a possible world structure $P \in \mathcal{W}_\Sigma$ is an extension of T iff there exists a possible world structure $P' \in \mathcal{W}_{\Sigma'}$ such that $P'|_\Sigma = P$ and P' is an extension of T' .

A final question that remains to be answered is what happens in the case where the above theorem is not applicable, i.e., when $\psi \in^+ F$ and $K\varphi \in^- \psi$. It turns out that in this case there is no correspondence. We demonstrate this by the following example.

Example 4.1. Let T be the theory $\{K\neg Kq\}$. If we replace Kq by p , we get $T' = \{K\neg p; Kq \Leftarrow p\}$.

Let us first look at the well-founded model of T . We start by applying the stable operator $\mathcal{C}_{\mathcal{D}_T}$ to the least precise pair $(\mathcal{I}_\Sigma \ \{\})$. To obtain a new underestimate P' , we construct $C_{\mathcal{D}_T}^\downarrow(\{\}) = \text{lfp}([\mathcal{D}_T(\cdot \ \{\})])$. We find that $C_{\mathcal{D}_T}^\downarrow(\{\}) = \mathcal{I}_\Sigma$, because:

$$[\mathcal{D}_T(\mathcal{I}_\Sigma, \{\})] = \text{Mod}(T\langle\{\} \ \mathcal{I}_\Sigma\rangle) = \text{Mod}(\mathbf{t}) = \mathcal{I}_\Sigma.$$

For the new overestimate, we have that $C_{\mathcal{D}_T}^\uparrow(\mathcal{I}_\Sigma) = \mathcal{I}_\Sigma$, because:

$$[\mathcal{D}_T(\mathcal{I}_\Sigma, \mathcal{I}_\Sigma)] = \text{Mod}(T\langle\mathcal{I}_\Sigma \ \mathcal{I}_\Sigma\rangle) = \text{Mod}(\mathbf{t}) = \mathcal{I}_\Sigma.$$

We therefore have that $\mathcal{C}_{\mathcal{D}_T}(\mathcal{I}_\Sigma \ \{\}) = (\mathcal{I}_\Sigma \ \mathcal{I}_\Sigma)$. Moreover, since by symmetry of \mathcal{D}_T , $C_{\mathcal{D}_T}^\downarrow = C_{\mathcal{D}_T}^\uparrow$, we now also see that $\mathcal{C}_{\mathcal{D}_T}(\mathcal{I}_\Sigma \ \mathcal{I}_\Sigma) = (\mathcal{I}_\Sigma \ \mathcal{I}_\Sigma)$. Therefore, this belief pair is the well-founded model of T , which is also its unique stable model.

We now perform a similar construction for T' . Firstly, $C_{\mathcal{D}_{T'}}^\downarrow(\{\}) = \mathcal{I}_\Sigma$, because:

$$[\mathcal{D}_{T'}(\mathcal{I}_\Sigma, \{\})] = \text{Mod}(T'\langle\{\} \ \mathcal{I}_\Sigma\rangle) = \text{Mod}(\mathbf{t}; \mathbf{t} \Leftarrow p) = \mathcal{I}_\Sigma$$

Secondly, $C_{\mathcal{D}_{T'}}^\uparrow(\mathcal{I}_\Sigma) = \{\}$, as can be seen from the following computation:

$$\begin{aligned} [\mathcal{D}_{T'}(\mathcal{I}_\Sigma \ \mathcal{I}_\Sigma)] &= \text{Mod}(\mathbf{f}; \mathbf{f} \Leftarrow p) = \{\} \\ [\mathcal{D}_{T'}(\mathcal{I}_\Sigma \ \{\})] &= \text{Mod}(\mathbf{f}; \mathbf{f} \Leftarrow p) = \{\} \end{aligned}$$

Therefore, the well-founded model of T' is $(\mathcal{I}_{\{p,q\}} \ \{\})$. The restriction of this to the original alphabet Σ is $(\mathcal{I}_\Sigma \ \{\})$, which does not coincide with the well-founded model of T . Moreover, the partial stable models of T' are $(\mathcal{I}_{\{p,q\}} \ \{\})$, $(\{\} \ \mathcal{I}_{\{p,q\}})$ and $(\{\{q\}, \{\}\} \ \{\{q\}, \{\}\})$, which do not correspond to those of T either.

As a side note, we remark that if we were to ignore our analysis of Section 4.1 and take F_p to be the formula $K\varphi \Rightarrow p$ instead of $K\varphi \Leftarrow p$, then we would not get a correspondence either. Indeed, it can easily be checked that the well-founded model of $\{K\neg p; Kq \Leftarrow p\}$ is also $(\mathcal{I}_\Sigma \ \{\})$.

The results of our analysis of predicate introduction for autoepistemic logic can now be summarized by the table in Figure 3.

$\psi \in F$	$K\varphi \in \psi$	F_p	(part.) expansion	K-K	(part.) extension	wfm
+	+	$K\varphi \Rightarrow p$	\checkmark^*	\checkmark^*	\checkmark^*	\checkmark^*
-	+	$K\varphi \Leftarrow p$	\checkmark	\checkmark	\checkmark	\checkmark
+	-	$K\varphi \Rightarrow p$	\checkmark	\checkmark	\times	\times
-	-	$K\varphi \Leftarrow p$	\checkmark	\checkmark	\checkmark	\checkmark

(*): Holds only if φ is objective.

Figure 3. Semantics preserved by replacing $K\varphi$ by p .

4.3. Discussion and related work

The nesting of modal operators is a source of computational complexity when evaluating autoepistemic theories in possible world structures, and therefore also when constructing models of such theories. Moreover, it also obscures the relation between this logic and other, related languages, such as logic programming and default logic. Indeed, both the Konolige transformation [6] from default logic into autoepistemic logic and, for instance, the transformations of logic programming into autoepistemic logic considered in [1] map into the fragment without nested modal operators. In [7], a transformation is presented that, at least under the semantics of expansions, can reduce any theory to an equivalent one that does not have such nestings. This transformation preserves the original alphabet of the theory, but might lead to an exponential blow-up in its size, since it uses the standard propositional normalization technique of distributing disjunction over conjunction. Our results on predicate introduction can be used to achieve the opposite effect of avoiding such a blow-up, at the expense of an increase in the alphabet. A simple algorithm that does this, would be the following. As long as there are formulas of K -rank at least 2, select a formula $K\varphi$ with maximum K -rank and replace this by a new atom, in the way previously described. This algorithm reduces a theory T to a theory T'' without nested K operators, whose size is linear in the size of the original theory. Our results show that T' is equivalent to T on the original alphabet of T under the semantics of expansions, partial expansions, and Kripke-Kleene semantics. For the semantics of (partial) extensions and the well-founded semantics, this result does, however, not hold. Indeed, here, our results do not give us a way of getting rid of nestings $K\varphi \in^- \psi$, where $\psi \in^+ F$.

Our analysis of the problem of predicate introduction in autoepistemic logic shows that our algebraic theorems also allow meaningful and useful results to be derived for this logic. Moreover, the algebraic concepts we have defined, i.e., those of fixpoint extension and part-to-part, part-to-whole, and whole-to-part monotonicity, have also proven to be useful analysis tools in this case. The use of these concepts reveals some interesting similarities to predicate introduction for logic programming, which might otherwise have gone unnoticed. Indeed, Figure 4 shows four cases of predicate introduction, in which, at the algebraic level, what happens in logic programming is precisely the same as what happens in autoepistemic logic. As such, the results of this section provide convincing evidence for the fact that our algebraic theory of fixpoint extensions is not only a convenient way of proving results for logic programming, but is also more widely applicable abstraction of a general knowledge representation principle.

Logic programming		Autoepistemic logic		Preserves stable fixpoints
Δ	Δ'	T	T'	
$\{R \leftarrow \neg R\}$	$\left\{ \begin{array}{l} R \leftarrow \neg P. \\ P \leftarrow R. \end{array} \right\}$	$\{KKr\}$	$\left\{ \begin{array}{l} Kp \\ Kr \Rightarrow p \end{array} \right\}$	✓
$\{R \leftarrow \neg R\}$	$\left\{ \begin{array}{l} R \leftarrow P. \\ P \leftarrow \neg R. \end{array} \right\}$	$\{\neg K\neg Kr\}$	$\left\{ \begin{array}{l} \neg Kp \\ Kr \Leftarrow p \end{array} \right\}$	✓
$\{R \leftarrow R\}$	$\left\{ \begin{array}{l} R \leftarrow P. \\ P \leftarrow R. \end{array} \right\}$	$\{\neg KKr\}$	$\left\{ \begin{array}{l} \neg Kp \\ Kr \Rightarrow p \end{array} \right\}$	✓
$\{R \leftarrow R\}$	$\left\{ \begin{array}{l} R \leftarrow \neg P. \\ P \leftarrow \neg R. \end{array} \right\}$	$\{K\neg Kr\}$	$\left\{ \begin{array}{l} Kp \\ Kr \Leftarrow p \end{array} \right\}$	×

Figure 4. Correspondences between logic programming and autoepistemic logic.

5. Conclusion

In a companion paper [10], we developed a theory of fixpoint extension in the algebraic framework of approximation theory and applied this to logic programming. In this paper, we have studied the application of these results to autoepistemic logic. Concretely, we examined a transformation to reduce the nesting depth of the modal operator K . We showed that, at the algebraic level, there are some remarkable parallels between the effects of this transformation and what happens in the case of logic programming. We were able to prove that this transformation preserves equivalence under the semantics of (partial) expansions and Kripke-Kleene semantics. Moreover, we also showed that, in a large number of cases, though not all, equivalence is also preserved under the well-founded semantics and the semantics of (partial) extensions. In summary, we have demonstrated that our abstract concepts, defined at the level of approximation theory, can be used to analyze the problem of predicate introduction and prove equivalence results for different non-monotonic logics and under different kinds of fixpoint semantics for these logics.

References

- [1] Bonatti, P.: Autoepistemic logics as a unifying framework for the semantics of logic programs, *Journal of Logic Programming*, **22**, 1995, 91–149.
- [2] Denecker, M., Marek, V., Truszczyński, M.: Fixpoint 3-valued semantics for autoepistemic logic, *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, MIT Press / AAAI-Press, 1998.

- [3] Denecker, M., Marek, V., Truszczyński, M.: Approximating operators, stable operators, well-founded fix-points and applications in nonmonotonic reasoning, in: *Logic-based Artificial Intelligence* (J. Minker, Ed.), chapter 6, Kluwer Academic Publishers, 2000, 127–144.
- [4] Denecker, M., Marek, V., Truszczyński, M.: Uniform semantic treatment of default and autoepistemic logics, *Artificial Intelligence*, **143**(1), January 2003, 79–122.
- [5] Konolige, K.: On the Relation between Default and Autoepistemic Logic, in: *Readings in Nonmonotonic Reasoning* (M. L. Ginsberg, Ed.), Kaufmann, Los Altos, CA, 1987, 195–226.
- [6] Konolige, K.: On the relation between default and autoepistemic logic, *Artificial Intelligence*, **35**, 1988, 343–382.
- [7] Marek, V. W., Truszczyński, M.: Autoepistemic Logic., *J. ACM*, **38**(3), 1991, 588–619.
- [8] Meyer, J.-J., van der Hoek, W.: *Epistemic Logic for Computer Science and Artificial Intelligence*, Cambridge University Press, 1995.
- [9] Moore, R.: Possible-World Semantics for Autoepistemic Logic, *Proc. of the Non-Monotonic Reasoning Workshop*, AAAI Press, Mohonk, N.Y, 1984.
- [10] Vennekens, J., Wittocx, J., Mariën, M., Denecker, M., Bruynooghe, M.: Predicate Introduction for Logics with Fixpoint Semantics. Part I: Logic Programming, *Fundamenta Informaticae*.