Distributed Autoepistemic Logic and its Application to Access Control

Pieter Van Hertum
KU Leuven
Leuven, Belgium
pieter.vanhertum@cs.kuleuven.be

Marcos Cramer
Universtiy of Luxembourg
Luxembourg, Luxembourg
marcos.cramer@uni.lu

Bart Bogaerts
Aalto University
Espoo, Finland
bart.bogaerts@aalto.fi

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

Abstract

In this paper we define and study an extension of autoepistemic logic (AEL) called *distributed autoepistemic logic* (dAEL) with multiple agents that have full introspection in their own knowledge as well as in that of others. This mutual full introspection between agents is motivated by an application of dAEL in access control. We define 2- and 3-valued semantic operators for dAEL. Using these operators, approximation fixpoint theory, an abstract algebraic framework that unifies different knowledge representation formalisms, immediately yields us a family of semantics for dAEL, each based on different intuitions that are well-studied in the context of AEL. The application in access control also motivates an extension of dAEL with inductive definitions (dAEL(ID)). We explain a use-case from access control to demonstrate how dAEL(ID) 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. Multiple 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. $k\ says\ \varphi$ is usually rendered as "k supports φ", which can be interpreted to mean that k has issued statements that – together with some 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.

We argue that it is reasonable to assume that the statements issued by a principal are a complete characterization of what the agent supports (Section 5). 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. To tackle this problem, we define *distributed autoepistemic logic (dAEL)*. The application of access control also motivates an extension of dAEL with *inductive definitions* called *dAEL(ID)*.

Autoepistemic logic was designed to model *knowledge*, including knowledge derived from reasoning about knowledge, but can be applied to model other modalities too. Note that when making a claim about an agent's knowledge, we make a claim about his internal state of mind. However, the formalism of AEL does not presuppose that its K modality represents an agent's internal state of mind.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, we assume agents to 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 the $K\phi$ is interpreted to mean that an agent has issued statements that imply ϕ.

2 Preliminaries

We assume familiarity with the basic concepts of first-order logic. 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 $\mathbf{u}^{-1} = \mathbf{u}$. We assume throughout this paper that a first-order vocabulary Σ is fixed and use $\mathcal{L}$ for the language of standard first-order logic over Σ. 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 treated as abbreviations in the standard way.

2.1 Autoepistemic Logic

The language $\mathcal{L}_k$ of *autoepistemic logic* [Moore, 1985b] is defined recursively using the standard rules for the syntax of first-order logic, augmented with: $K(\psi) \in \mathcal{L}_k$ if $\psi \in \mathcal{L}_k$.

An AEL theory T is a set of formulas over $\mathcal{L}_k$. 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. Furthermore, we assume that for each $d \in D$, d is a constant symbol of $\mathcal{L}_k$ whose interpretation in all structures is d. A *possible world structure* is a set of structures. 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 less structures possess more knowledge, or formally $Q_1 \leq_K Q_2$ holds if and only if $Q_2 \subseteq Q_1$.

The semantics of AEL is based on the S5 truth assignment. The *value* of a sentence $\varphi \in \mathcal{L}_k$ with respect to a possible world structure Q and a structure I (denoted $\varphi^{Q,I}$) is defined using the recursive rules for first-order logic augmented with:

$$(K\varphi)^{Q,I} = \mathbf{t} \qquad \text{if } \varphi^{Q,J} = \mathbf{t} \text{ for each } J \in Q.$$

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.

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 abstract algebraical framework *approximation fixpoint theory* (AFT) that a uniform view on those different semantics was obtained.

2.2 Approximation Fixpoint Theory

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

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 $\mathrm{glb}(S)$. 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* if $O(x) = x$. Every monotone operator O in a complete lattice has a least fixpoint, denoted $\mathrm{lfp}(O)$.

Given a lattice L, AFT makes use of the bilattice L^2. 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 $[x, y]$ is non-empty and 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 $(x, x) \in L^c$. The *precision order* on L^2 is 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). 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 can also derive the 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 and 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.

2.3 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 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 and S is an overestimation of Q. That is, P contains all knowledge the agent certainly has and S all knowledge the agent possibly has. From now on, we assume all belief pairs to be consistent. Belief pairs are ordered by a precision ordering $\leq_p$.

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. The *value* of φ with respect to belief pair B and interpretation I (notation $\varphi^{B,I}$) is defined inductively:

$$
\begin{aligned}
(P(\bar{t}))^{B,I} &= \quad \bar{t}^I \in P^I \\
(\neg\varphi)^{B,I} &= \quad (\varphi^{B,I})^{-1} \\
(\varphi \wedge \psi)^{B,I} &= \quad \mathrm{glb}_{\leq_t}(\varphi^{B,I}, \psi^{B,I}) \\
(\forall x : \varphi)^{B,I} &= \quad \mathrm{glb}_{\leq_t}\{\varphi[x/d]^{B,I} \mid x \in D\} \\
(K\varphi)^{(P,S),I} &= \quad \begin{cases} \mathbf{t} & \text{if } \varphi^{(P,S),I'} = \mathbf{t} \text{ for all } I' \in P \\ \mathbf{f} & \text{if } \varphi^{(P,S),I'} = \mathbf{f} \text{ for some } I' \in S \\ \mathbf{u} & \text{otherwise} \end{cases}
\end{aligned}
$$

The logical connectives combine truth values based on Kleene's truth tables [Kleene, 1938]. DMT (2000) defined the bilattice operator D_T^* that maps (P, S) to (P', S') where

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

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. 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.

3 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. We assume $\mathcal{A}$ to be a subset of the domain D over which all structures are defined.

3.1 Syntax and Basic Semantic Notions

Definition 3.1. We define the language $\mathcal{L}_d$ of distributed autoepistemic logic using the standard recursive rules of first-order logic, augmented with:

$$K_A(\psi) \in \mathcal{L}_d \text{ if } \psi \in \mathcal{L}_d \text{ and } A \in \mathcal{A}$$

In a distributed setting, different agents each have their own theory describing their beliefs or knowledge about the world. To represent the knowledge of multiple agents, we generalise the notion of a possible world structure. 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 DPWS's. One DPWS contains more knowledge than another if each agent has more knowledge: given two DPWS's $\mathcal{Q}^1$ 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 3.2. The *value* of a sentence φ in $\mathcal{Q}, I$ (denoted $\varphi^{\mathcal{Q},I}$) is defined inductively by the standard recursive rules for first-order logic, augmented with:

$$(K_A\varphi)^{\mathcal{Q},I} = \mathbf{t} \qquad \text{if } \varphi^{\mathcal{Q},J} = \mathbf{t} \text{ for each } J \in \mathcal{Q}_A$$

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

Definition 3.3. A *distributed belief pair* is an indexed family $\mathcal{B} = (\mathcal{B}_A)_{A\in\mathcal{A}}$, where for each $A \in \mathcal{A}$, $\mathcal{B}_A$ is a pair (P_A, S_A) of possible world structures.

The precision order on distributed belief pairs is a pointwise extension of the precision order on belief pairs. By abuse of notation, we sometimes identify $\mathcal{B}$ with a pair of distributed possible world structures $(\mathcal{B}^c, \mathcal{B}^l)$. The following proposition follows easily from the equivalent result in AEL.

Proposition 3.4. *The set of all DPWS's 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 $\leq_p$. The latter is the bilattice of the former.*

As before, we assume that all distributed belief pairs are consistent. The notion of three-valued valuations is extended to the distributed setting by evaluating each modal operator with respect to the correct agent.

Definition 3.5. The *value* of φ with respect to distributed belief pair $\mathcal{B}$ and interpretation I (notation $\varphi^{\mathcal{B},I}$) is defined inductively by replacing the fifth rule in the recursive definition of the three-valued valuation of an AEL formula by:

$$(K_A\varphi)^{\mathcal{B},I} = \begin{cases} \mathbf{t} & \text{if } \varphi^{\mathcal{B},I'} = \mathbf{t} \text{ for all } I' \in \mathcal{B}^c_A \\ \mathbf{f} & \text{if } \varphi^{\mathcal{B},I'} = \mathbf{f} \text{ for some } I' \in \mathcal{B}^l_A \\ \mathbf{u} & \text{otherwise} \end{cases}$$

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

3.2 Semantics of dAEL through AFT

The two- and three-valued valuations form the building blocks to extend the semantic operator and its approximator from AEL to dAEL.

Definition 3.6. 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. 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 3.7. 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

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

Theorem 3.8. *$\mathcal{D}^*_\mathcal{T}$ is an approximator of $\mathcal{D}_\mathcal{T}$.*

The stable operator $\mathcal{D}^{st}_\mathcal{T}$ is defined for dAEL as $\mathcal{D}^{st}_\mathcal{T}(\mathcal{Q}) = \text{lfp}(\mathcal{D}^*_\mathcal{T}(\cdot, \mathcal{Q})^c)$. Different fixpoints of these operators lead to different semantics as discussed in Section 2.2;

Definition 3.9. Let $\mathcal{T}$ be a distributed theory.

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

Example 3.10. 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

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

The child, who has an inherent comprehension of dAEL, now has to choose between the various semantics. The following analysis helps him choose. 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: $\{\{c\}\}$ **(3)** The disbelief of c: $\{\emptyset\}$ **(4)** The lack of knowledge: $\{\emptyset, \{c\}\}$, denoted as $\bot$. There are two *supported models*, namely $(\bot_D, \bot_M)$ and $(\{\{c\}\}_D, \{\{c\}\}_M)$: either both Dad and Mom agree to giving candy or none of them does. The *Kripke-Kleene model* is $((\bot, \{\{c\}\})_D, (\bot, \{\{c\}\})_M)$. So in the Kripke-Kleene semantics it is unknown for both Dad and Mom whether the child can have candy. However, from none of their theories it follows that the child can have no candy. The DPWS $\bot := (\bot_D, \bot_M)$ is the (unique) *stable model*: $(\bot, \bot)$ is a *partial stable* model, since $\bot = \text{lfp}(\mathcal{D}^*_{\mathcal{T}}(\cdot, \bot))^c$ and $\bot = \text{lfp}(\mathcal{D}^*_{\mathcal{T}}(\cdot, \bot))^l$. $(\bot, \bot)$ is the *well-founded model*.

The stable and well-founded semantics only derive knowledge that is "grounded" 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.

4 dAEL with Inductive Definitions

In this section, we discuss how to extend dAEL with (inductive) definitions (IDs). There are two main reasons for this extension: **(1)** IDs are a common concept in all branches of mathematics; as such, we expect them to be useful as well when reasoning about knowledge **(2)** in the application to access control, the need for IDs arises naturally (see Section 5).

4.1 Preliminaries: Inductive Definitions

A definition Δ over a language $\mathcal{L}$ is a set of rules δ of the form: $P(\bar{t}) \leftarrow \varphi$ with $\varphi \in \mathcal{L}$. We call $P(\bar{t})$ the head ($head(\delta)$) and φ the body ($body(\delta)$) of that rule. We say that Δ *defines* Q if Δ contains a rule δ with $head(\delta) = Q(\bar{t})$. We use $Def(\Delta)$ to denote all symbols defined in Δ; all other symbols are called *parameters*; the set of all parameters is denoted $Par(\Delta)$. If $\mathcal{O}$ is an interpretation of $Par(\Delta)$ and I a (partial) interpretation of $Def(\Delta)$, we use $\mathcal{O} + I$ to interprete symbols in $Par(\Delta)$ as in $\mathcal{O}$ and other symbols as in I. We assume that an interpretation $\mathcal{O}$ of the parameters is fixed.

AFT defines a family of semantics for inductive definitions based on a slight generalisation of the immediate consequence operator defined by van Emden and Kowalski [1976]. $T_{\Delta,\mathcal{O}}$ maps an interpretation I of $Def(\Delta)$ to an interpretation I' of $Def(\Delta)$ such that for defined symbols P:

$$P(\bar{d})^{I'} = \bigvee_{\{\delta \in \Delta | head(\delta) = P(\bar{t}) \wedge \bar{t}^{I+\mathcal{O}} = \bar{d}\}} body(\delta)^{I+\mathcal{O}}. \quad (1)$$

This operator was extended by Fitting [1985] for the three-valued setting to an operator $\Psi_{\Delta,\mathcal{O}}$, mapping a *partial* $Def(\Delta)$-interpretation I to a partial $Def(\Delta)$-interpretation I' such that also Equation (1) holds, now simply replacing two-valued truth valuation by Kleene-valuation.

DMT (2000) showed that for each $\mathcal{O}$, $\Psi_{\Delta,\mathcal{O}}$ is an approximator of $T_{\Delta,\mathcal{O}}$ and obtained a family of semantics for such definitions. Denecker and Vennekens [2014] have argued that the well-founded semantics correctly formalises our intuition and hence that it is actually the *right* semantics. Following them, we use the well-founded semantics for IDs; our work can be generalised to allow other semantics for IDs. Given an interpretation $\mathcal{O}$ of the parameters of Δ, we write $wfm_\Delta(\mathcal{O})$ for the interpretation $\mathcal{O} + I$, where I is the $\Psi_{\Delta,\mathcal{O}}$-well-founded fixpoint of $T_{\Delta,\mathcal{O}}$. The well-founded model is also defined if the parameters are only interpreted partially, i.e., in case $\mathcal{O}$ is a partial interpretation. In this case, $wfm_\Delta(\mathcal{O})$ is a partial interpretation as well.

4.2 Syntax and semantics of dAEL(ID)

dAEL(ID) extends dAEL with modal inductive definitions, where the bodies of rules can contain modal operators. We use $\mathcal{L}_d(ID)$ as shorthand for this language. Formally, $\mathcal{L}_d(ID)$ consists of logical formulas as in $\mathcal{L}_d$ and modal inductive definitions. A distributed theory with inductive definitions is an indexed family $\mathcal{T} = (\mathcal{T}_A)_{A \in \mathcal{A}}$ of theories, i.e. sets containing $\mathcal{L}_d$ formulas and modal inductive definitions. The semantics of definitions mainly remains unchanged. All we need to take care of is evaluate modal literals with respect to the distributed belief pair. As such, we define the immediate consequence (bilattice) operator $\Psi_{\Delta,\mathcal{O},\mathcal{B}}$ that maps a partial $Def\,\Delta$-interpretation I to I' such that

$$P(\bar{d})^{I'} = \bigvee_{\{\delta \in \Delta | head(\delta) = P(\bar{t}) \wedge \bar{t}^{I+\mathcal{O}} = \bar{d}\}} body(\delta)^{\mathcal{B},I+\mathcal{O}}.$$

We write $wfm_\Delta(\mathcal{B}, \mathcal{O})$ for the (possibly three-valued) $\Psi_{\Delta,\mathcal{B},\mathcal{O}}$-well-founded fixpoint.

We first define a valuation of such definitions with respect to distributed belief pairs (this can also be used for DPWSs).

$$\Delta^{\mathcal{B},I} = \begin{cases} \mathbf{t} & \text{if } I = wfm_\Delta(\mathcal{B}, I|_{Par(\Delta)}) \\ \mathbf{f} & \text{if } I \not\geq_p wfm_\Delta(\mathcal{B}, I|_{Par(\Delta)}) \\ \mathbf{u} & \text{otherwise} \end{cases}$$

The intuition here is: $\mathcal{B}$ provides partial information about the state of mind of agents. If I is $wfm_\Delta(\mathcal{B}, I|_{Par(\Delta)})$, then this partial information suffices to determine that the definition is satisfied. Similarly, if $I \not\geq_p wfm_\Delta(\mathcal{B}, I|_{Par(\Delta)})$, this information is enough to determine that the definition is not satisfied. Otherwise, the truth of the definition is still unknown (more information on $\mathcal{B}$ is needed to determine it).

The knowledge operator and approximator for $\mathcal{L}_d(ID)$ are, using this valuation, simple extensions of those in dAEL. The approximation is defined similarly by evaluating all formulas φ and definitions Δ with respect to $\mathcal{B}, I$.

5 Applying dAEL(ID) to Access Control

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 iff it is logically entailed by the policy.

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. $k\ says\ \varphi$ 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.

We illustrate the $says$-modality in access control by showing how it is employed to delegate authority. Suppose that principal A has control over a resource r, i.e., that any principal i has access to r if and only if A says that i has access to r. Now A can delegate to principal C the decision whether B has access to r by issuing the statement

$$(C \; says \; access(B, r)) \Rightarrow access(B, r). \tag{2}$$

If C issues $access(B, r)$, then (2) implies $access(B, r)$, i.e. $A \; says \; access(B, r)$, so B has access to r.

Note that we used the fact that C said $access(B, r)$ in order to derive what A supports from what A explicitly said: we assumed $A \; says \; (C \; says \; access(B, r))$ based on $C \; says \; access(B, r)$. To make delegation work in general, practically all $says$-based logics statements allow us to derive $j \; says \; (k \; says \; \varphi)$ from $k \; says \; \varphi$. Note that in epistemic terminology, by identifying $k \; says$ with K_k, this can be called mutual positive introspection between principals.

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. For example BL, an access control logic with support for system state and explicit time [Garg, 2009; Garg and Pfenning, 2012], is an intuitionistic modal logic with support for mutual positive introspection but not for mutual negative introspection. dAEL(ID), on the other hand, is based on classical logic, and supports mutual negative introspection between principals. In order to justify our claim that dAEL(ID) is a good access control logic, we discuss these two differences between BL and dAEL(ID).

We illustrate the advantage of mutual negative introspection by showing how it allows to correctly handle statements whose goal it is to deny or revoke access rights. 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. A natural way for A to do this 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 him access. However, this effect can only be achieved if we assume the $says$-modality to allow mutual negative introspection: A must know that C does *not* issue a statement $\neg access(B, r)$ to derive that B has access rights.

In order to derive statements of the form $\neg k \; says \; \varphi$, we have to assume the statements issued to be a complete characterization of what the agent supports, like the "All I Know" assumption for AEL [Levesque, 1990]. Together with support for mutual positive and negative introspection, this motivates the use of dAEL as a viable access control logic.

The addition of inductive definitions to dAEL allows principals to *define* access rights and other properties relevant for access control through inductive (recursive) definitions. Denecker et al. [2000] showed that in classical logics, adding definitions leads to a strictly more expressive language. We illustrate the advantage of inductive definitions for access control by showing how a certain access control problem related to the revocation of delegated rights can be modelled in a natural and concise way in dAEL(ID).

When principals delegate access rights to others, delegation chains can form. There are different ways to treat these delegation chains when revoking rights, which give rise to different revocation schemes [Hagström *et al.*, 2001; Cramer *et al.*, 2015]. 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 revocations and whose effect propagates forward. Our dAEL(ID) model of SGN revocation behaves precisely like it was defined by Cramer *et al.* [2015].

Suppose that A controls a resource r and that A wants to delegate access right to other principals, along with SGN revocation right. 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 $revoke(j)$. Assuming that access will be granted to a principal k iff $A \; says \; access(k, r)$, A can ensure that the statements of the form $deleg\_to(j)$ and $revoke(j)$ will be interpreted as delegation and SGN revocation by issuing the following inductive definition of the predicate $access$:[1]

$$\left\{ \begin{array}{l} access(A, r). \\ access(j, r) \leftarrow \\ \quad \exists k \; (A \; says \; access(k, r) \wedge k \; says \; deleg\_to(j)) \wedge \\ \quad \neg \exists i \; (A \; says \; access(i, r) \wedge i \; says \; revoke(j)). \end{array} \right\}$$

If $k \; says \; \varphi$ is interpreted as $K_k \varphi$ with the well-founded semantics, this definition has the intended interpretation.

We now argue using two example scenarios why we believe the well-founded semantics to be the best choice when applying dAEL(ID) to access control. In both scenarios, we assume that A controls r and that A has issued the inductive definition above to delegate access right to other principals and allow them to perform SGN revocation.

Given a certain semantics for dAEL(ID), it is reasonable to grant k access to r only if $K_A \, access(k, r)$ holds in all models. With this interpretation of the semantics, the partial stable semantics and the well-founded semantics coincide. Therefore we ignore the partial stable semantics for the discussion of semantics in this section.

In the first scenario, we suppose A has issued the statements $deleg\_to(B)$ and $deleg\_to(C)$, that B has issued the statements $revoke(C)$ and $deleg\_to(D)$, and that C has issued the statements $revoke(B)$ and $deleg\_to(D)$. By issuing $revoke(C)$, B is attempting to revoke C's access right (and vice versa). Of course, this attempt is only successful

[1] We use $\exists k$ in this definition as an abbreviation for a disjunction, containing a disjunct for each instantiation of k by an agent.

if B has access. So C should have access right iff B does not. Since the scenario is symmetric between B and C, they should either both be granted or both be denied access right. 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 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. To avoid security risks neither B nor C should have access. 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 this 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]. Consider the statements issued by the principals as a distributed theory in dAEL(ID). This theory has different models depending on the choice of semantics. We present the models by presenting a set of expressions X^t where X is a principle and t the truth value of $K_A access(X, r)$ in the model. There are two supported models $\{A^{\mathbf{t}}, B^{\mathbf{t}}, C^{\mathbf{f}}, D^{\mathbf{t}}\}$ and $\{A^{\mathbf{t}}, B^{\mathbf{f}}, C^{\mathbf{t}}, D^{\mathbf{t}}\}$. These are also the stable models. The Kripke-Kleene model and the well-founded model are identical: $\{A^{\mathbf{t}}, B^{\mathbf{u}}, C^{\mathbf{u}}, D^{\mathbf{u}}\}$. This model is not exact: The truth-value of the statements $A\ says\ access(X, r)$, with $X \in \{B, C, D\}$ is unknown. Note that all supported and stable models both grant access to D. Given our above argument against granting access to D, these semantics cannot be considered viable for this application.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. Thus these semantics, while not based on intuitionistic logic, are faithful to the motivation that Garg and Pfenning [2012] gave for using intuitionistic logic in access control. Furthermore, they exhibit the existing conflict between B and C by making their access right status undefined.

Consider a second scenario, in which 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 $\{A^{\mathbf{t}}, B^{\mathbf{u}}, C^{\mathbf{u}}, D^{\mathbf{f}}\}$ of the distributed theory corresponding to this scenario is not exact; it is unknown whether B and C have access; this clearly diverges from our requirements. The well-founded model $\{A^{\mathbf{t}}, B^{\mathbf{t}}, C^{\mathbf{f}}, D^{\mathbf{f}}\}$ on the other hand correctly computes this desired outcome.

From these scenarios, we can see that the only semantics for dAEL(ID) 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 in AEL.

6 Related Work

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 generalises 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. They do not consider distributed theories and do not assume introspection. Belle et al. [2015] also studied multi-agent theories in the same setting but added only knowing and common knowledge constructs. Toyama et al. [2002] has distributed theories, does assume introspection, uses propositional logic and only uses one of the semantics we discussed: supported semantics. Permpoontanalarp and Jiang [1995] study a number of logics and develop a proof theory that extends the logic of Morgenstern et al.; they do not define a semantics. A motivation of them is that the logic of Morgenstern et al. has some undesirable properties if reduced to the single agent case, where it differs from AEL. The logic we defined is equivalent to AEL when reduced to the single agent case. Vlaeminck *et al.* [2012] defined an extension to AEL with multiple agents, but this extension requires a global stratification on the agents, which is undesirable for a 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; 2015b], *well-founded set semantics* [Bogaerts, 2015], *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.

7 Conclusion and Future Work

Motivated by an application in access control, we extended AEL to a distributed setting, resulting in *dAEL*: distributed autoepistemic logic. dAEL allows for a set of agents to each have their own theory and refer to each others knowledge. For this, the K operator of AEL is replaced by an indexed operator K_A, where A refers to an agent. We defined the semantics of this logic using AFT. Further motivated by applications in access control, we defined dAEL(ID): the extension of dAEL with inductive definitions. We discussed the usability of this new logic in access control and illustrated it with examples.

We conclude with an overview of possible future research based on the findings of this paper. In this paper we study the semantics of dAEL(ID), but for practical applications, a decision procedure for dAEL(ID) or an expressively rich subset of it needs to be developed. The complexity of determining access rights based on a theory written in dAEL(ID) should be studied. Furthermore, the relation of dAEL(ID) to various existing access control models needs to be studied further.

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*.

Bart Bogaerts is supported by the Finnish Center of Excellence in Computational Inference Research (COIN) funded by the Academy of Finland (under grant #251170).

References

[Abadi, 2003] Martín Abadi. Logic in Access Control. In *Proceedings of LICS*, pages 228–233, 2003.

[Abadi, 2008] Martín Abadi. Variations in Access Control Logic. In *Proceedings of DEON*, pages 96–109, 2008.

[Belle and Lakemeyer, 2015] Vaishak Belle and Gerhard Lakemeyer. Only knowing meets common knowledge. In *Proceedings of IJCAI*, pages 2755–2761, 2015.

[Bogaerts *et al.*, 2015a] Bart Bogaerts, Joost Vennekens, and Marc Denecker. Grounded fixpoints and their applications in knowledge representation. *AIJ*, 224:51–71, 2015.

[Bogaerts *et al.*, 2015b] Bart Bogaerts, Joost Vennekens, and Marc Denecker. Partial grounded fixpoints. In *Proceedings of IJCAI*, pages 2784–2790, 2015.

[Bogaerts, 2015] Bart Bogaerts. *Groundedness in logics with a fixpoint semantics*. PhD thesis, 2015.

[Cramer *et al.*, 2015] Marcos Cramer, Diego Agustin Ambrossio, and Pieter Van Hertum. A Logic of Trust for Reasoning about Delegation and Revocation. In *Proceedings of SACMAT*, pages 173–184. ACM, 2015.

[Denecker and Vennekens, 2014] Marc Denecker and Joost Vennekens. The well-founded semantics is the principle of inductive definition, revisited. In *Proceedings of KR*, pages 22–31, 2014.

[Denecker *et al.*, 1998] Marc Denecker, Victor Marek, and Mirosław Truszczyński. Fixpoint 3-valued semantics for autoepistemic logic. In *Proceedings of AAAI*, pages 840–845. MIT Press, 1998.

[Denecker *et al.*, 2000] Marc Denecker, Victor Marek, and Mirosław Truszczyński. Approximations, stable operators, well-founded fixpoints and applications in nonmonotonic reasoning. In *Logic-Based Artificial Intelligence, Springer*, volume 597, pages 127–144, 2000.

[Denecker *et al.*, 2003] Marc Denecker, Victor Marek, and Mirosław Truszczyński. Uniform semantic treatment of default and autoepistemic logics. *AIJ*, 143(1):79–122, 2003.

[Denecker *et al.*, 2004] 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.

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

[Denecker, 2000] Marc Denecker. Extending classical logic with inductive definitions. In *Proceedings of CL*, pages 703–717, 2000.

[Fitting, 1985] Melvin Fitting. A Kripke-Kleene semantics for logic programs. *The Journal of Logic Programming*, 2(4):295 – 312, 1985.

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

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

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

[Gurevich and Neeman, 2008] Yuri Gurevich and Itay Neeman. DKAL: Distributed-knowledge authorization language. In *Proceedings CSF*, pages 149–162. IEEE, 2008.

[Hagström *et al.*, 2001] Åsa Hagström, Sushil Jajodia, Francesco Parisi-Presicce, and Duminda Wijesekera. Revocations – A Classification. In *Proceedings of CSF*, pages 44–58. IEEE, 2001.

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

[Kleene, 1938] S. C. Kleene. On notation for ordinal numbers. *The Journal of Symbolic Logic*, 3(4):150–155, 1938.

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

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

[Moore, 1985a] Robert C. Moore. A Formal Theory of Knowledge and Action. In *Formal Theories of the Commonsense World*, pages 319–358. Springer-Verlag, 1985.

[Moore, 1985b] Robert C. Moore. Semantical considerations on nonmonotonic logic. *AIJ*, 25(1):75–94, 1985.

[Morgenstern, 1990] Leora Morgenstern. A formal theory of multiple agent nonmonotonic reasoning. In *Proceedings of AAAI*, pages 538–544, 1990.

[Niemelä, 1991] Ilkka Niemelä. Constructive tightly grounded autoepistemic reasoning. In *Proceedings of IJCAI*, pages 399–405, 1991.

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

[Strass, 2013] Hannes Strass. Approximating operators and semantics for abstract dialectical frameworks. *AIJ*, 205:39–70, 2013.

[Toyama *et al.*, 2002] Katsuhiko Toyama, Takahiro Kojima, and Yasuyoshi Inagaki. Translating multi-agent autoepistemic logic into logic program. In *Proceedings of CLIMA*, pages 49–62, 2002.

[van Emden and Kowalski, 1976] Maarten H. van Emden and Robert A. Kowalski. The semantics of predicate logic as a programming language. *J. ACM*, 23(4):733–742, 1976.

[Vlaeminck *et al.*, 2012] Hanne Vlaeminck, Joost Vennekens, Maurice Bruynooghe, and Marc Denecker. Ordered Epistemic Logic: Semantics, complexity and applications. In *Proceedings of KR*, pages 369–379, 2012.