Distributed Autoepistemic Logic: Semantics, Complexity, and Applications to Access Control

Marcos Cramer
TU Dresden
Dresden, Germany
marcos.cramer@tu-dresden.de
Tel.: +49 351 463 38426

Pieter Van Hertum
pietervanhertum@gmail.com

Bart Bogaerts
Vrije Universiteit Brussel (VUB)
Brussels, Belgium
bart.bogaerts@vub.be

Marc Denecker
KU Leuven
Leuven, Belgium
marc.denecker@kuleuven.be

December 18, 2019

Abstract

In this paper we define and study a multi-agent extension of autoepistemic logic (AEL) called *distributed autoepistemic logic* (dAEL). We define the semantics of dAEL using approximation fixpoint theory, an abstract algebraic framework that unifies different knowledge representation formalisms by describing their semantics as fixpoints of semantic operators. We define 2- and 3-valued semantic operators for dAEL. Using these operators, approximation fixpoint theory allows us to define a class of semantics for dAEL, each based on different intuitions that are well-studied in the context of AEL. We define a mapping from dAEL to AEL and identify the conditions under which the mapping preserves semantics, and furthermore argue that when it does not, the dAEL semantics is more desirable than the AEL-induced semantics since dAEL manages to contain inconsistencies.

The development of dAEL has been motivated by an application in the domain of *access control*. We explain how dAEL can be fruitfully applied to this domain and discuss how well-suited the different semantics are for the application in access control.

1 Introduction

Access control is concerned with methods to determine which principal (i.e. user or program) has the right to access a resource, e.g. the right to read or modify a file. Many logics have been proposed for distributed access control [Abadi, 2003; Gurevich and Neeman, 2008; Abadi, 2008; Garg and Pfenning, 2012; Genovese, 2012]. Most of these logics use a modality k *says* indexed by a principal k. *says*-based access control logics are designed for systems in which different principals can issue statements that become part of the access control policy. $k \mathit{says} \varphi$ is usually rendered as "k supports φ", which can be interpreted to mean that k has issued statements that – together with additional information present in the system – imply φ. Different access control logics vary in their account of which additional information may be assumed in deriving the statements that k supports.

In Section 2, we argue that it is reasonable to assume that the statements issued by a principal are a complete characterization of what the agent supports. This is similar to the well-known "All I know"-assumption [Levesque, 1990] in autoepistemic logic (AEL) [Moore, 1985b; Denecker *et al.*, 2011], which states that an AEL theory is considered to be a complete characterization of what the agent knows. As such, one might wonder if AEL can be a suitable logic for representing access control policies. A first restriction that prohibits this type of applications is that AEL is designed to only model the state of mind of a *single agent*, while in the domain of access control, typically multiple agents are in play. An extension to AEL with multiple agents has been defined by Vlaeminck *et al.* [2012], but this extension requires a global stratification on the agents, i.e., an order on the agents, where agents higher in the order can only refer to knowledge/statements of agents lower in the order. This is undesirable for a distributed system; e.g., in Section 6 we present situations where such an order simply does not exist. Therefore, we extend AEL to a truly distributed multi-agent setting, and name our extension *distributed autoepistemic logic* (*dAEL*). We argue in Section 6 that the proposed extension provides a good formal model of the $says$-modality.

As the term "autoepistemic logic" suggests, AEL was designed to model (a single agent's) *knowledge*, including knowledge derived from reasoning about knowledge. However, the formalism of AEL can be applied to model other modalities too. Note that claims about an agent's knowledge are claims about that agent's internal state of mind. However, the formalism of AEL does not presuppose that its K modality represents an internal state of mind of an agent. For example, we can interpret the K modality to refer to the public commitments of an agent, i.e. interpret $K\phi$ to mean that the agent in question has publicly made statements that imply ϕ, and as such identify K with the $says$ modality. In what follows, we will keep the AEL terminology and refer to K as "knowledge" without thereby implying that it represents an internal state of mind.

In dAEL, agents have full (positive and negative) introspection into other agents' knowledge. This is of course an unreasonable assumption when the K modality represents an internal state of mind like actual knowledge. It is, however, reasonable when $K\phi$ is interpreted to mean that an agent has (publicly) issued statements that imply ϕ.

Section 2 gives some preliminary motivation for the design choices of dAEL based on the access control application that we have in mind. Section 3 contains preliminaries from AEL and approximation fixpoint theory. In Section 4, we first define the syntax of dAEL, then define a 2-valued and a 3-valued semantic operator for dAEL, and then show how approximation fixpoint theory can be applied to these operators to define a class of semantics for dAEL corresponding to equally-named, well-known semantics for AEL. In Section 5, we define a mapping from dAEL to AEL and show that for a subset of the logic defined by a consistency requirement, the mapping preserves all semantics. The class of theories in which the semantics coincide are, intuitively, those in which all agents have consistent knowledge; outside of this class, we show that our new logic dAEL manages to contain inconsistencies within a single agent, while in AEL this is impossible. Next, in Section 6, we discuss some use cases of applying dAEL to access control, and in Section 7, we study complexity of inference in our logic. After discussing related work in Section 8, we conclude the paper in Section 9 with remarks about possible topics for future work.

Publication History A preliminary version of this paper was presented at the IJCAI conference [Van Hertum *et al.*, 2016]. The current paper extends the previous work with examples, proofs, a detailed account of the semantical relationship between dAEL and AEL and a more detailed discussion of the applicability of dAEL to access control. In contrast to the conference paper, the current version no longer defines how the construct of inductive definitions can be incorporated into dAEL, as we found that it complicated the logic without being necessary for our applications.

2 Motivation

Before defining the syntax and semantics of dAEL, we discuss some general features of the $says$-based approach to access control to give some preliminary motivations for the formalism.

An *access control policy* is a set of norms defining which principal is to be granted access to which resource under which circumstances. Specialized logics called *access control logics* were developed for representing policies and access requests and reasoning about them. A general principle adopted by most logic-based approaches to access control is that access is granted if and only if it is logically entailed by the policy.

2.1 $says$-based access control logics and denial

There is a large variety of access control logics, but most of them use a modality $k\, says$ indexed by a principal k [Genovese, 2012]. $says$-based access control logics are designed for systems in which different principals can issue statements that become part of the access control policy. $k\, says\, \phi$ is usually explained informally to mean that k supports ϕ [Abadi, 2008; Garg and Pfenning, 2012; Genovese, 2012]; this means that k has issued statements that – together with additional information present in the system – imply ϕ. Different access control logics vary in their account of which rules of inference and which additional information may be used in deriving statements that k supports from the statements that k has explicitly issued. For instance, if $Alice$ issues the statement $(Bob\ says\ ok) \Rightarrow ok$ and Bob issues the statement ok, it is to be expected that also $Alice\ says\ ok$ holds.

Many state-of-the-art $says$-based access control logics, e.g., Binder Logic (BL) [Garg and Pfenning, 2012], are designed for application in a system based on *proof-carrying authorization* [Appel and Felten, 1999]. In such a system, an access request is always submitted together with a proof that establishes that the requester has access, and the task of the reference monitor is only to check the validity of this proof. If the proof is based on the assumption that some other principal k supports some formula ϕ, the proof will contain the signed certificate that establishes that k has made statements implying ϕ. In this way, assumptions of the form $k\ says\ \phi$ can be discharged. However, there is no way to discharge of assumptions of the form $\neg k\ says\ \phi$. If k has not made any statements implying ϕ, there is no way to prove this to the reference monitor by submitting some certificates issued by k, as the reference monitor can never be convinced that there are no other statements made by k, which have not been presented to the reference monitor. For this reason, many state-of-the-art $says$-based access control logics do not provide the means for deriving statements of the form $\neg k\ says\ \phi$ or $j\ says\ \neg k\ says\ \phi$ on the basis of the observation that k has not issued any statements that could imply ϕ.

However, precisely formulas of this form make it possible to model access denials naturally in a $says$-based access control logic, as illustrated in the following example.

Example 2.1. Suppose A is a professor with control over a resource r, B is a PhD student of A who needs access to r, and C is a postdoc of A supervising B. A wants to grant B access to r, but wants to grant C the right to deny B's access to r, for example in case B misuses her rights. A natural way for A to do this using the $says$-modality is to issue the statement $\neg C\ says\ \neg access(B, r) \Rightarrow access(B, r)$. This should have the effect that B has access to r unless C denies her access. However, this effect can only be achieved if our logic allows A to derive $\neg C\ says\ \neg access(B, r)$ from the fact that C has not issued any statements implying $\neg access(B, r)$. ▲

Such denials can be realized in a system in various ways: One way is to have a central server where all statements belonging to the access control policy are stored, independently of who has issued them. In this case, the reference monitor can confirm that C has not issued a statement implying $\neg access(B, r)$ and thus grant B access. This way the access control system is not truly distributed, even though the access control policy is still produced in a distributed way.

Such denials can also be realized in a truly distributed system if a certain degree of cooperativity of the principals with the reference monitor is assumed. Suppose for example that C does not want to deny B access right to r. In this case he will not issue any statement implying $\neg access(B, r)$. Furthermore, it is reasonable to assume that he will be cooperative with the reference monitor in this respect: If the reference monitor asks C whether he has issued statements implying $\neg access(B, r)$, he will say no. If C were not cooperative in this way, it would have the same effect as him stating $\neg access(B, r)$, which goes against his goal of not denying B access right. So given that the cooperativity needed here is in the interest of the concerning principals, we do not consider this cooperativity assumption problematic.

Note that the applicability of our logic does not depend on how precisely the access control system is realized in practice.

2.2 Autoepistemic Logic

The derivation of $\neg C\ says\ \neg access(B, r)$ described above, i.e., its derivation from the fact that C has not issued any statements implying $\neg access(B, r)$, is non-monotonic: If C later issues a statement implying $\neg access(B, r)$, the formula $\neg C\ says\ \neg access(B, r)$ can no longer be derived. In other words, adding a formula to the access control policy causes that something previously implied by the policy is no longer implied. Existing $says$-based access control logics are monotonic and hence they cannot support the type of reasoning described above for modelling denial with the $says$-modality.

In order to derive statements of the form $\neg k\ says\ \phi$, we have to assume the statements issued by a principal to be a complete characterization of what the principal supports. This is similar to the motivation behind Moore's autoepistemic logic (AEL) to consider an agent's theory to be a complete characterization of what the agent knows

[Moore, 1985b; Levesque, 1990; Niemelä, 1991; Denecker *et al.*, 2011]. This motivates an application of AEL to access control.

However, AEL cannot model more than one agent. In order to extend it to the multi-agent case, one needs to specify how the knowledge of the agents interacts. Most state-of-the-art access control logics allow $j\ says\ (k\ says\ \phi)$ to be derived from $k\ says\ \phi$, as this is required for standard delegation to be naturally modelled using the $says$-modality. In the knowledge terminology of AEL, this can be called mutual positive introspection between agents. In order to also model denial as described above, we also need mutual negative introspection, i.e., that $j\ says\ \neg k\ says\ \phi$ to be derived from $\neg k\ says\ \phi$.

2.3 Approximation Fixpoint Theory

Several semantics have been proposed for AEL. Approximation fixpoint theory (AFT) (see Subsection 3.2) is an algebraic framework that captures most of those. Furthermore, AFT provides us with a unified methodology for lifting AEL semantics to a distributed setting. What is required to apply AFT is to define semantic operators for our distributed version of AEL.

Among the many semantics induced by AFT, we find, in line with the claims from Denecker *et al.* [2011], the well-founded semantics to be best suited when applying dAEL to access control. Unlike other widely studied semantics of autoepistemic logic like the Moore's original expansion semantics, the Kripke-Kleene semantics and the stable semantics, the well-founded semantics is both *grounded* [Bogaerts *et al.*, 2015a], meaning that derivable formulas are supported by cycle-free justifications, and *constructive* [Denecker and Vennekens, 2007], meaning that the model can be characterized as the limit of a construction process. Both of these features are important for the access control application, as it means that agents can only access, or delegate control over a resource if there is a (non-cyclic) reason for it, and furthermore, that the provenance of this access can be traced back (by means of following the construction process). In Section 6 we discuss application scenarios that illustrate these advantages of the well-founded semantics over other semantics.

3 Formal Preliminaries

We assume familiarity with the basic concepts of first-order logic. We assume throughout this paper that a first-order vocabulary Σ is fixed, use $\mathbb{T}$ for the set of terms over Σ (which we call Σ-*terms*) and $\mathcal{L}$ for the language of standard first-order logic over Σ. Furthermore, we assume that Σ is the disjoint union of Σ_o and Σ_s, where Σ_o represents a set of *objective* symbols and Σ_s a set of *subjective* symbols. Symbols in Σ_o could for instance be arithmetic symbols, equality, or other symbols whose interpretation is shared among all involved agents. We assume that an infinite supply of variables is available and fixed throughout the paper. A variable assignment a assigns to each variable an object of a given domain. If x is a variable and d an element of the given domain, we use $a[x : d]$ for the variable assignment that assigns d to x and otherwise equals a. We consider the set of logical symbols of $\mathcal{L}$ to formally consist of $\wedge$, $\neg$ and $\forall$. The symbols $\vee$, $\Rightarrow$, $\Leftrightarrow$ and $\exists$ are, as usual, treated as abbreviations in the standard way:

$$\begin{aligned}(\varphi \vee \psi) &= \neg(\neg\varphi \wedge \neg\psi)\\ (\varphi \Rightarrow \psi) &= (\neg\varphi \vee \psi)\\ (\varphi \Leftrightarrow \psi) &= ((\varphi \Rightarrow \psi) \wedge (\psi \Rightarrow \varphi))\\ \exists x : \varphi &= \neg\forall x : \neg\varphi.\end{aligned}$$

Brackets may be dropped when this does not lead to ambiguity.

We use truth values $\mathbf{t}$ for truth, $\mathbf{f}$ for falsity and additionally, in a three-valued setting, we use $\mathbf{u}$ for unknown. The truth order $<_t$ on truth values is induced by $\mathbf{f} <_t \mathbf{u} <_t \mathbf{t}$. The precision order $<_p$ on truth values is induced by $\mathbf{u} <_p \mathbf{t}, \mathbf{u} <_p \mathbf{f}$. We define $\mathbf{t}^{-1} = \mathbf{f}, \mathbf{f}^{-1} = \mathbf{t}$ and $\mathbf{u}^{-1} = \mathbf{u}$.

3.1 Autoepistemic Logic

The language $\mathcal{L}_k$ of *autoepistemic logic* [Moore, 1985b][1] is defined recursively using the standard rules for the syntax of first-order logic, augmented with one modal rule. This syntax is standard in modal logics. The language is thus defined by

$$\begin{array}{ll} P(\bar{t}) \in \mathcal{L}_k & \text{if } P \text{ is an } n\text{-ary predicate in } \Sigma \text{ and } \bar{t} \text{ an } n\text{-tuple of terms} \\ (\varphi \land \psi) \in \mathcal{L}_k & \text{if } \varphi \in \mathcal{L}_k \text{ and } \psi \in \mathcal{L}_k \\ \neg\varphi \in \mathcal{L}_k & \text{if } \varphi \in \mathcal{L}_k \\ \forall x : \varphi \in \mathcal{L}_k & \text{if } \varphi \in \mathcal{L}_k \\ K\varphi \in \mathcal{L}_k & \text{if } \varphi \in \mathcal{L}_k \end{array}$$

An AEL theory T is a set of sentences (that is, formulas without free occurrences of variables) in $\mathcal{L}_k$. AEL uses the semantic concepts of standard modal logic. A *structure* is defined as usual in first-order logic. It formally represents a potential state of affairs of the world. We assume a domain D, shared by all structures, to be fixed throughout the paper. We also assume a Σ_o-structure I_o is fixed, representing the shared knowledge among all involved agents (currently, there is only one, but in the next section, there will be multiple agents). A *possible world structure* is a set of Σ-structures that coincide with I_o on all symbols of Σ_o. It contains all structures that are consistent with an agent's knowledge. Possible world structures are ordered with respect to the amount of knowledge they contain. Possible world structures that contain fewer structures possess more knowledge. Indeed, an agent A "knows" that a certain claim holds if this claim holds in all worlds A deems possible. Thus, in smaller possible world structures, more knowledge is present. Formally, given possible world structures Q_1 and Q_2, we define $Q_1 \leq_K Q_2$ to hold if and only if $Q_2 \subseteq Q_1$.

The semantics of AEL is based on the standard S5 truth assignment [Lewis and Langford, 1932; Hughes and Cresswell, 1996]. The *value* of a formula $\varphi \in \mathcal{L}_k$ with respect to a possible world structure Q, a structure I and a variable assignment a (denoted $\varphi^{Q,I,a}$) is defined using the standard recursive rules for first-order logic augmented with one additional rule for the modal operation. Formally, we define

$$\begin{array}{ll} (P(\bar{t}))^{Q,I,a} & = \begin{cases} \mathbf{t} \text{ if } \bar{t}^{I,a} \in P^I \\ \mathbf{f} \text{ otherwise} \end{cases} \\ (\neg\varphi)^{Q,I,a} & = (\varphi^{Q,I,a})^{-1} \\ (\varphi \land \psi)^{Q,I,a} & = \text{glb}_{\leq_t}(\varphi^{Q,I,a}, \psi^{Q,I,a}) \\ (\forall x : \varphi)^{Q,I,a} & = \text{glb}_{\leq_t}\{\varphi^{Q,I,a[x:d]} \mid d \in D\} \\ (K\varphi)^{Q,I,a} & = \begin{cases} \mathbf{t} & \text{if } \varphi^{Q,I',a} = \mathbf{t} \text{ for all } I' \in Q \\ \mathbf{f} & \text{otherwise} \end{cases} \end{array}$$

If φ is an AEL sentence, it is easy to see that $\varphi^{Q,I,a}$ is independent of a. In this case we use $\varphi^{Q,I}$ to denote this value.

Moore proposed to formalise the intuition that an AEL theory T expresses "all the agent knows" in semantic terms, as a condition on the possible world structure Q representing the agent's belief state. The condition is: a world I is possible according to Q if and only if I satisfies T given Q. Formally, Moore defines that Q is an *autoepistemic expansion* of T if for every world I, it holds that $I \in Q$ if and only if $T^{Q,I} = \mathbf{t}$.

The above definition is essentially a fixpoint characterisation. The underlying operator D_T maps Q to

$$D_T(Q) = \{I \mid T^{Q,I} = \mathbf{t}\}.$$

Autoepistemic expansions are exactly the fixpoints of D_T; they are the possible world structures that, according to Moore, express candidate belief states of an autoepistemic agent with knowledge base T.

[1] Technically, Moore only defined the propositional fragment of the logic we define below. Here, we are interested in a first-order variant of Moore's logic. Also the extension with *objective information* (the distinction between Σ_o and Σ_s) was not part of the original presentation.

Soon, researchers pointed out certain "anomalies" in the expansion semantics [Halpern and Moses, 1985; Konolige, 1988]. In the following years, many different semantics for AEL were proposed. It was only with the introduction of the abstract algebraical framework *approximation fixpoint theory* (AFT) that a uniform view on those different semantics was obtained. We define several of the semantics of AEL later, after introducing the algebraical preliminaries on AFT.

3.2 Approximation Fixpoint Theory

We recall the basics of lattice theory and approximation fixpoint theory by Denecker, Marek and Truszczyński [2000] (further shortened as DMT).

Lattices and Operators A *complete lattice* $\langle L, \leq \rangle$ is a set L equipped with a partial order $\leq$, such that every set $S \subseteq L$ has both a least upper bound and a greatest lower bound, denoted $\mathrm{lub}(S)$ and $\mathrm{glb}(S)$ respectively. A complete lattice has a least element $\bot$ and a greatest element $\top$. An operator $O : L \to L$ is *monotone* if $x \leq y$ implies that $O(x) \leq O(y)$. An element $x \in L$ is a *fixpoint* of O if $O(x) = x$. Every monotone operator O in a complete lattice has a least fixpoint, denoted $\mathrm{lfp}(O)$, which is the limit (i.e., the least upper bound) of the sequence x_α given by:

$$\begin{aligned}
&x_0 = \bot \\
&x_{\alpha+1} = O(x_\alpha) \\
&x_\lambda = \mathrm{lub}(\{x_\alpha \mid \alpha < \lambda\}), \text{ with } \lambda \text{ a limit ordinal}
\end{aligned}$$

Approximation Fixpoint Theory Given a lattice L, AFT makes use of the set L^2. We call elements of L^2 *approximations*. We define *projections* for pairs as usual: $(x, y)_1 = x$ and $(x, y)_2 = y$. Pairs $(x, y) \in L^2$ are used to approximate all elements in the interval $[x, y] = \{z \mid x \leq z \wedge z \leq y\}$. We call $(x, y) \in L^2$ *consistent* if $x \leq y$, i.e. if $[x, y]$ is non-empty. We use L^c to denote the set of consistent elements. Elements $(x, x) \in L^c$ are called *exact*. We identify a point $x \in L$ with the exact bilattice point $(x, x) \in L^c$. L^2 is equipped with a *precision order*, defined as $(x, y) \leq_p (u, v)$ if $x \leq u$ and $v \leq y$. If (u, v) is consistent, the latter means that (x, y) approximates all elements approximated by (u, v), or in other words that $[u, v] \subseteq [x, y]$. If L is a complete lattice, then so is $\langle L^2, \leq_p \rangle$.

AFT studies fixpoints of lattice operators $O : L \to L$ through operators approximating O. An operator $A : L^2 \to L^2$ is an *approximator* of O if it is $\leq_p$-monotone, and has the property that for all x, $O(x) \in [x', y']$, where $(x', y') = A(x, x)$. Approximators map L^c into L^c. As usual, we restrict our attention to *symmetric* approximators: approximators A such that for all x and y, $A(x, y)_1 = A(y, x)_2$. DMT [2004] showed that the consistent fixpoints of interest (supported, stable, well-founded) are uniquely determined by an approximator's restriction to L^c, hence, sometimes we only define approximators on L^c. Given an approximator A, we define the (complete) stable operator $S_A : L \to L : S_A(x) = \mathrm{lfp}(A(\cdot, x)_1)$, where $A(\cdot, y)_1$ denotes the operator $L \to L : x \mapsto A(x, y)_1$.

AFT studies fixpoints of O using fixpoints of A.

1. The *A-Kripke-Kleene fixpoint* is the $\leq_p$-least fixpoint of A. It approximates all fixpoints of O.
2. A *partial A-stable fixpoint* is a pair (x, y) such that $x = S_A(y)$ and $y = S_A(x)$.
3. An *A-stable fixpoint* of O is a fixpoint x of O such that (x, x) is a partial A-stable fixpoint.
4. The *A-well-founded fixpoint* is the least precise partial A-stable fixpoint.

The A-Kripke-Kleene fixpoint of O can be constructed by iteratively applying A, starting from $(\bot, \top)$. For the A-well-founded fixpoint, the following constructive characterisation has been worked out by Denecker and Vennekens [2007].

Definition 3.1. An *A-refinement* of (x, y) is a pair $(x', y') \in L^2$ satisfying one of the following two conditions:

- $(x, y) \leq_p (x', y') \leq_p A(x, y)$, or
- $x' = x$ and $A(x, y')_2 \leq y' \leq y$.

An A-refinement is *strict* if $(x, y) \neq (x', y')$.

Definition 3.2. A *well-founded induction* of A is a sequence $(x_i, y_i)_{i\leq\beta}$ with β an ordinal such that

- $(x_0, y_0) = (\bot, \top)$;
- (x_{i+1}, y_{i+1}) is an A-refinement of (x_i, y_i), for all $i < \beta$;
- $(x_\lambda, y_\lambda) = \mathrm{lub}_{\leq_p}\{(x_i, y_i) \mid i < \lambda\}$ for each limit ordinal $\lambda \leq \beta$.

A well-founded induction is *terminal* if its limit (x_β, y_β) has no strict A-refinements.

For an approximator A, there are many different terminal well-founded inductions of A. Denecker and Vennekens [2007] showed that they all have the same limit, and that this limit equals the A-well-founded fixpoint of O. If A is symmetric, the A-well-founded fixpoint of O (and in fact, every tuple in a well-founded induction of A) is consistent.

An alternative constructive characterisation of the A-well-founded fixpoint is the following. Denecker *et al.* [2000] also defined the four-valued stable operator $S^*_A : L^2 \to L^2$ by $S^*_A((x, y)) = (S_A(x), S_A(y))$. Then the A-well-founded fixpoint is the $\leq_p$-least fixpoint of S^*_A, so it can be constructed by a transfinite iterative application of S^*_A to $(\bot, \top)$ until a fixpoint is reached. We make use of this characterization of the A-well-founded fixpoint in some of the proofs that are included in the appendix.

Since the A-well-founded fixpoint is a consistent partial A-stable fixpoint, there always exists at least one consistent partial A-stable fixpoint. Furthermore, it easily follows from the definition of partial A-stable fixpoints that partial A-stable fixpoints are always fixpoints of A. These two properties together imply that if A has a unique consistent fixpoint, this is also the unique consistent partial A-stable fixpoint.

3.3 AFT and Autoepistemic Logic

DMT [1998] showed that many semantics from AEL can be obtained by direct applications of AFT. In order to do this, they defined a three-valued version of the semantic operator.

In order to approximate an agent's state of mind, i.e., to represent partial information about possible world structures, DMT [1998] defined a *belief pair* as a tuple (P, S) of two possible world structures. They say that a belief pair *approximates* a possible world structure Q if $P \leq_K Q \leq_K S$, or equivalently if $S \subseteq Q \subseteq P$. Intuitively, P is an underestimation or a *conservative* bound of the agent's knowledge, and S is an overestimation or *liberal* bound of the agent's knowledge. That is, P contains all interpretations that are potentially contained in the agent's possible world structure, and S all interpretations that are certainly contained in the agent's possible world structure. Stated even differently, P represents the knowledge the agent certainly has and S the knowledge the agent possibly has. We call a belief pair (P, S) *consistent* if $P \leq_K S$, i.e., if it approximates at least one possible world structure. From now on, we assume all belief pairs to be consistent. Belief pairs can be ordered by a precision ordering $\leq_p$: Given two belief pairs (P, S) and (P', S'), we say that (P, S) is less precise than (P', S') (notation $(P, S) \leq_p (P', S')$) if $P \leq_K P'$ and $S' \leq_K S$.

We now define a three-valued valuation of sentences with respect to a belief pair (which represents an approximation of the state of mind of an agent) and a structure, representing the state of the world.

Definition 3.3. The *value* of φ with respect to belief pair B, an interpretation I and a variable assignment a (notation $\varphi^{B,I,a}$) is defined inductively as follows:

$$(P(\bar{t}))^{B,I,a} = \begin{cases} \mathbf{t} & \text{if } \bar{t}^{I,a} \in P^I \\ \mathbf{f} & \text{otherwise} \end{cases}$$

$$(\neg\varphi)^{B,I,a} = (\varphi^{B,I,a})^{-1}$$

$$(\varphi \land \psi)^{B,I,a} = \mathrm{glb}_{\leq_t}(\varphi^{B,I,a}, \psi^{B,I,a})$$

$$(\forall x : \varphi)^{B,I,a} = \mathrm{glb}_{\leq_t}\{\varphi^{B,I,a[x:d]} \mid d \in D\}$$

$$(K\varphi)^{(P,S),I,a} = \begin{cases} \mathbf{t} & \text{if } \varphi^{(P,S),I',a} = \mathbf{t} \text{ for all } I' \in P \\ \mathbf{f} & \text{if } \varphi^{(P,S),I',a} = \mathbf{f} \text{ for some } I' \in S \\ \mathbf{u} & \text{otherwise} \end{cases}$$

$A \wedge B$		B		
		t	**f**	**u**
A	**t**	**t**	**f**	**u**
	f	**f**	**f**	**f**
	u	**u**	**f**	**u**

$A \vee B$		B		
		t	**f**	**u**
A	**t**	**t**	**t**	**t**
	f	**t**	**f**	**u**
	u	**t**	**u**	**u**

		$\neg A$
A	**t**	**f**
	f	**t**
	u	**u**

Figure 1: The Kleene truth tables [Kleene, 1938].

The logical connectives combine three-valued truth values based on Kleene's truth tables (see Figure 1). As before, in case φ is a sentence, $\varphi^{B,I,a}$ is independent of a and we use $\varphi^{B,I}$ for $\varphi^{B,I,a}$ for any a.

Let T be a fixed AEL theory. DMT [2000] defined the approximating operator D_T^* that maps a belief pair (P, S) to another belief pair (P', S') where

$$P' = \{I \mid T^{(P,S),I} \neq \mathbf{f}\} \text{ and } S' = \{I \mid T^{(P,S),I} = \mathbf{t}\}$$

Intuitively, the new conservative bound contains all worlds in which the theory evaluates to true (with the current knowledge) and the new liberal bound all worlds in which T does not evaluate to false. Thus P' contains all knowledge that can *certainly* be derived from the current state of mind and Q' all knowledge that can *possibly* be derived from it. DMT showed that D_T^* is an approximator of D_T. Hence, the operators induce a class of semantics for AEL: Moore's expansion semantics (supported fixpoints), Kripke-Kleene expansion semantics [DMT 1998] (Kripke-Kleene fixpoints), (partial) stable extension semantics ((partial) stable fixpoints) and well-founded extension semantics (well-founded fixpoints) [DMT 2003]. The latter two were new semantics induced by AFT.

4 dAEL: Syntax and Semantics

In this section, we describe the syntax and semantics of distributed autoepistemic logic. Theories in this logic describe the knowledge of a set of different agents. Throughout the rest of this paper, we assume a set of agents $\mathcal{A}$ to be fixed, with $\mathcal{A}$ a subset of the domain D over which all structures are defined. The reason for this assumption is that it allows us to reuse the quantifications from first-order logic to quantify over the set of agents at hand. Furthermore, we assume that for each agent $A \in \mathcal{A}$, there is a constant $A \in \Sigma_o$, interpreted as A in the objective structure I_o.

4.1 Syntax and Basic Semantic Notions

Definition 4.1. We define the language $\mathcal{L}_d$ of distributed autoepistemic logic recursively as follows.

$P(\bar{t}) \in \mathcal{L}_d$	if P is an n-ary predicate in Σ and $\bar{t}$ an n-tuple of terms
$(\varphi \wedge \psi) \in \mathcal{L}_d$	if $\varphi \in \mathcal{L}_d$ and $\psi \in \mathcal{L}_d$
$\neg\varphi \in \mathcal{L}_d$	if $\varphi \in \mathcal{L}_d$
$\forall x : \varphi \in \mathcal{L}_d$	if $\varphi \in \mathcal{L}_d$
$K_t(\psi) \in \mathcal{L}_d$	if $\psi \in \mathcal{L}_d$ and $t \in \mathbb{T}$

This definition consists of the standard recursive rules of first-order logic, augmented with a modal operator. The intuitive reading of $K_t(\psi)$ is "t is an agent and t knows ψ". Hence, if the term t does not denote an agent, $K_t(\psi)$ will be interpreted to be false.

We assume that Σ_o contains a dedicated unary predicate Agt, whose interpretation is assumed to always be $\mathcal{A}$.

In a distributed setting, different agents each have their own theory describing their beliefs or knowledge about the world:

Definition 4.2. A *distributed theory* is an indexed family $\mathcal{T} = (\mathcal{T}_A)_{A \in \mathcal{A}}$ where each $\mathcal{T}_A$ is a set of sentences in $\mathcal{L}_d$.

Example 4.3. Consider a situation with three agents, A, B and C who will vote openly on some issue. Agent A decides to vote yes if at least one of the other agents votes yes; otherwise he votes no. Agent B decides to follow the crowd: if all other agents are unanimous, she follows their vote. Otherwise, she abstains. Agent C decides to vote yes no matter what the other agents vote. The intended result of this vote is clear. C votes yes, hence A follows and in the end, the result is unanimous: every agent votes yes.

In dAEL, we model this situation as follows. We assume a single nullary predicate symbol yes and use $K_A\mathrm{yes}$ (respectively $K_A\neg\mathrm{yes}$) to denote the fact that agent A votes yes (respectively no). In this example, we thus use $K_A\varphi$ to denote public announcements (not knowledge) of agents. Consider the following three theories.

$$\begin{aligned}
\mathcal{T}_A &= \left\{\ (K_B\mathrm{yes} \lor K_C\mathrm{yes}) \Leftrightarrow \mathrm{yes}\ \right\}\\
\mathcal{T}_B &= \left\{ \begin{array}{l} (K_A\mathrm{yes} \land K_C\mathrm{yes}) \Rightarrow \mathrm{yes} \\ (K_A\neg\mathrm{yes} \land K_C\neg\mathrm{yes}) \Rightarrow \neg\mathrm{yes} \end{array} \right\}\\
\mathcal{T}_C &= \{\mathrm{yes}\}.
\end{aligned}$$

Now, $\mathcal{T} = (\mathcal{T}_A, \mathcal{T}_B, \mathcal{T}_C)$ is a distributed autoepistemic theory. As we shall show later (in Example 4.15), all semantics we define for dAEL agree on this theory. Furthermore, its unique model equals the intended model sketched above, i.e., it is such that $K_A\mathrm{yes}$, $K_B\mathrm{yes}$ and $K_C\mathrm{yes}$ are all true while $K_A\neg\mathrm{yes}$, $K_B\neg\mathrm{yes}$ and $K_C\neg\mathrm{yes}$ are all false. ▲

To represent the knowledge of multiple agents, we generalise the notion of a possible world structure:

Definition 4.4. A *distributed possible world structure (DPWS)* is an indexed family $\mathcal{Q} = (\mathcal{Q}_A)_{A\in\mathcal{A}}$, where $\mathcal{Q}_A$ is a possible world structure for each $A \in \mathcal{A}$.

The knowledge order can be extended pointwise to DPWSs. One DPWS contains more knowledge than another if each agent has more knowledge:

Definition 4.5. Given two DPWSs $\mathcal{Q}^1$ and $\mathcal{Q}^2$, we define $\mathcal{Q}^1 \leq_K \mathcal{Q}^2$ if $\mathcal{Q}^1_A \leq_K \mathcal{Q}^2_A$ for each $A \in \mathcal{A}$.

The value of a sentence is obtained like in AEL by evaluating each modal operator with respect to the right agent.

Definition 4.6. The *value* of a sentence φ with respect to a DPWS $\mathcal{Q}$, an interpretation I and a variable assignment a (denoted $\varphi^{\mathcal{Q},I,a}$) is defined inductively by the following recursive rules:

$$\begin{aligned}
(P(\bar{t}))^{\mathcal{Q},I,a} &= \left\{ \begin{array}{l} \mathbf{t} \text{ if } \bar{t}^{I,a} \in P^I \\ \mathbf{f} \text{ otherwise} \end{array} \right.\\
(\neg\varphi)^{\mathcal{Q},I,a} &= (\varphi^{\mathcal{Q},I,a})^{-1}\\
(\varphi \land \psi)^{\mathcal{Q},I,a} &= \mathrm{glb}_{\leq_t}(\varphi^{\mathcal{Q},I,a}, \psi^{\mathcal{Q},I,a})\\
(\forall x : \varphi)^{\mathcal{Q},I,a} &= \mathrm{glb}_{\leq_t}\{\varphi^{\mathcal{Q},I,a[x:d]} \mid d \in D\}\\
(K_t\varphi)^{\mathcal{Q},I,a} &= \left\{ \begin{array}{ll} \mathbf{t} & \text{if } t^{I,a} \in \mathcal{A} \text{ and} \\ & \quad \varphi^{\mathcal{Q},J,a} = \mathbf{t} \text{ for each } J \in \mathcal{Q}_{t^{I,a}} \\ \mathbf{f} & \text{otherwise} \end{array} \right.
\end{aligned}$$

As before, if φ is a sentence, $\varphi^{\mathcal{Q},J,a}$ is independent of a and we omit a in the notation.

In order to generalise this valuation to a partial setting, we define a generalisation of belief pairs.

Definition 4.7. A *distributed belief pair* is a pair $\mathcal{B} = (\mathcal{P}, \mathcal{S})$ of distributed possible world structures.

If $\mathcal{B} = (\mathcal{P}, \mathcal{S})$ is a distributed belief pair, we denote the conservative bound $\mathcal{P}$ as $\mathcal{B}^c$ and the liberal bound $\mathcal{S}$ as $\mathcal{B}^l$. Furthermore, we use $\mathcal{B}_A$ to denote the belief pair $(\mathcal{P}_A, \mathcal{S}_A)$. The precision order on the approximating lattice is defined as usual, in the following definition.

Definition 4.8. If $\mathcal{B}^1$ and $\mathcal{B}^2$ are two distributed belief pairs, we say that $\mathcal{B}^1$ is *less precise* than $\mathcal{B}^2$ if $\mathcal{B}^1_A \leq_p \mathcal{B}^2_A$ for each agent A. We denote this fact by $\mathcal{B}^1 \leq_p \mathcal{B}^2$.

The following proposition follows easily from the equivalent result in AEL.

Proposition 4.9. *The set of all DPWSs forms a complete lattice when equipped with the order $\leq_K$. The set of all distributed belief pairs forms a lattice when equipped with the order $\leq_p$.*

Proof. Follows immediately from the fact the order $\leq_K$ is the product order of the knowledge orders for each agent and the same for $\leq_p$.

□

As before, we restrict our attention to *consistent* distributed belief pairs. Note that for a set $\mathcal{S}$ of DPWSs,

$$\text{lub}_{\leq_K}(\mathcal{S}) = \left(\bigcap_{\mathcal{Q} \in \mathcal{S}} \mathcal{Q}_A \right)_{A \in \mathcal{A}}$$

The notion of three-valued valuations is extended to the distributed setting by evaluating each modal operator with respect to the correct agent.

Definition 4.10. The *value* of φ with respect to a distributed belief pair $\mathcal{B}$, an interpretation I and a variable assignment a (notation $\varphi^{\mathcal{B},I,a}$) is defined inductively as follows:

$$
\begin{aligned}
(P(\bar{t}))^{B,I,a} &= \begin{cases} \mathbf{t} & \text{if } \bar{t}^{I,a} \in P^I \\ \mathbf{f} & \text{otherwise} \end{cases} \\
(\neg\varphi)^{B,I,a} &= (\varphi^{B,I,a})^{-1} \\
(\varphi \wedge \psi)^{B,I,a} &= \text{glb}_{\leq_t}(\varphi^{B,I,a}, \psi^{B,I,a}) \\
(\forall x : \varphi)^{B,I,a} &= \text{glb}_{\leq_t}\{\varphi^{B,I,a[x:d]} \mid d \in D\} \\
(K_t\varphi)^{\mathcal{B},I,a} &= \begin{cases} \mathbf{t} & \text{if } t^{I,a} \in \mathcal{A} \text{ and} \\ & \quad \varphi^{\mathcal{B},J,a} = \mathbf{t} \text{ for all } J \in \mathcal{B}^c_{t^{I,a}} \\ \mathbf{f} & \text{if } t^{I,a} \notin \mathcal{A} \text{ or} \\ & \quad \varphi^{\mathcal{B},J,a} = \mathbf{f} \text{ for some } J \in \mathcal{B}^l_{t^{I,a}} \\ \mathbf{u} & \text{otherwise} \end{cases}
\end{aligned}
$$

Note that this definition differs from the recursive definition of the three-valued valuation of an AEL formula only in the the fifth rule.

This valuation essentially provides us with the means to apply AFT to lift the class of semantics of AEL to dAEL.

4.2 Semantics of dAEL through AFT

Recall that we assume that a structure I_o interpreting the domain and all symbols in Σ_o is fixed and hence shall not be repeated in all definitions. The two- and three-valued valuations form the building blocks to extend the semantic operator and its approximator from AEL to dAEL.

Definition 4.11. The knowledge revision operator for a distributed theory $\mathcal{T}$ is a mapping from the set of distributed possible world structures to itself, defined by

$$\mathcal{D}_{\mathcal{T}}(\mathcal{Q}) = (\{I \mid (\mathcal{T}_A)^{\mathcal{Q},I} = \mathbf{t}\})_{A \in \mathcal{A}}$$

This revision operator revises the knowledge of all agents simultaneously, given their current states of mind. Fixpoints represent states of knowledge of the agents that cannot be revised any further. Or, in other words, distributed possible world structures that are consistent with the theories of all agents.

Definition 4.12. The approximator for a distributed theory $\mathcal{T}$ on a distributed belief pair $\mathcal{B}$ is defined by $\mathcal{D}^*_{\mathcal{T}}(\mathcal{B}) = (\mathcal{D}^c_{\mathcal{T}}(\mathcal{B}), \mathcal{D}^l_{\mathcal{T}}(\mathcal{B}))$, where

$$
\begin{aligned}
\mathcal{D}^c_{\mathcal{T}}(\mathcal{B}) &= (\{I \mid (\mathcal{T}_A)^{\mathcal{B},I} \neq \mathbf{f}\})_{A \in \mathcal{A}}, \\
\mathcal{D}^l_{\mathcal{T}}(\mathcal{B}) &= (\{I \mid (\mathcal{T}_A)^{\mathcal{B},I} = \mathbf{t}\})_{A \in \mathcal{A}}.
\end{aligned}
$$

Theorem 4.13. *$\mathcal{D}^*_\mathcal{T}$ is an approximator of $\mathcal{D}_\mathcal{T}$.*

Proof. One can easily see from Definition 4.10 that the valuation $\mathcal{B} \mapsto (\mathcal{T}_A)^{\mathcal{B},I}$ is $\leq_p$-monotone. This implies that when $\mathcal{B} \leq_p \mathcal{B}'$, $\mathcal{D}^c_\mathcal{T}(\mathcal{B}) \supseteq \mathcal{D}^c_\mathcal{T}(\mathcal{B}')$ and $\mathcal{D}^l_\mathcal{T}(\mathcal{B}) \subseteq \mathcal{D}^l_\mathcal{T}(\mathcal{B}')$, i.e. $\mathcal{D}^c_\mathcal{T}(\mathcal{B}) \leq_K \mathcal{D}^c_\mathcal{T}(\mathcal{B}')$ and $\mathcal{D}^l_\mathcal{T}(\mathcal{B}) \geq_K \mathcal{D}^l_\mathcal{T}(\mathcal{B}')$, i.e. $\mathcal{D}^*_\mathcal{T}(\mathcal{B}) \leq_p \mathcal{D}^*_\mathcal{T}(\mathcal{B}')$. Thus $\mathcal{D}^*_\mathcal{T}$ is $\leq_p$-monotone.

The fact that $\mathcal{D}^*_\mathcal{T}$ coincides with $\mathcal{D}_\mathcal{T}$ on two-valued belief pairs, follows from the fact that if $\mathcal{B} = (\mathcal{Q}, \mathcal{Q})$, then $\mathcal{T}^{\mathcal{B},I}_A = \mathcal{T}^{\mathcal{Q},I}_A$. □

The stable operator $S_{\mathcal{D}^*_\mathcal{T}}$ is defined for dAEL as $S_{\mathcal{D}^*_\mathcal{T}}(\mathcal{Q}) = \mathrm{lfp}(\mathcal{D}^c_\mathcal{T}(\cdot, \mathcal{Q}))$. Different fixpoints of these operators lead to different semantics as discussed in Section 3.2;

Definition 4.14. Let $\mathcal{T}$ be a distributed theory.

- A *supported model* of $\mathcal{T}$ with respect to I_o is a fixpoint of $\mathcal{D}_\mathcal{T}$.
- The *Kripke-Kleene model* of $\mathcal{T}$ with respect to I_o is the $\leq_p$-least fixpoint of $\mathcal{D}^*_\mathcal{T}$.
- A *partial stable model* of $\mathcal{T}$ with respect to I_o is a distributed belief pair $\mathcal{B}$, such that $\mathcal{B}^c = S_{\mathcal{D}^*_\mathcal{T}}(\mathcal{B}^l)$ and $\mathcal{B}^l = S_{\mathcal{D}^*_\mathcal{T}}(\mathcal{B}^c)$.
- A *stable model* of $\mathcal{T}$ with respect to I_o is a DPWS $\mathcal{Q}$, such that $(\mathcal{Q}, \mathcal{Q})$ is a partial stable model of $\mathcal{T}$.
- The *well-founded model* of $\mathcal{T}$ with respect to I_o is the least precise partial stable model of $\mathcal{T}$.

We use the abbreviations Sup-*model*, KK-*model*, PSt-*model*, St-*model* and WF-*model* to refer to these five kinds of models respectively.

Example 4.15 (Example 4.3 continued)**.** As discussed before, the intended result is that all three agents vote yes. This intended result corresponds to the following DPWS:

$$(\{\{\text{yes}\}\}_A, \{\{\text{yes}\}\}_B, \{\{\text{yes}\}\}_C)\,.$$

Note that in this DPWS, $K_A$yes, $K_B$yes and $K_C$yes are all true, while $K_A\neg$yes, $K_B\neg$yes and $K_C\neg$yes are all false. We now show that this DPWS is indeed the only model of $\mathcal{T}$. For this we first establish that this DPWS viewed as an exact distributed belief pair is the only fixpoint of $\mathcal{D}^*_\mathcal{T}$.

Let $\mathcal{B}$ be any distributed belief pair. Then

$$\begin{aligned}
\mathcal{D}^c_\mathcal{T}(\mathcal{B})_C &= \{I \mid (\mathcal{T}_C)^{\mathcal{B},I} \neq \mathbf{f}\} \text{ (by Definition 4.12)}\\
&= \{I \mid \text{yes}^{\mathcal{B},I} \neq \mathbf{f}\}\\
&= \{I \mid I = \{\text{yes}\}\}\\
&= \{\{\text{yes}\}\}.
\end{aligned}$$

Similarly, $\mathcal{D}^l_\mathcal{T}(\mathcal{B})_C = \{\{\text{yes}\}\}$.

Now suppose $\mathcal{B}$ is a fixpoint of $\mathcal{D}^*_\mathcal{T}$. By the above, $\mathcal{B}_C = \{\{\text{yes}\}, \{\text{yes}\}\}$, i.e. $(K_C\text{yes})^\mathcal{B} = \mathbf{t}$ by Definition 4.10, i.e. $(K_B\text{yes} \vee K_C\text{yes})^\mathcal{B} = \mathbf{t}$. So for any fixpoint $\mathcal{B}$ of $\mathcal{D}^*_\mathcal{T}$, we get

$$\begin{aligned}
\mathcal{D}^c_\mathcal{T}(\mathcal{B})_A &= \{I \mid (\mathcal{T}_A)^{\mathcal{B},I} \neq \mathbf{f}\}\\
&= \{I \mid ((K_B\text{yes} \vee K_C\text{yes}) \Leftrightarrow \text{yes})^{\mathcal{B},I} \neq \mathbf{f}\}\\
&= \{I \mid \text{yes}^{\mathcal{B},I} \neq \mathbf{f}\} \text{ (since } (K_B\text{yes} \vee K_C\text{yes})^\mathcal{B} = \mathbf{t})\\
&= \{I \mid I = \{\text{yes}\}\}\\
&= \{\{\text{yes}\}\}.
\end{aligned}$$

Similarly, $\mathcal{D}^l_\mathcal{T}(\mathcal{B})_A = \{\{\text{yes}\}\}$. Now from this we get that $(K_A\text{yes})^\mathcal{B} = \mathbf{t}$, i.e. by the above we get that $(K_A\text{yes} \wedge K_C\text{yes})^\mathcal{B} = \mathbf{t}$. By a similar derivation as above, we can then conclude that $\mathcal{D}^c_\mathcal{T}(\mathcal{B})_B = \mathcal{D}^l_\mathcal{T}(\mathcal{B})_B = \{\{\text{yes}\}\}$, i.e. $\mathcal{B} = (\{\{\text{yes}\}\}_A, \{\{\text{yes}\}\}_B, \{\{\text{yes}\}\}_C)$. Thus $(\{\{\text{yes}\}\}_A, \{\{\text{yes}\}\}_B, \{\{\text{yes}\}\}_C)$ is indeed the only fixpoint of $\mathcal{D}^*_\mathcal{T}$.

Since it is the unique fixpoint of $\mathcal{D}^*_{\mathcal{T}}$, it is the KK-*model* and the unique PSt-*model* and thus also the WF-*model* of $\mathcal{T}$. Since this model is exact, it is also the unique Sup-*model* and St-*model* of $\mathcal{T}$. Thus, we see that in this example, all our semantics coincide with the intended model. ▲

Example 4.16. Suppose we have two agents, the mother and father of a six-year-old child: $\mathcal{A} = (M, D)$. A common situation is one where the child fancies candy and the father answers "You can have some candy if it is okay for mom", while the mother answers "You can have candy if your father says so". These statements can be modelled in dAEL as

$$\mathcal{T}_D = \{K_M(c) \Rightarrow c\} \qquad\qquad \mathcal{T}_M = \{K_D(c) \Rightarrow c\}.$$

There exist four possible world structures for each agent:

1. The empty possible world set or inconsistent belief: $\emptyset$, denoted as $\top$.
2. The belief of c: $\{\{c\}\}$, i.e., the fact that it follows from the public announcements made by the agent in question that the kid can have candy.
3. The disbelief of c: $\{\emptyset\}$, i.e., the fact that it follows from the public announcements made by the agent in question that the kid cannot have candy.
4. The lack of knowledge: $\{\emptyset, \{c\}\}$, denoted as $\bot$, i.e., the fact that no statements about being able to get candy follow from the announcements made by the agent in question.

The semantic operator associated to this theory is:

$$\begin{aligned}
\mathcal{D}_{\mathcal{T}}(\mathcal{Q}) &= (\{I \mid (\mathcal{T}_D)^{\mathcal{Q},I} = \mathbf{t}\}_D, \{I \mid (\mathcal{T}_M)^{\mathcal{Q},I} = \mathbf{t}\}_M) \text{ (by Definition 4.11)}\\
&= (\{I \mid (K_M(c) \Rightarrow c)^{\mathcal{Q},I} = \mathbf{t}\}_D, \{I \mid (K_D(c) \Rightarrow c)^{\mathcal{Q},I} = \mathbf{t}\}_M)\\
&= (\{I \mid (K_M(c))^{\mathcal{Q},I} = \mathbf{f} \text{ or } c^{\mathcal{Q},I} = \mathbf{t}\}_D, \{I \mid (K_D(c))^{\mathcal{Q},I} = \mathbf{f} \text{ or } c^{\mathcal{Q},I} = \mathbf{t}\}_M)\\
&= (\{I \mid \emptyset \in \mathcal{Q}_M \text{ or } I = \{c\}\}_D, \{I \mid \emptyset \in \mathcal{Q}_D \text{ or } I = \{c\}\}_M) \text{ (by Definition 4.6)}
\end{aligned}$$

Therefore

$$\mathcal{D}_{\mathcal{T}}(\mathcal{Q})_D = \begin{cases} \{c\} & \text{if } \emptyset \notin \mathcal{Q}_M;\\ \bot & \text{if } \emptyset \in \mathcal{Q}_M. \end{cases} \qquad\qquad \mathcal{D}_{\mathcal{T}}(\mathcal{Q})_M = \begin{cases} \{c\} & \text{if } \emptyset \notin \mathcal{Q}_D;\\ \bot & \text{if } \emptyset \in \mathcal{Q}_D. \end{cases}$$

From this it is obvious that there are two *supported models*, namely $(\{\{c\}\}_D, \{\{c\}\}_M)$ and $(\bot_D, \bot_M)$. In the first model, $K_D c$ and $K_M c$ hold, i.e. Mom and Dad agree that the kid can have candy. In the second model, $K_D c$, $K_D \neg c$, $K_M c$ and $K_M \neg c$ are all false, i.e. Mom and Dad do not make any claims about the kid being allowed or disallowed to have candy.

For computing the other semantics, we need to determine the approximator $\mathcal{D}^*_{\mathcal{T}}$. The first component of the approximator is:

$$\begin{aligned}
\mathcal{D}^c_{\mathcal{T}}(\mathcal{B}) &= (\{I \mid (\mathcal{T}_D)^{\mathcal{B},I} \neq \mathbf{f}\}_D, \{I \mid (\mathcal{T}_M)^{\mathcal{B},I} \neq \mathbf{f}\}_M)\\
&= (\{I \mid (K_M(c) \Rightarrow c)^{\mathcal{B},I} \neq \mathbf{f}\}_D, \{I \mid (K_D(c) \Rightarrow c)^{\mathcal{B},I} \neq \mathbf{f}\}_M)\\
&= (\{I \mid (K_M(c))^{\mathcal{B},I} \neq \mathbf{t} \text{ or } c^{\mathcal{B},I} \neq \mathbf{f}\}_D, \{I \mid (K_D(c))^{\mathcal{B},I} \neq \mathbf{t} \text{ or } c^{\mathcal{B},I} \neq \mathbf{f}\}_M)\\
&= (\{I \mid \emptyset \in \mathcal{B}^c_M \text{ or } I = \{c\}\}_D, \{I \mid \emptyset \in \mathcal{B}^c_D \text{ or } I = \{c\}\}_M)
\end{aligned}$$

Therefore

$$\mathcal{D}^c_{\mathcal{T}}(\mathcal{B})_D = \begin{cases} \{c\} & \text{if } \emptyset \notin \mathcal{B}^c_M;\\ \bot & \text{if } \emptyset \in \mathcal{B}^c_M. \end{cases} \qquad\qquad \mathcal{D}^c_{\mathcal{T}}(\mathcal{B})_M = \begin{cases} \{c\} & \text{if } \emptyset \notin \mathcal{B}^c_D;\\ \bot & \text{if } \emptyset \in \mathcal{B}^c_D. \end{cases}$$

Similarly, for the second component of the approximator we get:

$$\mathcal{D}^l_{\mathcal{T}}(\mathcal{B})_D = \begin{cases} \{c\} & \text{if } \emptyset \notin \mathcal{B}^l_M; \\ \bot & \text{if } \emptyset \in \mathcal{B}^l_M. \end{cases} \qquad \mathcal{D}^l_{\mathcal{T}}(\mathcal{B})_M = \begin{cases} \{c\} & \text{if } \emptyset \notin \mathcal{B}^l_D; \\ \bot & \text{if } \emptyset \in \mathcal{B}^l_D. \end{cases}$$

The *Kripke-Kleene model* can be computed by iterated applications of $\mathcal{D}^*_{\mathcal{T}}$ to $(\bot, \top)$ until a fixpoint is reached:

$$\mathcal{D}^*_{\mathcal{T}}(((\bot_D, \bot_M), (\top_D, \top_M))) = ((\bot_D, \bot_M), (\{c\}_D, \{c\}_M))$$
$$\mathcal{D}^*_{\mathcal{T}}(((\bot_D, \bot_M), (\{c\}_D, \{c\}_M))) = ((\bot_D, \bot_M), (\{c\}_D, \{c\}_M))$$

Thus $((\bot_D, \bot_M), (\{c\}_D, \{c\}_M))$ is the Kripke-Kleene model of $\mathcal{T}$. In this model $K_D c$ and $K_M c$ have truth-value $\mathbf{u}$, while $K_D \neg c$ and $K_M \neg c$ have truth value $\mathbf{f}$. So intuitively, it is undetermined whether Mom and Dad allow the kid to have candy, but they definitely do not disallow it.

Observation about $\mathcal{D}^c_{\mathcal{T}}$: The value of $\mathcal{D}^l_{\mathcal{T}}(\mathcal{B}^c, \mathcal{B}^l)$ depends only on $\mathcal{B}^c$; indeed, $\mathcal{D}^l_{\mathcal{T}}(\mathcal{B}^c, \mathcal{B}^l) = \mathcal{D}_{\mathcal{T}}(\mathcal{B}^c)$.

Now for any DPWS $\mathcal{Q}$,

$$\begin{aligned} S_{\mathcal{D}^*_{\mathcal{T}}} &= \text{lfp}(\mathcal{D}^c_{\mathcal{T}}(\cdot, \mathcal{Q})) \\ &= \text{lfp}(\mathcal{D}_{\mathcal{T}}) \text{ (by the above observation about } \mathcal{D}^c_{\mathcal{T}}) \\ &= (\bot_D, \bot_M) \end{aligned}$$

So the only *partial stable* model is $(\bot, \bot)$. This is therefore also the *well-founded model*. And since it is exact, it is also the only stable model. In this model, $K_D c$, $K_D \neg c$, $K_M c$ and $K_M \neg c$ are all false, i.e. Mom and Dad do not make any claims about the kid being allowed or disallowed to have candy. ▲

In the above example, it can be seen that supported models are very liberal in deriving knowledge, as knowledge may be supported by circular reasoning. For instance, the supported model $(\{\{c\}\}_D, \{\{c\}\}_M)$ essentially states that from the announcements of Mom and Dad, it follows that the kid is allowed to have candy. Whether this is a problematic interpretation in the case of this toy example may be debatable, but is is certainly not acceptable in access control: We do not want to allow access when the only reason to support the access is a circular justification that assumes that the access in question should be granted. Such circular justification could cause security problems! We will discuss an example of this in Section 6.

This criticism is similar to what has been said about Moore's original autoepistemic expansions (which are exactly the supported models). Other semantics, such as stable and well-founded semantics are more *grounded* [Bogaerts *et al.*, 2015a] in the sense that they derive only knowledge for which there is ground in the theory: knowledge is only derived if there is a non-self supporting reason. This is a reasonable way of deriving knowledge from the theories.

5 dAEL and AEL

We now describe a mapping from dAEL to AEL. We prove that for distributed theories that do not contain any inconsistency, the semantics for dAEL match the corresponding semantics for AEL. In the case of partially inconsistent distributed theories, the semantics do not coincide: dAEL allows for a single agent to have inconsistent beliefs, whereas AEL has no mechanism to encapsulate an inconsistency in a similar way. This capacity of dAEL to encapsulate inconsistencies is a desirable feature, for instance in access control, where it facilitates to *isolate* a faulty agent. We discuss this kind of isolation in detail in Section 6.1.

It has been noted before, by Vennekens *et al.* [2007a] and Vlaeminck *et al.* [2012] that natural embeddings of certain "stratified" languages in AEL fail when there is the possibility of inconsistent knowledge. They have presented the notion of *permaconsistent* theories as a criterion for their embeddings to work. In this section, we show

1. how to generalise permaconsistency to dAEL,
2. that for permaconsistent theories, our mapping indeed preserves semantics, and
3. that a weaker criterion (being universally consistent) works for supported, stable and partial stable semantics.

We first present a generalisation of the notion of permaconsistency to the distributed case.

Definition 5.1. A distributed theory $\mathcal{T}$ is *permaconsistent* if for each $A \in \mathcal{A}$ and each theory T' that can be constructed from $\mathcal{T}_A$ by replacing all occurrences of formulas $K_B\varphi$ not nested under a modal operator by $\mathbf{t}$ or $\mathbf{f}$, it holds that T' has at least one model that expands I_o.

The mapping from dAEL to AEL consists of a collection of translation functions, defined in Definitions 5.5 to 5.11. These functions translate syntactic constructs of dAEL like formulas and distributed theories and semantic constructs of dAEL like DPWSs and distributed belief pairs into the corresponding constructs of AEL. We denote each of these translation functions by τ with some subscript, where the subscript indicates the type of the output of the translation function.

Given a vocabulary Σ used for writing a distributed theory $\mathcal{T}$ in dAEL, the AEL translation of $\mathcal{T}$ will be written in a slightly modified vocabulary Σ':

Definition 5.2. Given a vocabulary Σ, we define Σ' to be the vocabulary consisting of

- all symbols in Σ_o,
- all symbols in Σ_s, but with an arity increased by one.

The additional argument of relation and function symbols in Σ_s refers to the agent whose beliefs about the relation/function symbol we are using to interpret the symbol. Given an n-ary function symbol $f \in \Sigma$, we will therefore interpret f as an $n{+}1$-ary function symbol in AEL, where $f(v_1, \ldots, v_n, a)$ should be interpreted as the interpretation of $f(v_1, \ldots, v_n)$ according to agent a. Our mapping from dAEL to AEL will be based on this intuition, namely in each formula, the extra argument will be used to represent the agent whose knowledge is referred to. Since we assume all functions to be total, $f(a_1, \ldots, a_n, a_{n+1})$ also needs to be interpreted when a_{n+1} is not an agent. Since it does not matter which value we give to $f(a_1, \ldots, a_n, a_{n+1})$ in this case (this will follow from our particular translation), we fix an arbitrary element δ in our domain D to assign to such defective terms.

Example 5.3 (Example 4.16 continued)**.** In this example, $\mathcal{A} = (M, D)$ and Σ consists of a proposition symbol $c/0$ and two constant symbols, namely, $M/0, D/0$, where M and D refer to mommy and daddy and have a fixed interpretation (i.e., $M, D \in \Sigma_o$). As such, Σ' consists of one unary predicate symbol $c/1$ and the same function symbols as Σ. The intended interpretation of $c(d)$ is that the child can have candy according to d. ▲

We use the following notational conventions in this section: ϕ denotes an $\mathcal{L}_d^{\Sigma}$-formula, φ denotes an $\mathcal{L}_k^{\Sigma'}$-formula, I denotes a Σ-structure, and J denotes a Σ'-structure.

Definition 5.4. Given a Σ-term t and a Σ'-term s, we define the Σ'-term t_s recursively as follows:

- $x_s := x$ for each variable x
- $(f(t_1, \ldots, t_n))_s := f(t_{1s}, \ldots, t_{ns}, s)$ for each $f \in \Sigma_s$
- $(f(t_1, \ldots, t_n))_s := f(t_{1s}, \ldots, t_{ns})$ for each $f \in \Sigma_o$

Definition 5.5. We define the function $\tau_{\mathit{formula}} : \mathbb{T}^{\Sigma'} \times \mathcal{L}_d^{\Sigma} \to \mathcal{L}_k^{\Sigma'}$ as follows:

- $\tau_{\mathit{formula}}(s, P(t_1, \ldots, t_n)) := P(t_{1s}, \ldots, t_{ns}, s)$
- $\tau_{\mathit{formula}}(s, \neg\phi) = \neg\tau_{\mathit{formula}}(s, \phi)$
- $\tau_{\mathit{formula}}(s, \phi \wedge \psi) = \tau_{\mathit{formula}}(s, \phi) \wedge \tau_{\mathit{formula}}(s, \psi)$
- $\tau_{\mathit{formula}}(s, \forall x : \phi) = \forall x : \tau_{\mathit{formula}}(s, \phi)$
- $\tau_{\mathit{formula}}(s, K_t\phi) = \exists x : (x = t_s \wedge \mathrm{Agt}(x) \wedge K\tau_{\mathit{formula}}(x, \phi))$ for a fresh variable x

Definition 5.6. For a distributed theory $\mathcal{T}$, we define $\tau_{\mathit{theory}}(\mathcal{T}) := \bigcup_{A\in\mathcal{A}} \tau_{\mathit{formula}}(A, \mathcal{T}_A)$.

Example 5.7 (Example 4.16 continued)**.** In this example,

$$\tau_{\textit{theory}}(\mathcal{T}) = \left\{ \begin{array}{l} (\exists a : a = M \land \mathrm{Agt}(a) \land Kc(a)) \Rightarrow c(D) \\ (\exists a : a = D \land \mathrm{Agt}(a) \land Kc(a)) \Rightarrow c(M) \end{array} \right\}$$

Given that M and D are in Σ_o, i.e. are have a fixed interpretation in all models, this AEL theory is equivalent to the following simpler one:

$$\tau_{\textit{theory}}(\mathcal{T}) = \left\{ \begin{array}{l} Kc(M) \Rightarrow c(D), \\ Kc(D) \Rightarrow c(M) \end{array} \right\}$$

Intuitively, this state that if Mom says candy is allowed, so does Dad and vice versa, i.e. this theory contains the same knowledge as the original example. ▲

Example 5.8. This example we illustrates why the AEL translation of $K_t\phi$ is $\exists x : (x = t_s \land \mathrm{Agt}(x) \land K\tau_{\textit{formula}}(x, \phi))$ and not the simpler formula $\mathrm{Agt}(t_s) \land K\tau_{\textit{formula}}(t_s, \phi)$. Consider the following dAEL theory $\mathcal{T}$:

$$\mathcal{T}_A = \{(d = A \land p) \lor (d = B \land \neg p)\}$$
$$\mathcal{T}_B = \{p\}$$

It can be easily verified that in all semantics defined in this paper, $\mathcal{T}$ has a unique model $\mathcal{B}$, and that $(K_A K_d p)^{\mathcal{B},I,a} = \mathbf{f}$ for all I, a. Note that $\mathcal{B}$ is exact, i.e. of the form $(\mathcal{Q}, \mathcal{Q})$, so what we just wrote about $K_A K_d p$ implies that $(K_d p)^{\mathcal{B},I,a} = \mathbf{f}$ for some $I \in \mathcal{Q}$ and some variable assignment a.

The AEL translation $\tau_{\textit{theory}}(\mathcal{T})$ of $\mathcal{T}$ is as follows:

$$\tau_{\textit{theory}}(\mathcal{T}) = \left\{ \begin{array}{l} (d(A) = A \land p(A)) \lor (d(A) = B \land \neg p(A)) \\ p(B) \end{array} \right\}$$

Again, all our semantics agree that this theory has a unique model B, and B is exact, i.e. of the form (Q, Q). Then $\tau_{\textit{formula}}(A, K_d p)^{\mathcal{B},I,a} = (\exists x : (x = d(A) \land \mathrm{Agt}(x) \land Kp(x)))^{B,I,a} = \mathbf{f}$ for some $I \in Q$ and some variable assignment a. This is in line with the above semantic analysis of $K_d p$. On the other hand $(\mathrm{Agt}(d(A)) \land Kp(d(A)))^{B,I,a} = \mathbf{t}$ for any I, a. This shows that the idea to use $\mathrm{Agt}(d(A)) \land Kp(d(A))$ as the AEL translation of $K_d p$ would not work. ▲

We now define a mapping of dAEL's semantic notions to AEL's semantic notions.

Definition 5.9. For an indexed family $\mathcal{I} = (I_A)_{A \in \mathcal{A}}$ of Σ-structures, we define the Σ'-structure $\tau_{\textit{structure}}(\mathcal{I})$ as follows: For each n-ary function symbol $f \in \Sigma_s$ and all $d_1, \ldots, d_n \in D'$,

$$f^{\tau_{\textit{structure}}(\mathcal{I})}(d_1, \ldots, d_n, d) := \begin{cases} f^{I_d}(d_1, \ldots, d_n) & \text{if } d \in \mathcal{A}; \\ \delta & \text{otherwise.} \end{cases}$$

For each n-ary function symbol $f \in \Sigma_o$ and all $d_1, \ldots, d_n \in D'$,

$$f^{\tau_{\textit{structure}}(\mathcal{I})}(d_1, \ldots, d_n) := f^{I_o}(d_1, \ldots, d_n).$$

For each n-ary relation symbol $R \in \Sigma_s$ and $d_1, \ldots, d_n \in D'$,

$$R^{\tau_{\textit{structure}}(\mathcal{I})}(d_1, \ldots, d_n, d) \text{ iff } d \in \mathcal{A} \text{ and } R^{I_d}(d_1, \ldots, d_n).$$

For each n-ary relation symbol $R \in \Sigma_o$ and $d_1, \ldots, d_n \in D'$,

$$R^{\tau_{\textit{structure}}(\mathcal{I})}(d_1, \ldots, d_n) \text{ iff } R^{I_o}(d_1, \ldots, d_n).$$

Definition 5.10. For a DPWS $\mathcal{Q}$, we define $\tau_{\textit{pws}}(\mathcal{Q}) := \{\tau_{\textit{structure}}((I_A)_{A \in \mathcal{A}}) \mid I_A \in \mathcal{Q}_A \text{ for every } A \in \mathcal{A}\}$.

Definition 5.11. For a distributed belief pair $\mathcal{B}$, define $\tau_{\textit{beliefpair}}(\mathcal{B}) := (\tau_{\textit{pws}}(\mathcal{B}^c), \tau_{\textit{pws}}(\mathcal{B}^l))$.

The above mapping from dAEL to AEL preserves all semantics in case $\mathcal{T}$ is permaconsistent.

Theorem 5.12. *Let $\sigma \in \{\mathit{Sup}, \mathit{KK}, \mathit{PSt}, \mathit{St}, \mathit{WF}\}$ be a semantics, let $\mathcal{T}$ be a permaconsistent distributed theory, and let $\mathcal{B}$ be a distributed belief pair. Then $\mathcal{B}$ is a σ-model of $\mathcal{T}$ iff $\tau_{\mathit{beliefpair}}(\mathcal{B})$ is a σ-model of $\tau_{\mathit{theory}}(\mathcal{T})$.*

The proof of this theorem as well as the two theorems below is in the appendix.

We also present a weaker criterion that preserves models for three out of the five semantics.

Definition 5.13. We call a DPWS $\mathcal{Q}$ *universally consistent* if $\mathcal{Q}_A \neq \emptyset$ for all $A \in \mathcal{A}$.

Definition 5.14. We call a distributed belief pair $\mathcal{B}$ *universally consistent* if $\mathcal{B}^l$ is universally consistent.

Note that since $\mathcal{B}^l \subset \mathcal{B}^c$, if $\mathcal{B}$ is universally consistent, so is $\mathcal{B}^c$.

Definition 5.15. Let $\sigma \in \{\mathsf{Sup}, \mathsf{KK}, \mathsf{PSt}, \mathsf{St}, \mathsf{WF}\}$ be a semantics. We call a distributed theory $\mathcal{T}$ *universally consistent under σ* iff every σ-model of $\mathcal{T}$ is universally consistent.

The following theorem states that the mapping from dAEL to AEL is faithful for universally consistent models of a distributed theory for three out of the five semantics.

Theorem 5.16. *Let $\sigma \in \{\mathit{Sup}, \mathit{PSt}, \mathit{St}\}$ be a semantics, let $\mathcal{T}$ be a distributed theory, and let $\mathcal{B}$ be a universally consistent distributed belief pair. Then $\mathcal{B}$ is a σ-model of $\mathcal{T}$ iff $\tau_{\mathit{beliefpair}}(\mathcal{B})$ is a σ-model of $\tau_{\mathit{theory}}(\mathcal{T})$.*

The next theorem clarifies the relationship between permaconsistency and universal consistency.

Theorem 5.17. *Let $\sigma \in \{\mathit{Sup}, \mathit{KK}, \mathit{PSt}, \mathit{St}, \mathit{WF}\}$. If T is permaconsistent, then T is universally consistent under σ.*

Example 5.18. The reverse of Theorem 5.17 does not hold as can be seen for example by a theory $\{p \Rightarrow K_A p, K_A p \Rightarrow p\}$ with one agent A. This theory is not permaconsistent because after replacing the first modal subformula by $\mathbf{f}$ and the second by $\mathbf{t}$, we get

$$p \Rightarrow \mathbf{f}, \mathbf{t} \Rightarrow p,$$

which clearly is not consistent. However, it is universally consistent under the 3 mentioned semantics. E.g., The unique stable model is $\{\{\}, \{p\}\}$ which is universally stable. ▲

Theorems 5.12, 5.16 and 5.17 are proven in the appendix.

6 Applying dAEL to Access Control

In this section, we discuss application scenarios of dAEL that illustrate the motivations from Section 2 and give reasons for our claim that the well-founded semantics is particularly suitable for an application in access control.

First, we show how a certain access control problem related to the revocation of delegated rights can be modelled in a natural and concise way in dAEL.

In ownership-based frameworks for access control, it is common to allow principals (users or processes) to grant both permissions and administrative rights to other principals in the system. Often it is desirable to grant a principal the right to further grant permissions and administrative rights to other principals. This may lead to delegation chains starting at a *source of authority* (the owner of a resource) and passing on certain permissions to other principals [Li *et al.*, 2003; Tamassia *et al.*, 2004; Chander *et al.*, 2004; Yao and Tamassia, 2009].

For simplicity, we assume access right and delegation right always go hand in hand. In that case, one can recursively define access right for a resource r as follows:

- The owner of r always has access for r.
- If a principal A with access right for r has granted an authorization for resource r to another principle B, then B has access right for r.

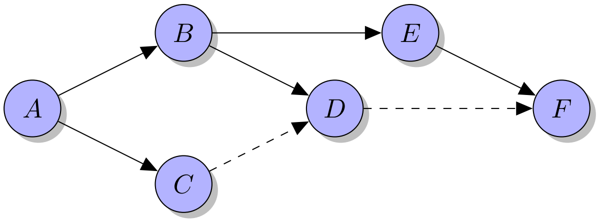

Figure 2: First example scenario of SGN. Full arrows represent delegations (positive authorizations), dashed arrows revocations (negative authorizations). A is the owner of the resource in question.

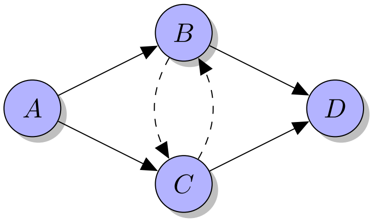

Figure 3: Second example scenario of SGN. Full arrows represent delegations, dashed arrows revocations. A is the owner of the resource in question.

Equivalently, one can say that a principle A has access right for r if there is a chain of authorizations for r starting in the owner of r and ending in principal A.

Furthermore, such frameworks commonly allow a principal to revoke a permission that she granted to another principal [Hagström *et al.*, 2001; Zhang *et al.*, 2003; Chander *et al.*, 2004; Barker *et al.*, 2014]. Depending on the reasons for the revocation, different ways to treat the delegation chain can be desirable [Hagström *et al.*, 2001; Cramer *et al.*, 2015; Cramer and Casini, 2017]. Any algorithm that determines which permissions to keep intact and which ones to delete when revoking a permission is called a *revocation scheme*. Of these revocation schemes, the one with the strongest effect is called the *Strong Global Negative* (SGN) revocation scheme: In this scheme, revocation is performed by issuing a negative authorization which dominates over positive authorizations and whose effect propagates forward.

Semi-formally, the effect of an SGN revocation can be characterized recursively as follows:

- The owner of r always has access for r.
- If a principal A with access right for r has issued a positive authorization for resource r to another principle B and no principal with access right for r has issued a negative authorization for r to B, then B has access right for r.

We illustrate the effect of SGN revocations the example depicted in Figure 2: In this example, A has issued positive authorizations to B and C, B has issued positive authorizations to D and E, E has issued a positive authorization to F, C has issued a negative authorization to D, and D has issued a negative authorization to F. Since A is the owner of the resource, A certainly has access by the first bullet item in the above semi-formal characterization of SGN. By the second bullet point, B and C have access, as A has issued positive authorizations to them and no one has issued a negative authorization to them. Similarly, since E have access, since B has issued a positive authorization to E, an no one has issued a negative authorization to E. Since C has access right and has issued a negative authorization to D, D certainly does not have access right despite the positive authorization issued to D by B. And since D does not have access, the negative authorization from D to F has no effect, so the access that E has granted to F takes effect.

In this example, the semi-formal characterization of SGN leads to clear results about who has acces and who does not. But this is not always the case. Consider for example the situation depicted in Figure 3.

Here B is attempting to revoke C's access right (and vice versa). According to the above characterization of the effect of an SGN revocation, this attempt is only successful if B has access. In other words, C should have access if and only if B does not have access. But since the scenario is symmetric between B and C, they should either both be granted or both be denied access right. However, this cannot be achieved without violating the above characterization of SGN revocation. Paradoxical situations like this one can arise whenever the authorization graph contains a cycle that contains at least one negative authorization (one revocation).

Existing papers that have covered SGN revocation have handled this issue in different ways:

- In the revocation framework of Hagström *et al.* [2001], this problem only arises when their *Strong Global Negative revocation* is combined with a *negative-takes-precedence* conflict resolution policy. In their paper, they do not describe in detail how Strong Global Negative revocation is supposed to work in the context of a *negative-takes-precedence* conflict resolution policy. In other words, their paper implicitly implies the existence of such problematic scenarios, but does not expicitly discuss them.

- The paper by Cramer *et al.* [2015] is the first one to explicitly discuss this problem. The problem is circumvented by disallowing problematic authorization graphs (basically any authorization graph with a cycle that contains at least one negative authorization).

- The paper by Cramer and Casini [2017] has a two-part inductive definition (Definitions 3 and 4) that directly corresponds to our above characterization of SGN revocation. In a footnote the paper specifies that this inductive definition is to be interpreted using the well-founded semantics for inductive definitions [Denecker, 1998]. The paper points out that there exist paradoxical cases in which the well-founded model of the inductive definition is three-valued rather than two-valued, so that for some principals it may be undecided whether they have access or not. The paper stipulate that in such cases *undecided* is to be treated in the same way as *false*, so that the principals directly affected by such a paradoxical situation will not have access until the paradoxical situation is resolved. (Applied to our above example this approach implies that formally the access right of B, C and D is *undefined*, which practically means access gets denied for all three of them.)

We will now model delegation and SGN revocation in dAEL. When interpreted with the well-founded semantics for dAEL, this model of delegation and revocation is equivalent to the formalization by Cramer and Casini [Cramer and Casini, 2017]. Furthermore, we will motivate why this behavior of SGN revocation corresponds better to general access control principles than the behavior that one would get if one used dAEL with another semantics than the well-founded semantics.

Our dAEL model of delegation and SGN revocation is based on statements issued by the various principals involved in a system: A principal k can delegate access right to a principal j by issuing the statement $deleg\_to(j)$, and can revoke access right from j by issuing the statement $revoke(j)$. We assume that the owner A of the resource wants to ensure that these delegation and revocation statements are interpreted in line with our above characterization of SGN revocation. The owner A can achieve this by issuing the following statements as part of its theory (together with the $deleg\_to$ and $revoke$ statements that A makes):

$$access(A, r)$$

$$(\exists k\ (K_A access(k, r) \land K_k deleg\_to(j)) \land \neg\exists i\ (K_A access(i, r) \land K_i revoke(j))) \Rightarrow access(j, r)$$

Now access is to be granted to a principal k if and only if the owner A believes the statement $access(k, r)$, i.e., if and only if $K_A access(k, r)$ holds in the well-founded model of the distributed theory given by the above base theory of owner A and all the statements issued by the various principals in the system.

We illustrate this using our first example scenario from Figure 2. Given that A is the owner of the resource in question, the distributed theory that represents the authorizations present in this example scenario is as follows:

$$\mathcal{T}_A = \left\{ \begin{array}{l} access(A,r) \\ (\exists k\ (K_A\, access(k,r) \wedge K_k\, deleg\_to(j)) \wedge \neg\exists i\ (K_A\, access(i,r) \wedge K_i revoke(j)))\ \Rightarrow access(j,r) \\ deleg\_to(B) \\ deleg\_to(C) \end{array} \right\}$$

$$\mathcal{T}_B = \left\{ \begin{array}{l} deleg\_to(D) \\ deleg\_to(E) \end{array} \right\}$$

$$\mathcal{T}_C = \left\{\ revoke(D)\ \right\}$$

$$\mathcal{T}_D = \left\{\ revoke(F)\ \right\}$$

$$\mathcal{T}_E = \left\{\ deleg\_to(F)\ \right\}$$

$$\mathcal{T}_F = \{\}$$

Let $\mathcal{B}_{\mathsf{WF}}$ be the well-founded model of $\mathcal{T}$. By formalizing the informal reasoning about this example that we presented above, one can show that $\mathcal{B}_{\mathsf{WF}}$ assigns $\mathbf{t}$ to the statements $K_A\, access(A,r)$, $K_A\, access(B,r)$, $K_A\, access(C,r)$, $K_A\, access(E,r)$, $K_A\, access(F,r)$, while it assigns $\mathbf{f}$ to the statement $K_A\, access(D,r)$. Therefore A, B, C, E and F will be granted access to resource r and D will be denied access to it.

In this application of dAEL, the information that we are interested in from a given model is only the information about which truth-values the model assigns to statements of the form $K_A\, access(X,r)$, i.e. which agents are given access to the recourse by the owner (A). For this reason, we present the relevant information as a set of expressions X^t where X is a principal and t the truth value of $K_A\, access(X,r)$ in the model. So in the above example, we would say that the well-founded model $\mathcal{B}_{\mathsf{WF}}$ satisfies $\{A^{\mathbf{t}}, B^{\mathbf{t}}, C^{\mathbf{t}}, D^{\mathbf{f}}, E^{\mathbf{t}}, F^{\mathbf{t}}\}$.[2]

For the above example, the other semantics presented in Section 4.2 give the same results as the well-founded semantics. However, that is not always the case. In the cases when the various semantics differ, the well-founded semantics is the only one that ensures that decisions about access are *grounded* [Bogaerts *et al.*, 2015a], meaning that derivable formulas are supported by cycle-free justifications, and that they satisfy a security principle that has been worded by Garg [2009] as follows: "When access is granted to a principal k, it should be known where k's authority comes from". For this reason, we consider the well-founded semantics to be preferable to the other semantics for applications of dAEL to access control. We will now illustrate these desirable feature of the well-founded semantics through two example scenarios.

First, let us consider again the scenario depicted in Figure 3. Here A is the owner of the resource r and has issued the statements $deleg\_to(B)$ and $deleg\_to(C)$, B has issued the statements $revoke(C)$ and $deleg\_to(D)$, and that C has issued the statements $revoke(B)$ and $deleg\_to(D)$. As explained above, attempting to apply the semi-formal characterization of SGN revocation to this scenario leads to paradoxical arguments about the access rights of B and C. So we may say that the scenario contains a conflict that cannot be automatically resolved. At this point, A as the principal with control over r will have to manually resolve the conflict by removing access from at least one of B and C depending on the cause for the conflict between them.

In practice, it may take A some time to study the situation and perform this manual resolution. During this time, the system should still respond to access requests. The intended behavior is that neither B nor C should have access, to avoid security risks. The situation for D is less clear: Given that D would have access no matter who of B and C has access, one could make a case for granting D access in this situation.

However, granting access right to D would violate the security principle mentioned above: "When access is granted to a principal k, it should be known where k's authority comes from" [Garg, 2009].

Now consider the statements issued by the principals as a distributed theory $\mathcal{T}$, with the two statements governing access included in the theory of the resource owner A. This theory has different models depending on the choice of semantics. There are two supported models satisfying $\{A^{\mathbf{t}}, B^{\mathbf{t}}, C^{\mathbf{f}}, D^{\mathbf{t}}\}$ and $\{A^{\mathbf{t}}, B^{\mathbf{f}}, C^{\mathbf{t}}, D^{\mathbf{t}}\}$ respectively.

[2] Note that this presentation of a model does not present all the information that is in the model, only the one that is relevant for our discussion about the access control application presented here. But in fact, this information suffices for figuring out the entire models, for more details on this, see the proof of Theorem 7.6

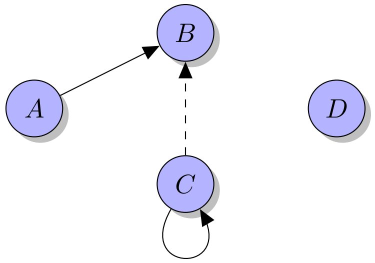

Figure 4: Illustration of the third scenario for SGN. Full arrows represent delegations, dashed arrows revocations. A is the owner of the resource in question.

These are also the stable models. The Kripke-Kleene model and the well-founded model are identical and satisfy $\{A^{\mathbf{t}}, B^{\mathbf{u}}, C^{\mathbf{u}}, D^{\mathbf{u}}\}$. This model is not exact: the truth-value of the statements $K_A\, access(X, r)$, with $X \in \{B, C, D\}$ is unknown.

When there is more than one model, the only safe approach in the access control application is to merge the information from the multiple models in a skeptical way, i.e. to grant access only if each model justifies granting access. According to this principle, the supported model semantics and stable semantics lead to access being granted to A and D in this scenario. Given our above argument against granting access to D, this means that these semantics cannot be considered viable semantics for this application of dAEL. Furthermore, note that when the skeptical way of combining information from multiple models is applied to the partial stable semantics, the result is always the same as the result of the well-founded semantics. For this reason, we do not consider the partial stable semantics separately in this section.

The Kripke-Kleene and well-founded model of this theory gives access precisely to the principal that should have access according to our above discussion. Furthermore, it exhibits the existing conflict between B and C by making their access right status undefined.

Now consider a third scenario as depicted in Figure 4, in which the resource owner A has issued the statement $deleg\_to(B)$ and C has issued the statements $deleg\_to(C)$ and $revoke(B)$. Here C should clearly not have access, because the only principal granting her access is C herself. Hence C's revocation of B's access right does not have any effect, so B should be granted access. The Kripke-Kleene model of the distributed theory corresponding to this scenario is not exact; it satisfies $\{A^{\mathbf{t}}, B^{\mathbf{u}}, C^{\mathbf{u}}, D^{\mathbf{f}}\}$. In this model, it is unknown whether B and C have access; this clearly diverges from our requirements. The well-founded model on the other hand correctly computes this desired outcome: it satisfies $\{A^{\mathbf{t}}, B^{\mathbf{t}}, C^{\mathbf{f}}, D^{\mathbf{f}}\}$. The reason why the well-founded semantics leads to a better outcome than the Kripke-Kleene semantics is that it is *grounded* [Bogaerts *et al.*, 2015a], meaning that derivable formulas are supported by cycle-free justifications.

From these scenarios, we can see that the only semantics for dAEL that behaves as desired in the access control application is the well-founded semantics. These findings are in line with the findings of Denecker *et al.* [2011], who strongly argued in favour of the well-founded semantics of AEL.

6.1 Faulty agents

In a distributed setting, it can happen that one of the agents either deliberately or accidentally *fails*, i.e. has a theory that – together with additional information present in the system – implies a contradiction. It is to be expected that such a failure has at least some influence on the rest of the system. However, the hope is that the rest of the system does not suffer too much from a failure of a single agent. In this section, we show how dAEL isolates faulty agents and contrast it to what happens when standard AEL is used to model a multi-agent scenario using the translation from Section 5. Consider again the principals from example 2.1. A is a professor with theory

$$T_A = \left\{ \begin{array}{l} access(A, r). \\ access(C, r). \\ \neg K_C \neg access(B, r) \Rightarrow access(B, r) \end{array} \right\}.$$

Now, first we will consider a situation in which the PhD student B is faulty (making inconsistent claims). I.e., consider the following theories

$$T_B^1 = \{access(B, r) \wedge \neg access(B, r)\}, \quad T_C^1 = \{\}$$

and the distributed autoepistemic theory

$$T^1 = (T_A, T_B^1, T_C^1).$$

Under all of the semantics we defined for dAEL, a model

$$(Q_A, Q_B, Q_C)$$

will have the property that $Q_B = \top$, i.e., that the agent B has inconsistent knowledge. This is to be expected since the theory T_B^1 that describes his knowledge is inconsistent. The interesting thing to investigate is how this inconsistency affects the other agents' knowledge. Luckily, it doesn't! In this example supported, Kripke-Kleene, well-founded, partial stable and stable semantics all agree that the unique model is given by

$$\begin{aligned}
&Q_A = \{\{access(A, r), access(C, r), access(B, r)\}\} \\
&Q_B = \top \\
&Q_C = \bot
\end{aligned}$$

That is, A knows that everyone can access resource r, C makes no claims about access to resources, while B has inconsistent knowledge. This example is one of the situations where the mapping from dAEL to AEL does *not* preserve semantics. Indeed, AEL has no mechanisms to isolate inconsistencies. If an AEL theory contains an inconsistency, this always results in a globally inconsistent possible world structure.

Now, let us consider another variation of the same theory, namely with

$$T_B^2 = \{\}, \quad T_C^2 = \{access(B, r) \wedge \neg access(B, r)\}$$

and the distributed autoepistemic theory

$$T^2 = (T_A, T_B^2, T_C^2).$$

I.e., we now consider what happens if C is a faulty agent. Now, all semantics for dAEL agree that the unique model is given by

$$\begin{aligned}
&Q_A = \{\ \{access(A, r), access(C, r)\}, \{access(A, r), access(B, r), access(C, r)\}\ \} \\
&Q_B = \bot \\
&Q_C = \top
\end{aligned}$$

That is, A knows that A and C have access to the resource r and that it does not follow that B has access. In this example, C has inconsistent knowledge and B has no knowledge. Thus, in this case, we can see that the inconsistency in C's theory *does* influence the knowledge of other agents. Indeed, A observes that $K_C \neg access(B, r)$ holds and thus the last constraint no longer entails access of B. However, it only influence knowledge of other players at places where they explicitly refer to the faulty agent. If there were a fourth agent, say another postdoc D and T_A also contains the constraint $access(D, r)$, the result would be that A grants D access, regardless of inconsistencies in other agent's knowledge. Thus, we conclude that our proposed formalism manages to isolate faulty agents as desired.

This desirable behaviour with respect to faulty agents is the same behaviour that other $says$-based acess control logic such as BL [Garg and Pfenning, 2012] exhibit. The point we made in this subsection is that if one tried to use standard AEL as an access control logic by modelling the multi-agent features using the translation from Section 5, one would not get this desirable behaviour concerning faulty agents, so that the extension of AEL to dAEL is really necessary for the access control application.

7 Complexity

We now study complexity of reasoning in dAEL. Given the argumentation in the previous section, we focus on the well-founded semantics. More particularly, we are concerned with the the following decision task:

Task 7.1. *Given a finite set of agents $\mathcal{A}$, a finite Σ_o-structure I_o with domain D, a finite dAEL theory T and a sentence φ of the form $K_t\psi$ with t a Σ_o term, determine if φ holds in the well-founded model of T.*

This task is well-defined: since t is a Σ_o term, φ can be evaluated in a belief pair. Stated in words, we are interested in evaluating whether a certain formula (ψ) holds in the knowledge of a certain agent (represented here by the term t). In the context of access control, this decision problem is indeed the one we are interested in: there the formula is ψ typically of the form $acces(b, r)$ and the term t is typically $owner(r)$. I.e., there we wish to query whether according to the owner of a given recourse r, a certain agent has access to that resource.

More concretely, we will be interested in the *data complexity* of this task, i.e., all complexity results will be for a fixed T and φ, and thus are measured in terms of the size of the domain of I_o.

After the publication of the conference version of this paper, Ambrossio and Cramer [2019] already defined a query-driven decision procedure for dAEL that tackles exactly Task 7.1. Their decision procedure is designed in such a way that it allows one to determine access rights while avoiding redundant information flow between principals in order to enhance security and reduce privacy concerns. Their decision procedure is query-driven in the following sense: A query in the form of a dAEL formula φ is posed to a principal A. A determines whether her theory contains enough information to verify φ. It can happen that A cannot verify φ just on the basis of her theory, but can determine that if a certain other principal supports a certain formula, her theory implies the query. For example, A's theory may contain the formula $K_B p \Rightarrow \varphi$. In this case, A can forward a remote sub-query to B concerning the status of p in B's theory. If B verifies the sub-query p and informs A about this, A can complete her verification of the original query φ.

In this generation of subqueries, loops may occur. For this reason, the decision procedure includes a loop detection mechanism. When a loop is detected, the query causing the loop (by being identical to a query that is an ancestor of it in the call graph) is labelled either with $\mathbf{f}$ or $\mathbf{u}$, depending on whether the loop is over a negation or not. The details are described in [Ambrossio and Cramer, 2019].

Keeping the distributed theory $\mathcal{T}$ fixed and varying the size of the domain, this decision procedure for dAEL and its restriction to dAEL have a worst case runtime that is exponential in the size of the domain. From a practical perspective, this is not an encouraging result. In the following theorem, we show that this complexity is not a coincidence.

Theorem 7.2. *Task 7.1 is NP-hard and co-NP-hard.*

Proof. It is well known that the graph coloring problem (the problem of determining whether a given graph is colorable by a given set of colors) is NP-complete. In fact, this is one of Karp's original 21 NP-complete problems [Karp, 1972]. We reduce both this problem and its negation to the Task 7.1.

Consider a set of agents $\mathcal{A} = \{a, b\}$, the vocabulary Σ_o with two constants a, b and predicates $Node/1$, $Color/1$, $Edge/2$. Also consider the vocabulary Σ_s consisting of predicate symbols $Coloring/2$ (with the informal interpretation that $Coloring(n, c)$ holds if node n is colored with color c) and $p/0$. Furthermore, as before, let Σ denote $\Sigma_o \cup \Sigma_s$. Let φ denote the first-order Σ-formula

$$\begin{aligned}&(\forall n : Node(n) \Rightarrow \exists c : Color(c) \land Coloring(n, c)) \land\\ &(\forall n1, n2 : Edge(n_1, n_2) \Rightarrow \neg\exists c : Coloring(n1, c) \land Coloring(n2, c)).\end{aligned}$$

It can be seen that φ holds in an interpretation I if and only if $Coloring^I$ is a coloring of $Edge^I$ with the colors $Color^I$. Let T_a be the empty theory and

$$T_b = \{p \Leftrightarrow K_a\neg\varphi\}.$$

Now consider $T = (T_a, T_b)$. For any graph (V, E) and set of colors C, let $I_{(V,E),C}$ denote the Σ_o interpretation with domain $\{a, b\} \cup V$ interpreting a as a, b as b and $Edge$ as E, $Node$ as V, and $Color$ as C.

Now, we claim that, given a Σ_o-interpretation I_o

- $K_a\neg\varphi$ holds in the well-founded model of T under $I_{(V,E),C}$ if and only if (V, E) is not colorable with C

- For any graph (V, E) and set of colors C, $K_b p$ holds in the well-founded model of T under $I_{(V,E),C}$ if and only if (V, E) admits no coloring with C, and
- $K_b \neg p$ holds in the well-founded model of T under $I_{(V,E),C}$ if and only if (V, E) admits a coloring with C.

Let $\mathcal{Q}$ denote the well-founded model of T. First of all, we note that $\mathcal{Q}_a = \bot$, i.e. $\mathcal{Q}_a$ is the set of all Σ-interpretations with domain $\{a, b\} \cup V$ that coincide with $I_{(V,E),C}$ on Σ_o. This is easy to see, since a has no knowledge whatsoever. Hence, the only way it can know $\neg\varphi$ is if (V, E) admits no coloring with C. From this fact, the first claim easily follows. Now, the other two claims follow from the first, since in b's theory, p is defined to be $K_a \neg\varphi$. □

It can be seen from the previous proof that the result also holds for the Kripke-Kleene model.

Given this complexity result, one might wonder how a logic like dAEL can be useful in practice, for instance for applications like access control. To this end, we develop a fragment of our logic, for which Task 7.1 can be solved in polynomial time. This restriction of dAEL has to be chosen in a careful way in order to find a good balance between expressivity and efficiency. In this subsection, we present one reasonable choice, called dAEL$_R$, from the rich space of potential restrictions of dAEL, and show that for this fragment, the task we are interested in, has polynomial data complexity.

In dAEL$_R$, the theories of the various agents may not contain arbitrary formulas, but only formulas that follow a certain syntax akin to that of rules in a logic program. We define these *rule formulas* as follows:

Definition 7.3. A *modal literal* is a dAEL formula of the form $K_A l$ or $\neg K_A l$ where A is an agent and l is a literal (an atom or its negation).

Definition 7.4. A *modal complex* is a dAEL formula in which every atom is a subformula of a modal literal.

Note that for a model complex φ, the interpretation $\varphi^{\mathcal{B},I,a}$ does not depend on I, so we may also write $\varphi^{\mathcal{B},a}$ for it.

Definition 7.5. A *rule formula* is a dAEL formula of the form $\forall \overline{x} : (\varphi \Rightarrow l)$, where φ is a modal complex and l is a literal

While dAEL$_R$ is significantly more restrictive than $\mathcal{L}_d$, we believe that for most access control applications it is sufficient. First of all, note that the access control application of dAEL discussed in Section 6 lies fully within dAEL$_R$. As a further example, it seems in principle to be possible to use dAEL$_R$ to handle the detailed case study that Garg [2009] presented in order to illustrate the applicability of his access control logic BL.[3]

Restricting our attention to dAEL$_R$, the task we are interested in has polynomial data complexity:

Theorem 7.6. *If each formula in T is a rule formula, then Task 7.1 can be performed in polynomial time.*

Proof. The clue to this proof is that in the special case where T only consists of rule formulas, only a small subset of all the possible world structures (and belief pairs) is relevant, in the sense that all others need not be considered in order to compute the well-founded model of T. Evaluating φ in this model is then also efficient.

To this end, we call a distributed possible world structure $\mathcal{Q}$ *literal-determined* if for each agent A, there exists a set of ground literals L_A such that $\mathcal{Q}_A$ is the least informative possible world structure in which all literals L_A are known. Stated differently

$$\mathcal{Q}_A = \{I \mid I \models l \text{ for all } l \in L_A\}.$$

We call a distributed belief pair *literal-determined* if both its possible world structures are.

Claim: *If T consists only of rule formulas, then for each distributed belief pair $\mathcal{B}$, it holds that $\mathcal{D}^*_{\mathcal{T}}(\mathcal{B})$ is literal-determined.* To see that this claim indeed holds, note that given a distributed belief pair $\mathcal{B}$ and an agent A, it holds that

$$\begin{aligned}
&\mathcal{D}^c_{\mathcal{T}}(\mathcal{B})_A = \{I \mid (\mathcal{T}_A)^{\mathcal{B},I} \neq \mathbf{f}\}\\
&= \{I \mid (\forall \overline{x} : (\varphi \Rightarrow l))^{\mathcal{B},I} \neq \mathbf{f} \text{ for all rule formulas } \forall \overline{x} : (\varphi \Rightarrow l) \text{ in } \mathcal{T}_A\}\\
&= \{I \mid I \models l[\overline{x} : \overline{d}] \text{ for all rule formulas } \forall \overline{x} : (\varphi \Rightarrow l) \text{ and all instantiations } \overline{x} : \overline{d} \text{ for which } \varphi^{\mathcal{B},[\overline{x}:\overline{d}]} \geq_t \mathbf{u}\}
\end{aligned}$$

[3] The BL theory that formalizes the access control policy of this case study consists of formulas of the form A *claims* $(\varphi_1 \wedge \cdots \wedge \varphi_n \wedge B_1$ *says* $\psi_1 \wedge \cdots \wedge B_m$ *says* $\psi_m \Rightarrow \chi)$. Given the intuitionistic nature of BL discussed in Section 8.2 below, a BL theory consisting of such formulas leads to the same access control decisions as the dAEL$_R$ distributed theory $\mathcal{T}$ such that for each formula of the form just mentioned, there is a corresponding rule formula $K_A\varphi_1 \wedge \cdots \wedge K_A\varphi_n \wedge K_{B_1}\psi_1 \wedge \cdots \wedge K_{B_m}\psi_m \Rightarrow \chi$ in $\mathcal{T}_A$.

I.e., that this set is literal-determined. An analogous argument yields that also $\mathcal{D}^l_{\mathcal{T}}(\mathcal{B})_A$ is literal-determined.

Now, from this claim, it follows that a well-founded induction exists that only uses literal-determined distributed belief pairs (indeed, an example of such induction is the maximal one). Each increasing (in precision) chain of distributed belief pairs has only polynomial length (there are only polynomally many ground literals). Furthermore, for each literal-determined distributed belief pair $\mathcal{B}$, $\mathcal{D}^*_{\mathcal{T}}(\mathcal{B})_A$ can be computed in polynomial time (this is easy to see by the equation in the proof of our claim: indeed, it suffices to evaluate the modal complexes φ occurring in all rule formulas for all instantiations $\overline{d}$ of the $\overline{x}$). From this, it follows that a well-founded induction can be constructed in polynomial time, and hence, the well-founded model of such a theory can be computed in polynomial time. Finally, in order to execute Task 7.1, we need to evaluate a single query in the well-founded model, also that consists of simply evaluating a single formula and hence, Task 7.1 is indeed polynomial. □

8 Related Work

In this section, we discuss two kinds of related work: In Subsection 8.1, we present other multi-agent extensions of autoepistemic logic that have been proposed in the literature, and compare them to dAEL. In Subsection 8.2, we discuss approaches in access control logic related to ours.

8.1 Other multi-agent extensions of AEL

Several extensions of autoepistemic logic, and other non-monotonic reasoning formalisms to the multi-agent case have been made [Morgenstern, 1990; Belle and Lakemeyer, 2015; Toyama *et al.*, 2002; Permpoontanalarp and Jiang, 1995]. Each of them starts from a particular dialect of the non-monotonic logic and generalizes it to multiple agents. Morgenstern [1990] made an extension to Moore's AEL [Moore, 1985a] and studied a centralized theory containing statements about the knowledge of different agents. She does not consider distributed theories and does not assume introspection. Belle and Lakemeyer [2015] also studied multi-agent theories in the same setting but added *only knowing* and *common knowledge* constructs. Toyama *et al.* [2002] developed a distributed variant on autoepistemic logic that also assumes introspection. Compared to our logic, it is quite limited in the sense that it is propositional and only introduces one of the many semantics we discussed, namely the supported model semantics (which corresponds to Moore's original expansions); as such, it also easily encountered the kind of problems with groundedness and cyclic support the original AEL suffered from [Halpern and Moses, 1985; Konolige, 1988; Bogaerts, 2015]. Permpoontanalarp and Jiang [1995] studied a number of logics and developed a proof theory that extends the logic of Morgenstern. Their main motivation is that the logic of Morgenstern has some undesirable properties if reduced to the single agent case, where it differs from AEL. Our logic on the other hand, when instantiated with only one agent, exactly coincides with AEL. Vlaeminck *et al.* [2012] defined two extensions to AEL with multiple agents, namely ordered epistemic logic (OEL) and distributed ordered epistemic logic (dOEL). Both of these logics require a partial order on the agents, where agents can only refer to knowledge of agents strictly lower in the order. If we add this restriction to our logic, we get exactly dOEL, i.e., dOEL is the fragment of dAEL for which there exists a stratification on the agents such that agents only refer to knowledge of "lower" agents. For such theories, all AFT semantics coincide and are equal to the semantics of dOEL as defined by Vlaeminck *et al.* [2012]. The logic OEL is close to dOEL, with the difference being that in OEL an agents knows everything any agent lower in the order knows. This behavior can be simulated in dAEL by adding the axiom scheme $K_A\phi \Rightarrow \phi$ to the theory of each agent greater than A in the order. In the context of an application to access control, the restriction of dOEL and OEL that a a global stratification on the agents is required is undesirable for a truly distributed system.

Our most important contribution with respect to other approaches that define multi-agent extensions of AEL is that we present a uniform, fundamental principle to lift various of those dialects to the multi-agent case using AFT. In this paper, we already lift 5 dialects, and it easily extends to more semantics. We can use the same approach to lift the family of *ultimate* semantics [Denecker *et al.*, 2000], *(partial) grounded fixpoint semantics* [Bogaerts *et al.*, 2015a,b], *well-founded set semantics* [Bogaerts *et al.*, 2016], *conflict-freeness*, *M-stable semantics* and *L-stable semantics* [Strass, 2013] from AEL to dAEL. This approach not only allows us to lift many semantics, it also provides a uniform principle for *comparing* various semantics and hence it brings *order* in the zoo of semantics for multi-agent AEL.

8.2 Related approaches in access control logic

Most access control logics proposed in the literature have been defined in a proof-theoretical way, i.e., by specifying which axioms and inference rules they satisfy. This contrasts with our approach of defining dAEL model-theoretically rather than proof-theoretically. Our main motivation for defining dAEL model-theoretically is that model-theoretic definitions are more basic: from a model-theoretic definition, a notion of entailment, and hence a proof-theoretic characterization can be derived, but not the other way around. We have already motivated the application of autoepistemic logic to access control in section 2.2, and the use of the well-founded semantics in section 6.

Garg and Abadí [Garg and Abadi, 2008] and Genovese [Genovese, 2012] have defined Kripke semantics for many of the access control logics discussed in the literature. However, these semantics are not meant to specify the meaning of the $says$-modality, but to be a tool for defining decision procedures for those access control logics. This contrasts with our approach of studying the meaning of the $says$-modality by showing that its intended use in access control justifies an application of the semantic principles of autoepistemic logic.

Hirsch and Clarkson [Hirsch and Clarkson, 2013] have defined a so-called *belief semantics* as well as a standard Kripke semantics for their access control logic FOCAL, arguing that the belief semantics corresponds better than the Kripke semantics to how principals reason in real-world systems. However FOCAL does not support mutual positive or negative introspection between principals, making it difficult to naturally model both delegation and denial.

We are not aware of any other $says$-based access control logic that allows to model the non-monotonic behavior of denials as straightforwardly as dAEL by allowing to derive formulas of the form $\neg k\ says\ \phi$ and supporting mutual negative introspection between principals. However, most state-of-the-art access control logics allow for mutual positive introspection between principals. For example BL, an access control logic with support for system state and explicit time, supports mutual positive introspection [Garg, 2009; Garg and Pfenning, 2012].

The only approach to access control based on a non-monotonic logical formalism that we are aware of is the unifying access control meta-model proposed by Barker [Barker, 2009]. This proposed meta-model is based on a rule language interpreted using Clark's completion, a non-monotonic logic programming semantics. Unlike dAEL, Barker's meta-model is not designed for distributed access control. If it is extended to support distributed access control policies and used in a straightforward way to implement our example, its behavior would correspond to the behavior that dAEL would have with the supported model semantics, which we have shown in section 6 to give undesirable results.

Barker and Genovese [2011] describe *Secommunity*, a framework for distributed access control based on Barker's access control meta-model. Secommunity is implemented in the Answer Set Programming system DLV, which works with the stable semantics. Thus in this distributed framework based on the Barker's meta-model, Clark's completion semantics has been replaced by stable semantics. But as described in section 6, the stable semantics also gives undesirable results when applied to a standard access control problem. To the best of our knowledge, Barker's meta-model has never been used under the well-founded semantics. We expect that, given the strong correspondences between logic programming and autoepistemic logic, induced by AFT, there will be a close relation between such a usage of Barker's model and our logic dAEL. Researching this is a topic for future work.

Classical vs. intuitionistic logic Many state-of-the-art access control logics are based on intuitionistic rather than classical logic. Garg [2009] justifies the use of intuitionistic logic in access control on the basis of the security principle that when access is granted to a principal k, it should be known where k's authority comes from. Autoepistemic logic, on the other hand, is based on classical logic. However, in Section 6 we argued that under the well-founded semantics, with its constructive semantics, this security principle is still satisfied.

Another justification for the use of intuitionistic logic has been put forward by Abadi [2008], who gives an overview over the design-space of access control logics, discussing advantages and disadvantages of certain axioms. One axiom discussed by Abadi is the Unit axiom $\phi \Rightarrow k\ says\ \phi$. Note that Unit implies mutual full introspection (i.e. mutual positive and mutual negative introspection) between principals, but is strictly stronger than mutual full introspection. Abadi showed that in classical logic, Unit implies Escalation, i.e. the property that $k\ says\ \phi$ implies that either ϕ or $k\ says\ \bot$. Abadi [2008] argued that Escalation embodies a rather degenerate interpretation of the $says$-modality, because it means that if a statement supported by a principal is actually false, the principal can be considered to support all statements.

Since dAEL builds on top of classical logic, we need to discuss the status of Unit and Escalation in dAEL. In dAEL, there is no objective interpretation of objective formulas, i.e., formulas without the $says$-modality. All we have is an introspective agent's interpretations of formulas, from which we can derive an objective interpretation of formulas of the form $k\ says\ \psi$, or, in our notation $K_k\psi$. Hence, when ϕ is an objective formula, neither Unit nor Escalation can be evaluated objectively in AEL. What we can do instead is to ask the following two questions:

1. Do Unit and Escalation hold for a formula ϕ of the form $K_k\psi$ or $\neg K_k\psi$? I.e., are the following formulas tautologies for all agents j and k and every formula ψ?

$$\begin{aligned} &K_k\psi \Rightarrow K_jK_k\psi && \text{(unit)}\\ &\neg K_k\psi \Rightarrow K_j\neg K_k\psi && \text{(unit)}\\ &K_jK_k\psi \Rightarrow (K_k\psi \vee K_j\bot) && \text{(escalation)}\\ &K_j\neg K_k\psi \Rightarrow (\neg K_k\psi \vee K_j\bot) && \text{(escalation)} \end{aligned}$$

2. Do Unit and Escalation hold within the belief of an agent? I.e., are the following formulas tautologies?

$$\begin{aligned} &K_k(\psi \Rightarrow K_j\psi) && \text{(unit)}\\ &K_k(K_j\psi \Rightarrow (\psi \vee K_k\bot)) && \text{(escalation)} \end{aligned}$$

The answer to question 1 is "yes". Indeed, Unit means that the agents in question have introspection in each other's knowledge. Escalation also holds, but simply due to the fact that $K_jK_k\psi$ and $K_k\psi$ are equivalent in our logic, as can easily be seen from the truth evaluation. In this case, escalation boils down to stating that it is impossible for an agent j to consistently make the claim that another agent k said something k did not actually say.

The answer to question 2 is "no": Principal A can know p, but not know that principal B knows p (unit). Also, A can know that B claims some property holds (say, that B has access to a recourse) and in the meanwhile A can claim that B does not have access to this recourse without thereby implying that B's claims are inconsistent.

9 Conclusion and Future Work

Motivated by an application in access control, we have extended AEL to a distributed setting, resulting in a logic called distributed autoepistemic logic (dAEL). dAEL allows for a set of agents to each have their own theory in which they refer to each others knowledge. For this, the knowledge operator K of AEL is replaced by an indexed operator K_A, where A refers to an agent. We have defined the semantics of this logic building on approximation fixpoint theory (AFT), a lattice-theoretic framework that captures the semantics of many non-monotonic logics. Using AFT has many practical advantages: first of all, it allows for a uniform lifting of many different semantics. Secondly, it ensures that all fundamental principles underlying these semantics remain preserved. And third, in doing so, we immediately obtain access to a wide variety of theoretic results. For instance, properties such as the fact that the well-founded model approximates all stable models was obtained *by definition*, since the corresponding result holds in the algebraic setting. Similarly, we can (but did not do so) apply algebraic stratification results [Vennekens *et al.*, 2006; Bogaerts *et al.*, 2016], predicate introduction results [Vennekens *et al.*, 2007b], or modularity results [Truszczyński, 2006] without effort. Also future progress in AFT will be directly applicable.

We have illustrated how dAEL can be applied to access control and have argued that one semantics is particularly suitable for modelling access control policies, namely the well-founded semantics. The non-monotonic behaviour of dAEL allowed us to model denial and revocation of access in dAEL, something that previous access control logics could not achieve due to their monotonicity. We have thus built a bridge between non-monotonic logic and access control. One of the tasks left for future research is to study whether this bridge may lead to further fruitful interaction between these fields additionally to the one already considered in this paper.

We have studied the complexity of reasoning with the well-founded semantics of dAEL and came to the unsettling conclusion that complexity of the considered task is quite high (both NP and coNP hard), thus making it unpractical for use in an access control setting. To overcome this limitation, we defined a fragment of our logic, which we called dAEL$_R$, in which reasoning becomes polynomial and that suffices to model the kind of application that motivated the paper in the first place.

Acknowledgements

This research was supported by project GOA 13/010 of the Research Fund KU Leuven and projects G.0489.10, G.0357.12, and G.0922.13 of the Research Foundation - Flanders and more specifically was part of the FNR-FWO project *Specification logics and Inference tools for verification and Enforcement of Policies*. Part of this work was also performed while Bart Bogaerts was a postdoctoral fellow of the Research Foundation – Flanders (FWO).

References

Martín Abadi. Logic in Access Control. In *Proceedings of the Eighteenth Annual IEEE Symposium on Logic in Computer Science*, pages 228–233, 2003.

Martín Abadi. Variations in Access Control Logic. In *9th International Conference on Deontic Logic in Computer Science*, pages 96–109, 2008.

Diego Agustín Ambrossio and Marcos Cramer. A Query-Driven Decision Procedure for Distributed Autoepistemic Logic with Inductive Definitions. *arXiv e-prints*, page arXiv:1910.04010, Oct 2019.

Andrew W Appel and Edward W Felten. Proof-carrying authentication. In *Proceedings of the 6th ACM Conference on Computer and Communications Security*, pages 52–62. ACM, 1999.

Steve Barker and Valerio Genovese. Secommunity: A Framework for Distributed Access Control. In *LPNMR*, pages 297–303, 2011.

Steve Barker, Guido Boella, Dov M. Gabbay, and Valerio Genovese. Reasoning about delegation and revocation schemes in answer set programming. *J. Log. Comput.*, 24(1):89–116, 2014.

Steve Barker. The next 700 access control models or a unifying meta-model? In *Proceedings of the 14th ACM symposium on Access control models and technologies*, SACMAT '09, pages 187–196. ACM, 2009.

Vaishak Belle and Gerhard Lakemeyer. Only knowing meets common knowledge. In Yang and Wooldridge [Yang and Wooldridge, 2015], pages 2755–2761.

Bart Bogaerts, Joost Vennekens, and Marc Denecker. Grounded fixpoints and their applications in knowledge representation. *Artif. Intell.*, 224:51–71, 2015.

Bart Bogaerts, Joost Vennekens, and Marc Denecker. Partial grounded fixpoints. In Yang and Wooldridge [Yang and Wooldridge, 2015], pages 2784–2790.

Bart Bogaerts, Joost Vennekens, and Marc Denecker. On well-founded set-inductions and locally monotone operators. *ACM Trans. Comput. Logic*, 17(4):27:1–27:32, September 2016.

Bart Bogaerts. *Groundedness in logics with a fixpoint semantics*. PhD thesis, Department of Computer Science, KU Leuven, June 2015. Denecker, Marc (supervisor), Vennekens, Joost and Van den Bussche, Jan (cosupervisors).

Ajay Chander, Drew Dean, and John C. Mitchell. Reconstructing trust management. *Journal of Computer Security*, 2004.

Marcos Cramer and Giovanni Casini. Postulates for Revocation Schemes. In Matteo Maffei and Mark Ryan, editors, *Principles of Security and Trust*, pages 232–252, Berlin, Heidelberg, 2017. Springer Berlin Heidelberg.

Marcos Cramer, Diego Agustin Ambrossio, and Pieter Van Hertum. A Logic of Trust for Reasoning about Delegation and Revocation. In *Proceedings of the 20th ACM Symposium on Access Control Models and Technologies* , pages 173–184. ACM, 2015.

Marc Denecker and Joost Vennekens. Well-founded semantics and the algebraic theory of non-monotone inductive definitions. In Chitta Baral, Gerhard Brewka, and John S. Schlipf, editors, *LPNMR*, volume 4483 of *Lecture Notes in Computer Science*, pages 84–96. Springer, 2007.

Marc Denecker, Victor Marek, and Mirosław Truszczyński. Fixpoint 3-valued semantics for autoepistemic logic. In Jack Mostow and Chuck Rich, editors, *AAAI'98*, pages 840–845, Madison, Wisconsin, July 26-30 1998. MIT Press.

Marc Denecker, Victor Marek, and Mirosław Truszczyński. Approximations, stable operators, well-founded fixpoints and applications in nonmonotonic reasoning. In Jack Minker, editor, *Logic-Based Artificial Intelligence*, volume 597 of *The Springer International Series in Engineering and Computer Science*, pages 127–144. Springer US, 2000.

Marc Denecker, Victor Marek, and Mirosław Truszczyński. Uniform semantic treatment of default and autoepistemic logics. *Artif. Intell.*, 143(1):79–122, 2003.

Marc Denecker, Victor Marek, and Mirosław Truszczyński. Ultimate approximation and its application in nonmonotonic knowledge representation systems. *Information and Computation*, 192(1):84–121, July 2004.

Marc Denecker, Victor Marek, and Mirosław Truszczyński. Reiter's default logic is a logic of autoepistemic reasoning and a good one, too. In Gerd Brewka, Victor Marek, and Mirosław Truszczyński, editors, *Nonmonotonic Reasoning – Essays Celebrating Its 30th Anniversary*, pages 111–144. College Publications, 2011.

Marc Denecker. The well-founded semantics is the principle of inductive definition. In Jürgen Dix, Luis Fariñas del Cerro, and Ulrich Furbach, editors, *JELIA*, volume 1489 of *LNCS*, pages 1–16. Springer, 1998.

Deepak Garg and Martín Abadi. A Modal Deconstruction of Access Control Logics. In Roberto Amadio, editor, *Foundations of Software Science and Computational Structures: 11th International Conference, FOSSACS 2008, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2008, Budapest, Hungary, March 29 - April 6, 2008. Proceedings*, pages 216–230, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

Deepak Garg and Frank Pfenning. Stateful Authorization Logic – Proof Theory and a Case Study. *Journal of Computer Security*, 20(4):353–391, 2012.

Deepak Garg. *Proof Theory for Authorization Logic and Its Application to a Practical File System*. PhD thesis, 2009.

Valerio Genovese. *Modalities in Access Control: Logics, Proof-theory and Application*. PhD thesis, 2012.

Yuri Gurevich and Itay Neeman. DKAL: Distributed-knowledge authorization language. In *Proceedings of the 2008 21st IEEE Computer Security Foundations Symposium*, pages 149–162. IEEE, 2008.

Åsa Hagström, Sushil Jajodia, Francesco Parisi-Presicce, and Duminda Wijesekera. Revocations – A Classification. In *Proceedings of the 14th IEEE Workshop on Computer Security Foundations*, pages 44–58. IEEE, 2001.

Joseph Y. Halpern and Yoram Moses. Towards a theory of knowledge and ignorance: Preliminary report. In Krzysztof R. Apt, editor, *Logics and Models of Concurrent Systems*, volume 13 of *NATO ASI Series*, pages 459–476. Springer Berlin Heidelberg, 1985.

Andrew K. Hirsch and Michael R. Clarkson. Belief Semantics of Authorization Logic. *CoRR*, abs/1302.2123, 2013.

G. E. Hughes and M. J. Cresswell. *A New Introduction To Modal Logic*. Routledge, 1996.

R. Karp. Reducibility among combinatorial problems. In R. Miller and J. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

Stephen Cole Kleene. On notation for ordinal numbers. *The Journal of Symbolic Logic*, 3(4):150–155, 1938.

Kurt Konolige. On the relation between default and autoepistemic logic. *Artif. Intell.*, 35(3):343–382, 1988.

Hector J. Levesque. All I know: A study in autoepistemic logic. *Artif. Intell.*, 42(2-3):263–309, 1990.

C.I. Lewis and C.H. Langford. *Symbolic logic*. Century philosophy series. The Century co., 1932.

Ninghui Li, Benjamin N. Grosof, and Joan Feigenbaum. Delegation Logic: A Logic-based Approach to Distributed Authorization. *ACM Transaction on Information and System Security*, 2003.

Robert C. Moore. A Formal Theory of Knowledge and Action. In J. R. Hobbs and R. C. Moore, editors, *Formal Theories of the Commonsense World*, pages 319–358. Springer-Verlag, 1985.

Robert C. Moore. Semantical considerations on nonmonotonic logic. *Artif. Intell.*, 25(1):75–94, 1985.

Leora Morgenstern. A formal theory of multiple agent nonmonotonic reasoning. In T. Dieterich and W. Swartout, editors, *Proceedings of the 8th National Conference on Artificial Intelligence. Boston, Massachusetts, July 29 - August 3, 1990, 2 Volumes.*, pages 538–544. AAAI/MIT Press, 1990.

Ilkka Niemelä. Constructive tightly grounded autoepistemic reasoning. In John Mylopoulos and Raymond Reiter, editors, *Proceedings of the 12th International Joint Conference on Artificial Intelligence. Sydney, Australia, August 24-30, 1991*, pages 399–405. Morgan Kaufmann, 1991.

Yongyuth Permpoontanalarp and John Yuejun Jiang. On multi-agent autoepistemic reasoning. In *WOCFAI*, pages 307–318, 1995.

Hannes Strass. Approximating operators and semantics for abstract dialectical frameworks. *Artif. Intell.*, 205:39–70, 2013.

Roberto Tamassia, Danfeng Yao, and William H. Winsborough. Role-Based Cascaded Delegation. In *Proceedings of the 9th ACM symposium on Access control models and technologies* , 2004.

Katsuhiko Toyama, Takahiro Kojima, and Yasuyoshi Inagaki. Translating multi-agent autoepistemic logic into logic program. In Jürgen Dix, João Alexandre Leite, and Ken Satoh, editors, *Computational Logic in Multi-Agent Systems: 3rd International Workshop, CLIMA'02, Copenhagen, Denmark, August 1, 2002, Pre-Proceedings*, volume 93 of *Datalogiske Skrifter*, pages 49–62. Roskilde University, 2002.

Mirosław Truszczyński. Strong and uniform equivalence of nonmonotonic theories - an algebraic approach. *Ann. Math. Artif. Intell.*, 48(3-4):245–265, 2006.

Pieter Van Hertum, Marcos Cramer, Bart Bogaerts, and Marc Denecker. Distributed autoepistemic logic and its application to access control. In Subbarao Kambhampati, editor, *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 1286–1292. IJCAI/AAAI Press, 2016.

Joost Vennekens, David Gilis, and Marc Denecker. Splitting an operator: Algebraic modularity results for logics with fixpoint semantics. *ACM Trans. Comput. Log.*, 7(4):765–797, 2006.

Joost Vennekens, David Gilis, and Marc Denecker. Erratum to splitting an operator: Algebraic modularity results for logics with fixpoint semantics (vol 7, pg 765, 2006), January 2007.

Joost Vennekens, Maarten Mariën, Johan Wittocx, and Marc Denecker. Predicate introduction for logics with a fixpoint semantics. Parts I and II. *Fundamenta Informaticae*, 79(1-2):187–227, 2007.

Hanne Vlaeminck, Joost Vennekens, Maurice Bruynooghe, and Marc Denecker. Ordered Epistemic Logic: Semantics, complexity and applications. In Gerhard Brewka, Thomas Eiter, and Sheila A. McIlraith, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Knowledge Representation and Reasoning, Rome, 10-14 July 2012*, pages 369–379. AAAI Press, 2012.

Qiang Yang and Michael Wooldridge, editors. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*. AAAI Press, 2015.

Danfeng Yao and Roberto Tamassia. Compact and Anonymous Role-Based Authorization Chain. *ACM Transactions on Information and System Security*, 2009.

Longhua Zhang, Gail-Joon Ahn, and Bei-Tseng Chu. A rule-based framework for role-based delegation and revocation. *ACM Transactions on Information and System Security*, 2003.

A Proofs of Theorems 5.12, 5.16 and 5.17

In order to prove Theorems 5.12, 5.16 and 5.17, we first need to define some additional notions and prove some lemmas.

Lemma A.1. *If $\mathcal{T}$ is permaconsistent, then $D^*_{\mathcal{T}}(\bot, \top)$ is universally consistent.*

Proof. If $\mathcal{T}$ is permaconsistent, then for each agent $A \in \mathcal{A}$ and each theory T' that can be constructed from $\mathcal{T}_A$ by replacing all non-nested occurrences of modal literals by $\mathbf{t}$ or $\mathbf{f}$ is consistent. Now, for each agent A, let $\mathcal{T}'_A$ the theory constructed from $\mathcal{T}_A$ by replacing all non-nested occurrences of modal literals by $\mathbf{t}$ if they occur in a negative context (under an odd number of negations) and by $\mathbf{f}$ otherwise. This theory is clearly stronger than $\mathcal{T}_A$. Since $\mathcal{T}$ is permaconsistent, $\mathcal{T}'_A$ is satisfiable, let I_A be a model of $\mathcal{T}'_A$. In this case, it holds that $\mathcal{T}_A^{(\bot,\top),I_A} = \mathbf{t}$ (since $\mathcal{T}_A$ is weaker than $\mathcal{T}'_A$).

From this, we find that for each agent A, $\{I \mid \mathcal{T}_A^{(\bot,\top),I}\}$ is non-empty and thus that $D^*(\bot, \top)$ is indeed universally consistent. □

Proof of Theorem 5.17. Suppose $\mathcal{T}$ is permaconsistent. By Lemma A.1, $D^*_{\mathcal{T}}(\bot, \top)$ is universally consistent. It follows directly from the definitions in AFT that each model of $\mathcal{T}$ (under any of the semantics), is more precise than $D^*_{\mathcal{T}}(\bot, \top)$. Furthermore, if $\mathcal{B}' \geq_p \mathcal{B}$ and $\mathcal{B}$ is universally consistent, then so is $\mathcal{B}'$. □

Definition A.2. Given a Σ'-structure J and a Σ'-term t, we write J_t for the Σ-structure defined by $s^{J_t}(d_1, \ldots, d_n) := s^J(d_1, \ldots, d_n, t^J)$ for every $s \in \Sigma$.

The following lemma states that for an indexed family of structures, the mapping does not discard any information, i.e., after applying the mapping, we can recover each agent's structure.

Lemma A.3. *Let $(I_A)_{A \in \mathcal{A}}$ be an indexed family of Σ-structures, and let $J = \tau_{\mathit{structure}}((I_A)_{A\in\mathcal{A}})$. Then $J_A = I_A$.*

Proof. Let $s \in \Sigma$. Then $s^{J_A}(d_1, \ldots, d_n) = s^J(d_1, \ldots, d_n, A) = s^{I_A}(d_1, \ldots, d_n)$. □

The following lemma generalizes Lemma A.3 to DPWSs.

Lemma A.4. *Let $\mathcal{Q}$ be a universally consistent DPWS, and let $A \in \mathcal{A}$. Then*

$$\{J_A \mid J \in \tau_{\mathit{pws}}(\mathcal{Q})\} = \mathcal{Q}_A$$

for each $A \in \mathcal{A}$.

Proof. We prove the equality by proving the subset relation in both directions.

Let $J \in \tau_{\mathit{pws}}(\mathcal{Q})$. Then there is an indexed family $(I_{A'})_{A'\in\mathcal{A}}$ s.t. $I_{A'} \in \mathcal{Q}_{A'}$ for each $A' \in \mathcal{A}$ and $J = \tau_{\mathit{structure}}((I_{A'})_{A'\in\mathcal{A}})$. Then by Lemma A.3, $J_A = I_A \in \mathcal{Q}_A$, as required.

To prove the other direction, let $I \in \mathcal{Q}_A$. Since $\mathcal{Q}$ is universally consistent, there is some indexed family $(I_{A'})_{A'\in\mathcal{A}}$ s.t. $I_A = I$ and $I_{A'} \in \mathcal{Q}_{A'}$ for all $A' \in \mathcal{A}$. Define $J := \tau_{\mathit{structure}}((I_{A'})_{A'\in\mathcal{A}})$. Note that $J \in \tau_{\mathit{pws}}(\mathcal{Q})$. Now $J_A = I_A$ by Lemma A.3, so $J_A = I$, as required. □

The following lemma says that the mapping is faithful to the valuations of AEL and dAEL formulas:

Lemma A.5. *For a Σ'-term t, a formula $\phi \in \mathcal{L}_d^{\Sigma}$, a universally consistent distributed belief pair $\mathcal{B}$ and a Σ'-strucutre J,*

$$\tau_{\mathit{formula}}(t, \phi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}),J} = \phi^{\mathcal{B},J_t}.$$

Proof. We prove the lemma by induction over the structure of ϕ.
$\underline{\phi = P(t_1, \ldots, t_n).}$ Then $\tau_{\mathit{formula}}(t, \phi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} = \mathbf{t}$
iff $P({t_1}_t, \ldots, {t_n}_t, t)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} = \mathbf{t}$
iff $({t_1}_t^J, \ldots, {t_n}_t^J, t^J) \in P^J$
iff $(t_1^{J_t}, \ldots, t_n^{J_t}, t^J) \in P^{J_t}$
iff $\phi^{\mathcal{B}, J_t} = \mathbf{t}$.
Analogously, we find $\tau_{\mathit{formula}}(t, \phi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} = \mathbf{f}$ iff $\phi^{\mathcal{B}, J_t} = \mathbf{f}$.
$\underline{\phi = \neg\psi.}$ Then $\tau_{\mathit{formula}}(t, \phi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} = \neg\tau_{\mathit{formula}}(t, \psi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J}$ and the result follows by the induction hypothesis.
$\underline{\phi = \phi_1 \wedge \phi_2.}$ Similarly.
$\underline{\phi = \forall x : \psi.}$ Similarly.
$\underline{\phi = K_s\psi.}$ Then $\tau_{\mathit{formula}}(t, \phi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} = \mathbf{t}$
iff $\exists x : (x = s_t \wedge \mathrm{Agt}(x) \wedge K\tau_{\mathit{formula}}(x, \psi))^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} = \mathbf{t}$
iff there is a $d \in \mathcal{A}$ with $d = s_t^J$ such that for each $J' \in \tau_{\mathit{pws}}(\mathcal{B}^c)$, $\tau_{\mathit{formula}}(d, \psi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J'} = \mathbf{t}$ (by Definition 3.3)
iff $s^{J_t} \in \mathcal{A}$ and for each $J' \in \tau_{\mathit{pws}}(\mathcal{B}^c)$, $\tau_{\mathit{formula}}(s^{J_t}, \psi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J'} = \mathbf{t}$ (since $s_t^J = s^{J_t}$)
iff $s^{J_t} \in \mathcal{A}$ and for each $J' \in \tau_{\mathit{pws}}(\mathcal{B}^c)$, $\psi^{\mathcal{B}, {J'}_{s^{J_t}}} = \mathbf{t}$ (by the induction hypothesis)
iff $s^{J_t} \in \mathcal{A}$ and for each $I \in \mathcal{B}^c_{s^{J_t}}$, $\psi^{\mathcal{B}, I} = \mathbf{t}$ (since $\mathcal{B}$ is universally consistent, using Lemma A.4)
iff $\phi^{\mathcal{B}, J_t} = \mathbf{t}$ (by Definition 4.10).
Analogously, we find $\tau_{\mathit{formula}}(t, \phi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} = \mathbf{f}$ iff $\phi^{\mathcal{B}, J_t} = \mathbf{f}$.

□

The following lemma states that the dAEL approximator $\mathcal{D}^*_{\mathcal{T}}$ is mapped to the AEL approximator $\mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}$, when restricted to universally consistent distributed belief pairs:

Lemma A.6. *For every distributed theory $\mathcal{T}$ and every universally consistent distributed belief pair $\mathcal{B}$,*

$$\tau_{\mathit{beliefpair}}(\mathcal{D}^*_{\mathcal{T}}(\mathcal{B})) = \mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}(\tau_{\mathit{beliefpair}}(\mathcal{B})).$$

Proof.

$$\begin{aligned}
\tau_{\mathit{beliefpair}}(\mathcal{D}^*_{\mathcal{T}}(\mathcal{B})) &= (\tau_{\mathit{pws}}((\{I \mid \phi^{\mathcal{B}, I} = \mathbf{t} \text{ for each } \phi \in \mathcal{T}_A\})_{A \in \mathcal{A}}), \\
&\qquad \tau_{\mathit{pws}}((\{I \mid \phi^{\mathcal{B}, I} \neq \mathbf{f} \text{ for each } \phi \in \mathcal{T}_A\})_{A \in \mathcal{A}})) \\
&= (\{J \mid \phi^{\mathcal{B}, J_A} = \mathbf{t} \text{ for each } A \in \mathcal{A} \text{ and } \phi \in \mathcal{T}_A\}, \\
&\qquad \{J \mid \phi^{\mathcal{B}, J_A} \neq \mathbf{f} \text{ for each } A \in \mathcal{A} \text{ and } \phi \in \mathcal{T}_A\}) \\
&\qquad \text{(by Definition 5.10 and Lemma A.3)} \\
&= (\{J \mid \tau_{\mathit{formula}}(A, \phi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} = \mathbf{t} \text{ for each } A \in \mathcal{A} \\
&\qquad\qquad \text{and } \phi \in \mathcal{T}_A\}, \\
&\qquad \{J \mid \tau_{\mathit{formula}}(A, \phi)^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} \neq \mathbf{f} \text{ for each } A \in \mathcal{A} \\
&\qquad\qquad \text{and } \phi \in \mathcal{T}_A\}) \text{ (by Lemma A.5)} \\
&= (\{J \mid \varphi^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} = \mathbf{t} \text{ for each } \varphi \in \tau_{\mathit{theory}}(\mathcal{T})\}, \\
&\qquad \{J \mid \varphi^{\tau_{\mathit{beliefpair}}(\mathcal{B}), J} \neq \mathbf{f} \text{ for each } \varphi \in \tau_{\mathit{theory}}(\mathcal{T})\}) \\
&= \mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}(\tau_{\mathit{beliefpair}}(\mathcal{B}))
\end{aligned}$$

□

The mapping maps the dAEL knowledge revision operator $\mathcal{D}_{\mathcal{T}}$ to the corresponding AEL knowledge revision operator $\mathcal{D}_T$:

Lemma A.7. *For every distributed theory $\mathcal{T}$ and every DPWS $\mathcal{Q}$,*

$$\tau_{\mathit{pws}}(\mathcal{D}_{\mathcal{T}}(\mathcal{Q})) = \mathcal{D}_{\tau_{\mathit{theory}}(\mathcal{T})}(\tau_{\mathit{pws}}(\mathcal{Q})).$$

Proof. Follows from Lemma A.6 and the fact that $D^*_T(\mathcal{Q},\mathcal{Q}) = (D_T(\mathcal{Q}), D_T(\mathcal{Q}))$ for each $\mathcal{Q}$. □

The mapping is faithful to the (universal) consistency of (distributed) possible world structures:

Lemma A.8. *A DPWS $\mathcal{Q}$ is universally consistent iff $\tau_{\mathit{pws}}(\mathcal{Q}) \neq \emptyset$.*

Proof. Trivial. □

The following lemma states that the restriction of τ_{pws} to universally consistent DPWSs is injective:

Lemma A.9. *If $\mathcal{Q}$ and $\mathcal{Q}'$ are DPWSs such that $\mathcal{Q}$ is universally consistent and $\tau_{\mathit{pws}}(\mathcal{Q}) = \tau_{\mathit{pws}}(\mathcal{Q}')$, then $\mathcal{Q} = \mathcal{Q}'$.*

Proof. By Lemma A.8, $\mathcal{Q}'$ is universally consistent too. By symmetry, it is enough to show that $\mathcal{Q}_A \subseteq \mathcal{Q}'_A$ for all $A \in \mathcal{A}$.

So let $I_A \in \mathcal{Q}_A$. Given that $\mathcal{Q}$ is universally consistent, we can choose an $I_B \in \mathcal{Q}_B$ for every $B \in \mathcal{A} \setminus \{A\}$. Then $\tau_{\mathit{structure}}((I_A)_{A\in\mathcal{A}}) \in \tau_{\mathit{pws}}(\mathcal{Q}) = \tau_{\mathit{pws}}(\mathcal{Q}')$, so $I_A \in \mathcal{Q}'_A$, as required. □

The following lemma makes an analogous statement of injectivity for $\tau_{\mathit{beliefpair}}$:

Lemma A.10. *If $\mathcal{B}$ and $\mathcal{B}'$ are universal belief pairs such that $\mathcal{B}$ is universally consistent and $\tau_{\mathit{beliefpair}}(\mathcal{B}) = \tau_{\mathit{beliefpair}}(\mathcal{B}')$, then $\mathcal{B} = \mathcal{B}'$.*

Proof. Follows trivially from Lemma A.9. □

The mapping is faithful to the knowledge order:

Lemma A.11. *If $\mathcal{Q} \leq_K \mathcal{Q}'$, then $\tau_{\mathit{pws}}(\mathcal{Q}) \leq_K \tau_{\mathit{pws}}(\mathcal{Q}')$.*

Proof. Let $J \in \tau_{\mathit{pws}}(\mathcal{Q}')$, i.e. for every $A \in \mathcal{A}$, $J_A \in \mathcal{Q}'_A$, i.e. $J_A \in \mathcal{Q}$. So $J \in \tau_{\mathit{pws}}(\mathcal{Q})$. □

The following lemma states that the mapping is faithful to $\leq_K$-least upper bounds and greatest lower bounds:

Lemma A.12. *For a set $\mathcal{S}$ of DPWSs, $\tau_{\mathit{pws}}(\mathrm{lub}_{\leq_K}(\mathcal{S})) = \mathit{lub}_{\leq_K}(\tau_{\mathit{pws}}(\mathcal{S}))$ and $\tau_{\mathit{pws}}(\mathrm{glb}_{\leq_K}(\mathcal{S})) = \mathrm{glb}_{\leq_K}(\tau_{\mathit{pws}}(\mathcal{S}))$.*

Proof. We prove the first equality; the second one can be proven similarly.

First we show that $\tau_{\mathit{pws}}(\mathrm{lub}(\mathcal{S}))$ is an upper bound of $\tau_{\mathit{pws}}(\mathcal{S})$: Let $Q \in \tau_{\mathit{pws}}(\mathcal{S})$. Then there is a DPWS $\mathcal{Q} \in \mathcal{S}$ such that $Q = \tau_{\mathit{pws}}(\mathcal{Q})$. Since $\mathcal{Q} \leq_K \mathrm{lub}\,\mathcal{S}$, Lemma A.11 implies that $Q \leq_K \tau_{\mathit{pws}}(\mathrm{lub}(\mathcal{S}))$.

Now we show that for each upper bound Q' of $\tau_{\mathit{pws}}(\mathcal{S})$, $\tau_{\mathit{pws}}(\mathrm{lub}(\mathcal{S})) \leq_K Q'$: Suppose that for every $Q \in \tau_{\mathit{pws}}(\mathcal{S})$, $Q \leq_K Q'$, i.e. $Q' \subseteq Q$. We need to show that $\tau_{\mathit{pws}}(\mathrm{lub}(\mathcal{S})) \leq_K Q'$, i.e. that $Q' \subseteq \tau_{\mathit{pws}}(\mathrm{lub}(\mathcal{S}))$. So let $J \in Q'$. Let $\mathcal{Q} \in \mathcal{S}$. Then $\tau_{\mathit{pws}}(\mathcal{Q}) \in \tau_{\mathit{pws}}(\mathcal{S})$, so $Q' \subseteq \tau_{\mathit{pws}}(\mathcal{Q})$. Hence $J \in \tau_{\mathit{pws}}(\mathcal{Q})$, i.e. $J\mid_A \in \mathcal{Q}_A$ for each $A \in \mathcal{A}$. Given that $\mathcal{Q}$ was an arbitrary element of $\mathcal{S}$, we have that $J\mid_A \in \bigcap\{Q \mid \text{for some } \mathcal{Q} \in \mathcal{S}, Q = \mathcal{Q}_A\}$. So $J \in \tau_{\mathit{pws}}((\bigcap\{Q \mid \text{for some } \mathcal{Q} \in \mathcal{S}, Q = \mathcal{Q}_A\})_{A\in\mathcal{A}}) = \tau_{\mathit{pws}}(\mathrm{lub}(\mathcal{S}))$, as required. □

The mapping is faithful to $\leq_p$-least upper bounds:

Lemma A.13. *For a set $\mathcal{S}$ of distributed belief pairs, $\tau_{\mathit{beliefpair}}(\mathrm{lub}_{\leq_p}(\mathcal{S})) = \mathrm{lub}_{\leq_p}(\tau_{\mathit{beliefpair}}(\mathcal{S}))$.*

Proof. This follows immediately from Lemma A.12 since

$$(\mathcal{P},\mathcal{S}) \leq_p (\mathcal{P}',\mathcal{S}')$$

if and only if

$$\mathcal{P} \leq_K \mathcal{P}' \text{ and } \mathcal{S} \geq_K \mathcal{S}'.$$

□

The following states that the stable revision of an element of the image of τ_{pws} is itself in the image of τ_{pws}:

Lemma A.14. *For any DPWS $\mathcal{Q}$, there is a DPWS $\mathcal{Q}'$ such that $\tau_{\mathit{pws}}(\mathcal{Q}') = S_{\mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}}(\tau_{\mathit{pws}}(\mathcal{Q}))$.*

Proof. We know that $S_{\mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}}(\tau_{\mathit{pws}}(\mathcal{Q})) = \mathrm{lfp}\, \mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))$. By induction, it is enough to show that for each ordinal number α, there is a DPWS $\mathcal{Q}'$ such that $\tau_{\mathit{pws}}(\mathcal{Q}') = \mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))^\alpha(\bot)$.

For $\alpha = 0$, let $\mathcal{Q}' := (\bot)_{A \in \mathcal{A}}$. Then $\tau_{\mathit{pws}}(\mathcal{Q}') = (\bot) = \mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))^0(\bot)$.

Suppose the result holds for α, i.e. there is a DPWS $\mathcal{Q}'$ such that $\tau_{\mathit{pws}}(\mathcal{Q}') = \mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))^\alpha(\bot)$. By Lemma A.6, $\tau_{\mathit{pws}}(\mathcal{D}^c_{\mathcal{T}}(\mathcal{Q}', \mathcal{Q})) = \mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\tau_{\mathit{pws}}(\mathcal{Q}'), \tau_{\mathit{pws}}(\mathcal{Q})) = \mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))^{\alpha+1}(\bot)$.

Let λ be a limit ordinal such that the result holds for every $\alpha < \lambda$. Define $\mathcal{S} := \{\mathcal{Q}' \mid \tau_{\mathit{pws}}(\mathcal{Q}') = \mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))^\alpha(\bot)$ for some $\alpha < \lambda\}$. By Lemma A.12, $\tau_{\mathit{pws}}(\mathrm{lub}(\mathcal{S})) = \mathrm{lub}(\tau_{\mathit{pws}}(\mathcal{S})) = \mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))^\lambda(\bot)$. □

The following lemma states that the mapping maps the stable dAEL knowledge revision operator $S_{\mathcal{D}^*_{\mathcal{T}}}$ to the stable AEL knowledge revision operator $S_{\mathcal{D}^*_{\mathcal{T}}}$:

Lemma A.15. *For every universally consistent distributed theory $\mathcal{T}$ and every DPWS $\mathcal{Q}$,*

$$S_{\mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}}(\tau_{\mathit{pws}}(\mathcal{Q})) = \tau_{\mathit{pws}}(S_{\mathcal{D}^*_{\mathcal{T}}}(\mathcal{Q})).$$

Proof. Let Q denote $S_{\mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}}(\tau_{\mathit{pws}}(\mathcal{Q}))$.

First suppose $Q = \emptyset$. We need to show that $\tau_{\mathit{pws}}(S_{\mathcal{D}^*_{\mathcal{T}}}(\mathcal{Q})) = \emptyset$, i.e. that $S_{\mathcal{D}^*_{\mathcal{T}}}(\mathcal{Q}) = \mathrm{lfp}(\mathcal{D}^c_{\mathcal{T}}(\cdot, \mathcal{Q}))$ is not universally consistent. For this it is enough to show that every universally consistent DPWS is not a fixpoint of $\mathcal{D}^c_{\mathcal{T}}(\cdot, \mathcal{Q})$. So suppose $\mathcal{Q}'$ is universally consistent. Then $\tau_{\mathit{pws}}(\mathcal{Q}') \neq \emptyset$, i.e. $\tau_{\mathit{pws}}(\mathcal{Q}') <_K Q$. Since Q is the least fixpoint of $\mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))$, $\mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\tau_{\mathit{pws}}(\mathcal{Q}'), \tau_{\mathit{pws}}(\mathcal{Q})) \neq \tau_{\mathit{pws}}(\mathcal{Q}')$. So by Lemma A.6, $\tau_{\mathit{pws}}(\mathcal{D}^c_{\mathcal{T}}(\mathcal{Q}', \mathcal{Q})) \neq \tau_{\mathit{pws}}(\mathcal{Q}')$, i.e. $\mathcal{D}^c_{\mathcal{T}}(\mathcal{Q}', \mathcal{Q}) \neq \mathcal{Q}'$, as required.

Now suppose $Q \neq \emptyset$. By Lemma A.14, there is a DPWS $\mathcal{Q}'$ such that $\tau_{\mathit{pws}}(\mathcal{Q}') = Q$. Note that by Lemma A.8, $\mathcal{Q}$ is universally consistent. $Q = \tau_{\mathit{pws}}(\mathcal{Q}')$ is a fixpoint of $\mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))$, i.e. $\mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\tau_{\mathit{pws}}(\mathcal{Q}'), \tau_{\mathit{pws}}(\mathcal{Q})) \neq \tau_{\mathit{pws}}(\mathcal{Q}')$. By Lemma A.6, $\tau_{\mathit{pws}}(\mathcal{D}^c_{\mathcal{T}}(\mathcal{Q}', \mathcal{Q})) = \tau_{\mathit{pws}}(\mathcal{Q}')$. By Lemma A.9, $\mathcal{D}^c_{\mathcal{T}}(\mathcal{Q}', \mathcal{Q}) = \mathcal{Q}'$, i.e. $\mathcal{Q}'$ is a fixpoint of $\mathcal{D}^c_{\mathcal{T}}(\cdot, \mathcal{Q})$. Let $\mathcal{Q}''$ denote the least fixpoint of $\mathcal{D}^c_{\mathcal{T}}(\cdot, \mathcal{Q})$. Then $\mathcal{Q}'' \leq_K \mathcal{Q}'$, so by Lemma A.11, $\tau_{\mathit{pws}}(\mathcal{Q}'') \leq_K \tau_{\mathit{pws}}(\mathcal{Q}')$. Additionally, $\mathcal{D}^c_{\mathcal{T}}(\mathcal{Q}'', \mathcal{Q}) = \mathcal{Q}''$, so $\tau_{\mathit{pws}}(\mathcal{D}^c_{\mathcal{T}}(\mathcal{Q}'', \mathcal{Q})) = \tau_{\mathit{pws}}(\mathcal{Q}'')$, so by Lemma A.6, $\tau_{\mathit{pws}}(\mathcal{Q}'')$ is a fixpoint of $\mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))$. Since $Q = \tau_{\mathit{pws}}(\mathcal{Q}')$ is the least fixpoint of $\mathcal{D}^c_{\tau_{\mathit{theory}}(\mathcal{T})}(\cdot, \tau_{\mathit{pws}}(\mathcal{Q}))$, $\tau_{\mathit{pws}}(\mathcal{Q}') \leq_K \tau_{\mathit{pws}}(\mathcal{Q}'')$. Combining the two inequalities, we get $\tau_{\mathit{pws}}(\mathcal{Q}') = \tau_{\mathit{pws}}(\mathcal{Q}'')$, so by Lemma A.9, $\mathcal{Q}' = \mathcal{Q}''$. So $\mathcal{Q}' = \mathrm{lfp}(\mathcal{D}^c_{\mathcal{T}}(\cdot, \mathcal{Q})) = S_{\mathcal{D}^*_{\mathcal{T}}}(\mathcal{Q})$, i.e. $Q = \tau_{\mathit{pws}}(\mathcal{Q}') = \tau_{\mathit{pws}}(S_{\mathcal{D}^*_{\mathcal{T}}}(\mathcal{Q}))$, as required. □

We are now ready to present the proofs of Theorems 5.12 and 5.16.

Proof of Theorem 5.16.
Case 1: $\sigma = \mathsf{Sup}$: Suppose $\mathcal{Q}$ is a universally consistent DPWS. $\mathcal{Q}$ is a Sup-model of $\mathcal{T}$
iff $\mathcal{D}_(\mathcal{Q}) = \mathcal{Q}$
iff $\tau_{\mathit{pws}}(\mathcal{D}_(\mathcal{Q})) = \tau_{\mathit{pws}}(\mathcal{Q})$ by Lemma A.9
iff $\mathcal{D}_{\tau_{\mathit{theory}}(\mathcal{T})}(\tau_{\mathit{pws}}(\mathcal{Q}))$ by Lemma A.7
iff $\tau_{\mathit{pws}}(\mathcal{Q})$ is a Sup-model of $\tau_{\mathit{theory}}(\mathcal{T})$.
Case 2: $\sigma = \mathsf{PSt}$: Similar to Case 1, but using Lemma A.15 instead of Lemma A.7.
Case 3: $\sigma = \mathsf{St}$: follows from Case 2 since St-models are two-valued PSt-models. □

Proof of Theorem 5.12.
Case 1: $\sigma \in \{\mathsf{Sup}, \mathsf{PSt}, \mathsf{St}\}$: follows by combining Theorems 5.16 and 5.17.
Case 2: $\sigma = \mathsf{KK}$: The KK-model of $\mathcal{T}$ is the $\leq_p$-least fixpoint of $\mathcal{D}^*_{\mathcal{T}}$ and the KK-model of $\tau_{\mathcal{T}}(\mathcal{T})$ is the $\leq_p$-least fixpoint of $\mathcal{D}^*_{\tau_{\mathcal{T}}(\mathcal{T})}$. So by Lemma A.10, it is enough to show that for each ordinal number $\alpha > 0$, ${\mathcal{D}^*_{\mathcal{T}}}^\alpha((\bot, \top)_{A \in \mathcal{A}})$ is universally consistent and $\tau_{\mathit{beliefpair}}({\mathcal{D}^*_{\mathcal{T}}}^\alpha((\bot, \top)_{A \in \mathcal{A}}) = {\mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}}^\alpha((\bot, \top))$. We prove this by transfinite induction.

For $\alpha = 1$, this is follows from Lemma A.1.

Suppose it is true for α. Then

$$\begin{aligned}\tau_{\mathit{beliefpair}}({\mathcal{D}^*_{\mathcal{T}}}^{\alpha+1}((\bot,\top)_{A\in\mathcal{A}}) &= \tau_{\mathit{beliefpair}}(\mathcal{D}^*_{\mathcal{T}}({\mathcal{D}^*_{\mathcal{T}}}^{\alpha}((\bot,\top)_{A\in\mathcal{A}})) \\ &= \mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}(\tau_{\mathit{beliefpair}}({\mathcal{D}^*_{\mathcal{T}}}^{\alpha}((\bot,\top)_{A\in\mathcal{A}})) \text{ by Lemma A.6} \\ &= {\mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}}^{\alpha+1}((\bot,\top)) \text{ by assumption about } \alpha.\end{aligned}$$

Now suppose it is true for all $\alpha < \lambda$. Then

$$\begin{aligned}&\tau_B({\mathcal{D}^*_{\mathcal{T}}}^{\lambda}((\bot,\top)_{A\in\mathcal{A}})) \\ &\quad= \tau_{\mathit{beliefpair}}(\mathrm{lub}(\{{\mathcal{D}^*_{\mathcal{T}}}^{\alpha}((\bot,\top)_{A\in\mathcal{A}}) \mid \alpha < \lambda\})) \\ &\quad= \mathrm{lub}(\tau_{\mathit{beliefpair}}[\{{\mathcal{D}^*_{\mathcal{T}}}^{\alpha}((\bot,\top)_{A\in\mathcal{A}}) \mid \alpha < \lambda\}]) \text{ by Lemma A.13} \\ &\quad= \mathrm{lub}(\{{\mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}}^{\alpha}((\bot,\top)) \mid \alpha < \lambda\}) \\ &\quad= {\mathcal{D}^*_{\tau_{\mathit{theory}}(\mathcal{T})}}^{\lambda}((\bot,\top))\end{aligned}$$

Case 3: $\sigma = \mathsf{WF}$: Similar to Case 2, but using Lemma A.15 instead of Lemma A.6. □