Decision Procedures for Epistemic Logic Exploiting Belief Bases

Emiliano Lorini, Benito Fabian Romero Jimenez

Decision Procedures for Epistemic Logic Exploiting Belief Bases

Emiliano Lorini
IRIT-CNRS, Toulouse University
Toulouse, France
Emiliano.Lorini@irit.fr

Fabian Romero
IRIT, Toulouse University
Toulouse, France
Benito.Romero@irit.fr

ABSTRACT

We provide tableau-based PSPACE satisfiability checking procedures for a family of multi-agent epistemic logics with a semantics defined in terms of belief bases. Such logics distinguish an agent's explicit beliefs, i.e., all facts included in the agent's belief base, from the agent's implicit beliefs, i.e., all facts deducible from the agent's belief base. We provide a simple dynamic extension for one of these logics by propositional assignments performed by agents. A propositional assignment captures a simple form of action that changes not only the environment but also the agents' beliefs depending on how they jointly perceive its execution. After having provided a PSPACE satisfiability checking procedure for this dynamic extension, we show how it can be used in human-robot interaction in which both the human and the robot have higher-order beliefs about the other's beliefs and can modify the environment by acting.

ACM Reference Format:
Emiliano Lorini and Fabian Romero. 2019. Decision Procedures for Epistemic Logic Exploiting Belief Bases. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 9 pages.

1 INTRODUCTION

In [28], a new epistemic logic called LDA (Logic of Doxastic Attitudes) has been introduced. LDA supports reasoning about explicit beliefs and implicit beliefs of multiple agents. The distinction between *explicit belief* and *implicit belief* has been widely discussed in the literature on knowledge representation and reasoning. According to [26], "...a sentence is explicitly believed when it is actively held to be true by an agent and implicitly believed when it follows from what is believed" (p.198). In other words, explicit beliefs correspond to an agent's *actual* beliefs, whereas implicit beliefs correspond to her *potential* ones. This distinction is also acknowledged by Fagin & Halpern (F&H)'s logic of general awareness (LGA) [13]: it defines explicit belief as a formula implicitly believed by an agent and of which the agent is aware.[1]

The logic LDA accounts for the connection between the concept of explicit belief and the concept of belief base: an agent's belief base, which is not necessarily closed under deduction, includes all facts that are explicitly believed by the agent. Differently from existing Kripke-style semantics for epistemic logic — exploited, among other logics, by F&H's logic of general awareness — in which the notion of doxastic alternative is primitive, in the LDA semantics the notion of doxastic alternative is defined from, and more generally grounded on, the concept of belief base.[2]

The main motivation behind the logic LDA is to bridge two traditions that have rarely talked to each other up to now. On the one hand, we have epistemic logic: it started in the 60ies with the seminal work of Hintikka [22] on the logics of knowledge and belief, it was extended to the multi-agent setting at the end of 80ies [12] and then furtherly developed during the last 20 years, the period of the "dynamic turn", with growing research on dynamic epistemic logic [42]. On the other hand, we have syntactic approaches to knowledge representation and reasoning mainly proposed in the area of artificial intelligence (AI). The latter includes for instance work on belief base and knowledge base revision [6, 17, 18], belief base merging [23], input-output logic [31], as well as more recent work on the so-called "database perspective" to the theory of intention by [37] and resource-bounded knowledge and reasoning about strategies [1]. All these approaches defend the idea that right level of abstraction for understanding and modelling cognitive processes and phenomena is the "belief base" level or, more generally, the "cognitive attitude base" level. The latter consists in identifying a cognitive agent with the sets of facts that she believes (belief base), desires (desire base) and intends (intention base) and in studying the interactions between the different bases.[3]

There is also a practical motivation behind the logic LDA in relation to modeling Theory of Mind (ToM) in human-machine interaction (HMI) applications including social robots [36, 44, 45] and intelligent virtual agents (IVAs) [9, 20, 33, 35]. An essential aspect of ToM is the general capacity of an agent to form higher-order beliefs about beliefs of other agents. Although existing robotic models of ToM take this aspect into consideration (see, e.g., [10, 25, 32]), they have some limitations. First of all, they only allow to represent higher-order beliefs of depth at most 2, where the depth of a higher-order belief is defined inductively as follows: (i) an agent's belief has depth 1 if and only if its content is an objective formula that does not mention beliefs of others (e.g., an agent i's belief that it is a sunny day); (ii) an agent's belief has depth n if and only it is a belief about a belief of depth $n-1$ of another agent (e.g., an agent i's belief of depth 2 that another agent j believes that it is a sunny day). Secondly, models of ToM used for robotic implementations do not have a high level of generality that makes them applicable

[1] For the connection between LDA and LGA, see [29], in which a polynomial embedding of the former into the latter is provided.

[2] Grounded semantics for epistemic logics have proposed in the AI literature. For instance, [27] provide a semantics based on the concept of interpreted system, while the approach by [11, 41] builds on propositional observability.

[3] This approach has also been used in linguistic work on modal expressions. For instance, according to [24], conversational common ground can be seen as set of formulas shared by the interlocutors and the set of worlds that are considered possible by the interlocutors are those worlds that satisfy all formulas in the common ground.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

to different scenarios and situations. As shown by [7], the standard epistemic logic approach [12] allows us to overcome these limitations by allowing to represent higher-order belief of any depth and by offering a general framework for formalizing a large variety of interactive situations between artificial robots and human agents. Unfortunately, it does not distinguish between explicit and implicit beliefs. It only allows to represent what a robot believes a human could *potentially* believe — if she had enough computational resources and time to infer it —, without representing what the robot believes a human *actually* believes. Another limitation of the standard epistemic logic approach arises when trying to model complex forms of information dynamics in a multi-agent setting. An example of this is the situation in which a certain action or event takes place, some agents in the system perceive it while others do not, some agents perceive that some agents perceive it, some agents perceive that some agents perceive that some agents perceive it, and so on so forth. Such scenarios are typically modeled in the context of dynamic epistemic logic (DEL) [42] — the dynamic extension of epistemic logic — with the help of so-called action models [4]. However, modeling complex information dynamics in DEL comes with a price. In DEL, whenever an agent *privately* receives a piece of information, the original epistemic model has to be duplicated by creating one copy of the model for the perceiver in which her beliefs have changed and one copy for the non-perceivers in which their beliefs have not changed [15]. Thus, the original epistemic model grows exponentially in the length of the sequence of private announcements. Furthermore, although extending multi-agent epistemic logic by simple notions of state eliminating public announcement or arrow eliminating private announcement does not increase its PSPACE complexity (see, e.g., [8, 30]), complexity increases if we move into the realm of full DEL, whose satisfiability problem is known to be NEXPTIME-complete [2].

The logic LDA provides a generalization of the standard epistemic approach in which the distinction between explicit and implicit belief can be captured: it allows us to represent both what a robot believes a human is explicitly believing in a given situation — which is the essential aspect of ToM — and what a robot believes a human can infer from what she explicitly believes. Moreover, differently from standard DEL, LDA can be easily extended to represent rich belief dynamics in a multi-agent setting, with no increase in complexity with respect to the static logic. The interesting aspect of LDA is that it allows us to model belief dynamics as operations modifying the belief bases of some agents. This leads to a 'parsimonious' account of private informative actions that — differently from traditional DEL — does not require to duplicate epistemic models and to make them exponentially larger.

The aim of the present paper is to explore the practical aspect of the logic LDA: we provide a decision procedure for LDA and for a dynamic extension of it, we discuss its potential applicability in human-robot interaction. The paper is organized as follows. In Section 2, we recall the syntax and the semantics of the logics in the LDA family. Then, in Section 3, we present tableau-based PSPACE satiability checking procedures for these logics. In Section 4, we introduce a dynamic extension of LDA, called DLDA, in which a simple form of action based on propositional assignment can be represented. In DLDA, an assignment is performed by an agent and its execution is *privately* perceived by some agents. In other words, an assignment changes both the environment and the agents' beliefs depending on perceptive situation, that is to say, on whether they perceive its execution, on whether they perceive other agents perceiving its execution, and so on. We show that the satisfiability problem for DLDA remains in PSPACE and we provide a PSPACE decision procedure for DLDA exploiting a polynomial reduction of DLDA satisfiability to LDA satisfiability. Moreover, we instantiate DLDA in a concrete example of human-robot interaction in a dynamic domain in which the human and the robot not only have higher-order beliefs about the beliefs of the other but also can modify the environment by acting. In Section 5, we conclude.

2 A LOGIC OF EXPLICIT AND IMPLICIT BELIEF

LDA is a logic for reasoning about explicit beliefs and implicit beliefs of multiple agents. Assume a countably infinite set of atomic propositions $Atm = \{p, q, \ldots\}$ and a finite set of agents $Agt = \{1, \ldots, n\}$.

We define the language of the logic LDA in two steps. We first define the language $\mathcal{L}_0(Atm, Agt)$ by the following grammar in Backus-Naur Form (BNF):

$$\alpha \quad ::= \quad \bot \mid p \mid \neg\alpha \mid \alpha_1 \wedge \alpha_2 \mid \triangle_i \alpha$$

where p ranges over Atm and i ranges over Agt. $\mathcal{L}_0(Atm, Agt)$ is the language for representing explicit beliefs of multiple agents. The formula $\triangle_i \alpha$ can be read as "agent i explicitly believes that α is true" or "α is in agent i's belief base". In this language, we can represent higher-order explicit beliefs, for example $\triangle_i \triangle_j \alpha$ express the fact that agent i explicitly believes that agent j explicitly believes that α is true.

Language $\mathcal{L}_{\mathsf{LDA}}(Atm, Agt)$, extends language $\mathcal{L}_0(Atm, Agt)$ by modal operators of implicit belief and is defined by the following grammar:

$$\varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i \varphi$$

with α ranging over $\mathcal{L}_0(Atm, Agt)$. For simplicity, we write $\mathcal{L}_0$ instead of $\mathcal{L}_0(Atm, Agt)$ and $\mathcal{L}_{\mathsf{LDA}}$ instead of $\mathcal{L}_{\mathsf{LDA}}(Atm, Agt)$, when the context is unambiguous.

The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from α, $\neg$ and $\wedge$ in the standard way.

The formula $\Box_i \varphi$ has to be read "agent i implicitly (or potentially) believes that φ is true". We define the dual operator $\Diamond_i$ as follows:

$$\Diamond_i \varphi \stackrel{\mathsf{def}}{=} \neg\Box_i \neg\varphi.$$

$\Diamond_i \varphi$ has to be read "φ is compatible (or consistent) with agent i's explicit beliefs".

2.1 Formal semantics

In this section, we present a semantics for the logic LDA in which the notion of doxastic alternative is not primitive but it is defined from the primitive concept of multi-agent belief base.

Definition 2.1 (Multi-agent belief base). A multi-agent belief base (MBB) is a tuple $B = (B_1, \ldots, B_n, S)$ where (i) for every $i \in Agt$, $B_i \subseteq \mathcal{L}_0$ is agent i's belief base, and (ii) $S \subseteq Atm$ is the actual state. The class of MBBs is denoted by $\mathbf{B}$.

Formulas of the language $\mathcal{L}_0$ are interpreted relative to multi-agent belief bases as follows.

Definition 2.2 (Satisfaction relation). Let $B = (B_1, \ldots, B_n, S)$ be a multi-agent belief base. Then, the satisfaction relation $\models$ between states and formulas in $\mathcal{L}_0$ is defined as follows:

$$\begin{aligned} B &\not\models \bot \\ B &\models p \iff p \in S \\ B &\models \neg\alpha \iff B \not\models \alpha \\ B &\models \alpha_1 \wedge \alpha_2 \iff B \models \alpha_1 \text{ and } B \models \alpha_2 \\ B &\models \triangle_i \alpha \iff \alpha \in B_i \end{aligned}$$

The following definition introduces the concept of doxastic alternative.

Definition 2.3 (Doxastic alternatives). Let $B = (B_1, \ldots, B_n, S)$ and $B' = (B'_1, \ldots, B'_n, S')$ be two multi-agent belief bases. Then, $B\mathcal{R}_i B'$ if and only if, for every $\alpha \in B_i$, $B' \models \alpha$, where the satisfaction relation $\models$ follows Definition 2.2.

$B\mathcal{R}_i B'$ means that B' is a doxastic alternative for agent i at B. The idea of the previous definition is that B' is a doxastic alternative for agent i at B if and only if, B' satisfies all facts that agent i explicitly believes at B.

A multi-agent belief model (MAB) is defined to be a multi-agent belief base supplemented with a set of multi-agent belief bases, called *context*. The latter includes all multi-agent belief bases that are compatible with the agents' common ground [38], i.e., the body of information that the agents commonly believe to be the case.

Definition 2.4 (Multi-agent belief model). A multi-agent belief model (MBM) is a pair (B, Cxt), where $B \in \mathbf{B}$ and $Cxt \subseteq \mathbf{B}$. The class of MBMs is denoted by $\mathbf{M}$.

The following definition generalizes the definition of the satisfaction relation $\models$ given in Definition 2.1 to the full language $\mathcal{L}_{\text{LDA}}$. Its formulas are interpreted with respect to MBMs. (Boolean cases are omitted, as they are defined in the usual way.)

Definition 2.5 (Satisfaction relation (cont.)). Let $(B, Cxt) \in \mathbf{M}$. Then:

$$\begin{aligned} (B, Cxt) &\models \alpha \iff B \models \alpha \\ (B, Cxt) &\models \Box_i \varphi \iff \forall B' \in Cxt : \text{if } B\mathcal{R}_i B' \text{ then} \\ &\qquad\qquad (B', Cxt) \models \varphi \end{aligned}$$

Figure 1 illustrates the general idea behind the logic LDA, especially for what concerns the relationship between the agents' belief bases and the agents' common ground (or context) and the relationship between the latter and the agents' implicit beliefs. While an agent's belief base captures the agent's private information, the common ground captures the agents' public information. An agent's implicit belief corresponds to a fact that the agent can deduce from the public information and her private information.

Note that the previous semantics does not guarantee that an agent's belief base is globally consistent with the agents' common ground, as it might be the case that there is no $B' \in Cxt$ such that, for all $\alpha \in B_i$, $B' \models \alpha$. In the latter case, we have that $(B, Cxt) \models \Box_i \bot$ which means that agent i's belief base at (B, Cxt) is globally inconsistent with the agent's common ground.

The global consistency property for multi-agent belief models is defined as follows.

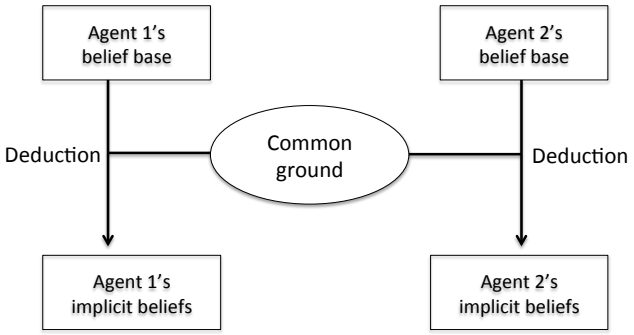

Figure 1: Conceptual framework

Definition 2.6 (Global consistency). Let $(B, Cxt) \in \mathbf{M}$. We say that (B, Cxt) satisfies global consistency (GC) if and only if, for every $B' \in \{B\} \cup Cxt$, there exists $B'' \in Cxt$ such that $B'\mathcal{R}_i B''$.

In some situations, it may be useful to assume that agents' beliefs are correct, i.e., what an agent believes is true. When talking about correct (or true) explicit and implicit beliefs, it is usual to call them explicit and implicit knowledge. Indeed, the terms "true belief", "correct belief" and "knowledge" are usually assumed to be synonyms. The following definition introduces belief correctness for multi-agent belief models.

Definition 2.7 (Belief correctness). Let $(B, Cxt) \in \mathbf{M}$. We say that (B, Cxt) satisfies belief correctness (BC) if and only if $B \in Cxt$ and, for every $B' \in Cxt$, $B'\mathcal{R}_i B'$.

Clearly, belief correctness implies global consistency.

As the following proposition highlights, belief correctness for multi-agent belief models is completely characterized by the fact that the actual world is included in the agents' common ground and that the agents' explicit beliefs cannot be wrong.

Proposition 2.8. *Let $(B, Cxt) \in \mathbf{M}$. Then, (B, Cxt) satisfies BC if and only if $B \in Cxt$ and, for all $i \in Agt$, $B' \in Cxt$ and $\alpha \in B_i$, we have $B' \models \alpha$.*

For every $X \subseteq \{GC, BC\}$, we let $\mathbf{M}_X$ denote the class of models satisfying all properties in X. Clearly, $\mathbf{M}_\emptyset = \mathbf{M}$. Let $\varphi \in \mathcal{L}$, we say that φ is valid for the class $\mathbf{M}_X$, denoted by $\models_X \varphi$, if and only if, for every $(B, Cxt) \in \mathbf{M}_X$, we have $(B, Cxt) \models \varphi$. For notational convenience, we write $\models \varphi$ instead of $\models_\emptyset \varphi$. We say that φ is satisfiable for the class $\mathbf{M}_X$ if and only if $\neg\varphi$ is not valid for the class $\mathbf{M}_X$.

3 TABLEAUX

In this section, we present a tableau method that can be used for satisfiability of formulas of the logic LDA. As we will deal with sets for the rest of this section, if we use a single formula on a set operation, we mean the singleton containing such formula and we alternatively use ";" as the union operator.

Definition 3.1 (Tableau Rule). A tableau rule consists of a set Γ above a line called the numerator, and a list of distinct sets $\Gamma_1, .., \Gamma_n$ separated by |, called the denominators:

$$\frac{\Gamma}{\Gamma_1 \mid \ldots \mid \Gamma_n}$$

The following definition specifies the conditions under which a rule is applicable.

Definition 3.2 (Applicable rule and saturated set). A tableau rule is applicable to a set Γ if Γ is an instance of its numerator and Γ is not an instance of one of its denominators. We say that a set Γ is saturated if there is no rule applicable to it.

The condition requiring that for a tableau rule to be applicable to a set Γ, Γ does not have to be an instance of one of its denominators, guarantees that when constructing a tableau we do not loop indefinitely by applying the same rule infinitely many times.

In the following definition, we introduce the static rules for our tableau method.

Definition 3.3 (Static rules). Let X be a finite set of formulas from $\mathcal{L}_{\mathsf{LDA}}$, then:

$$\bot\text{-rule: } \frac{\psi; \neg\psi; X}{\bot} \qquad \wedge\text{-rule: } \frac{\psi \wedge \varphi; X}{\psi; \varphi; \psi \wedge \varphi; X}$$

$$\neg\text{-rule: } \frac{\neg\neg\psi; X}{\psi; \neg\neg\psi; X} \qquad \triangle_i\text{-rule: } \frac{\triangle_i \alpha; X}{\Box_i \alpha; \triangle_i \alpha; X}$$

$$\vee\text{-rule: } \frac{\neg(\psi \wedge \varphi); X}{\neg\psi; \neg(\psi \wedge \varphi); X \mid \neg\varphi; \neg(\psi \wedge \varphi); X}$$

The following extra rules are used for the KD and KT variants of the logic LDA.

Definition 3.4 (T-rule and D-rule). Let X be a finite set of formulas from $\mathcal{L}_{\mathsf{LDA}}$, then:

$$\text{T-rule: } \frac{\Box_i \psi; X}{\psi; \Box_i \psi; X} \qquad \text{D-rule: } \frac{\Box_i \psi; X}{\Diamond_i \psi; \Box_i \psi; X}$$

The D-rule corresponds to the property of global consistency (GC) on multi-agent belief models, while the T-rule corresponds to the property of belief correctness (BC). This correspondence is captured by the function cf such that:

$$cf(\text{D-rule}) = GC$$
$$cf(\text{T-rule}) = BC$$

The transitional rule allows to generate a new successor for a certain agent i.

Definition 3.5 (Transitional rule). Let X be a finite set of formulas from $\mathcal{L}_{\mathsf{LDA}}$, then:

$$\Diamond_i\text{: } \frac{\Diamond_i \psi; X}{\psi; \{\varphi \mid \Box_i \varphi \in X\}}$$

The following definition introduces the concept of tableau.

Definition 3.6 (Tableau). Let $X \subseteq \{\text{T-rule}, \text{D-rule}\}$. A LDA_X-tableau for Γ is a tree such that each vertex v carries a pair (Γ', ρ), where Γ' is a set of formulas and ρ is one of the following: (i) an instance of a static rule applicable to Γ', (ii) an instance of a rule from X applicable to Γ', (iii) a transitional rule applicable to Γ' or iv) the empty rule nihil. The root of the tableau carries a pair (Γ, ρ) for some tableau rule ρ. Moreover, for every vertex v, if v carries the pair (Γ', ρ), then the following conditions hold:

- if Γ' is not saturated then $\rho \neq$ nihil, and
- if ρ has k denominators $\Gamma_1, \ldots, \Gamma_k$ then v has exactly k children $v_1, \ldots, v_k$ such that, for every $1 \leq h \leq k$, v_h carries (Γ_h, ρ') for some tableau rule ρ'.

The following definition introduces the concept of closed tableau.

Definition 3.7 (Closed tableau). A branch in a tableau is a path from the root of the tableau to an end vertex, where an end vertex is a vertex carrying a pair (Γ', nihil). A branch in a tableau is closed if its end node is of the form $(\{\bot\}, \text{nihil})$. A tableau is closed if all its branches are closed, otherwise it is open.

The following theorem highlights that our tableau method is sound.

Theorem 3.8. *Let $\varphi \in \mathcal{L}_{LDA}$ and let $X \subseteq \{T\text{-rule}, D\text{-rule}\}$. Then, if φ is satisfiable for the class $\mathbf{M}_{\{cf(x):x\in X\}}$ then all LDA_X-tableaux for $\{\varphi\}$ are open.*

Proof. By means of the truth conditions given in Definitions 2.2 and 2.5, it is straightforward to show that every static rule, every transitional rule and every rule in X applied to a $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable set of formulas Γ generates some $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable set of formulas, where a set of formulas Γ is $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable iff there exists $(B, Cxt) \in \mathbf{M}_{\{cf(x):x\in X\}}$ such that $(B, Cxt) \models \varphi$ for all $\varphi \in \Gamma$. More precisely, if Γ is $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable and rule ρ is applicable to Γ, then there exists a possible result Γ' of applying ρ to Γ such that Γ' is $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable.[4]

Thanks to the previous property, by induction on the height of the tableau, it is easy to show that if Γ is $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable and Θ is any LDA_X-tableau for Γ with finite height then Θ is open.

(Base case). Suppose Θ is a LDA_X-tableau for Γ with height 0. Then, (Γ, nihil). Since Γ is $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable, $\Gamma \neq \{\top\}$. Hence, Θ is open.

(Induction case). Suppose Θ is a LDA_X-tableau for Γ with height $k+1$. By the previous property and the fact that Γ is $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable, there exists a vertex v such that v is a child of Θ's root, v is the root of a sub-tableau Θ' of Θ with height equal or less than k and v carries a pair (Γ', ρ), where Γ' is $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable. By induction hypothesis, the sub-tableau Θ' is open. Consequently, the tableau Θ is open too.

Since every LDA_X-tableau for $\{\varphi\}$ has finite height, it follows that if φ is $\mathbf{M}_{\{cf(x):x\in X\}}$-satisfiable then all tableaux for $\{\varphi\}$ are open. □

In the rest of this section, we are going to prove that our tableau-based method is complete. Our proof exploits the proof theory of LDA provided in [28]. In particular, in [28], it is shown that the logic $\mathsf{LDA}_{\{\mathbf{D}_{\Box_i}\}}$, i.e., the KD variant of the logic LDA, defined by all tautologies of propositional calculus together with the following axioms and rules of inference is sound and complete relative to the

[4] Γ' is a possible result of applying some tableau rule to Γ iff there exists a tableau rule ρ such that Γ is its numerator and Γ' is one of its denominators.

class $\mathbf{M}_{\{GC\}}$:

$$(\Box_i\varphi \wedge \Box_i(\varphi \rightarrow \psi)) \rightarrow \Box_i\psi \qquad (\mathbf{K}_{\Box_i})$$

$$\neg(\Box_i\varphi \wedge \Box_i\neg\varphi) \qquad (\mathbf{D}_{\Box_i})$$

$$\triangle_i\alpha \rightarrow \Box_i\alpha \qquad (\mathbf{Int}_{\triangle_i,\Box_i})$$

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi} \qquad (\mathbf{MP})$$

$$\frac{\varphi}{\Box_i\varphi} \qquad (\mathbf{Nec}_{\Box_i})$$

Let

$$\Box_i\varphi \rightarrow \varphi \qquad (\mathbf{T}_{\Box_i})$$

Then, it is easy to generalize the completeness proof given in [28] to show that, for every $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$, the logic LDA_X is sound and complete for the class $\mathbf{M}_{\{g(x):x\in X\}}$ with $g(\mathbf{D}_{\Box_i}) = GC$ and $g(\mathbf{T}_{\Box_i}) = BC$, and where LDA_X is the logic defined by all tautologies of propositional calculus, the previous principles $\mathbf{K}_{\Box_i}$, $\mathbf{Int}_{\triangle_i,\Box_i}$, **MP**, $\mathbf{Nec}_{\Box_i}$, and all principles in X. Furthermore, it is easy to show that alternative sound and complete axiomatizations for these logics consist in replacing Axiom $\mathbf{K}_{\Box_i}$ and the rule of inference $\mathbf{Nec}_{\Box_i}$ by the following rule of inference:

$$\frac{\varphi \rightarrow (\psi_1 \vee \ldots \vee \psi_m)}{\Diamond_i\varphi \rightarrow (\Diamond_i\psi_1 \vee \ldots \vee \Diamond_i\psi_m)} \qquad (\mathbf{Trans}_{\Box_i})$$

Indeed, the rule $\mathbf{Trans}_{\Box_i}$ preserves validity. Moreover, Axiom $\mathbf{K}_{\Box_i}$ and the rule $\mathbf{Nec}_{\Box_i}$ are derivable from it.

Therefore, for every $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$, we have that if $\models_{\{g(x):x\in X\}} \varphi$ then $\vdash_{\mathsf{LDA}_X} \varphi$, where $\vdash_{\mathsf{LDA}_X} \varphi$, means that there is a proof of φ in LDA_X, that is, there is a sequence of formulas $(\varphi_1, \ldots, \varphi_n)$ such that:

- $\varphi_n = \varphi$,
- for every $1 \leq k \leq n$, either φ_k is an instance of one of the axiom schema of LDA_X or there are formulas $\varphi_{k_1}, \ldots, \varphi_{k_m}$ such that $k_1, \ldots, k_m < k$ and $\frac{\varphi_{k_1},\ldots,\varphi_{k_m}}{\varphi_k}$ is an instance of some inference rule of LDA_X.

In what follows, we are going to prove that if $\vdash_{\mathsf{LDA}_X} \varphi$ then there exists a LDA_X-tableau for $\{\neg\varphi\}$ which is closed. Before entering into the proof, let us define the collection of all LDA_X tableau-consistent sets of formulas.

Definition 3.9 (Tableau-consistent sets of formulas). Let $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. We define TC_{LDA_X} to be the largest collection of sets of formulas from the language $\mathcal{L}_{\mathsf{LDA}}$ which satisfies the following conditions. For every $\Gamma \in TC_{\mathsf{LDA}_X}$:

- Γ does not contain $\bot$,
- if $\Gamma \in TC_{\mathsf{LDA}_X}$ and $\varphi, \neg\varphi \in \Gamma$ then $\Gamma \cup \{\bot\} \in TC_{\mathsf{LDA}_X}$,
- if $\Gamma \in TC_{\mathsf{LDA}_X}$ and $\varphi \wedge \psi \in \Gamma$ then $\Gamma \cup \{\varphi, \psi\} \in TC_{\mathsf{LDA}_X}$,
- if $\Gamma \in TC_{\mathsf{LDA}_X}$ and $\neg\neg\varphi \in \Gamma$ then $\Gamma \cup \{\varphi\} \in TC_{\mathsf{LDA}_X}$,
- if $\Gamma \in TC_{\mathsf{LDA}_X}$ and $\varphi \in \Gamma$ then $\{\neg\varphi\} \cup \Gamma \notin TC_{\mathsf{LDA}_X}$,
- if $\Gamma \in TC_{\mathsf{LDA}_X}$ and $\neg(\varphi \wedge \psi) \in \Gamma$ then $\Gamma \cup \{\neg\varphi\} \in TC_{\mathsf{LDA}_X}$ or $\Gamma \cup \{\neg\psi\} \in TC_{\mathsf{LDA}_X}$,
- if $\Gamma \in TC_{\mathsf{LDA}_X}$ and $\triangle_i\alpha \in \Gamma$ then $\Gamma \cup \{\Box_i\alpha\} \in TC_{\mathsf{LDA}_X}$,
- if $\Gamma \in TC_{\mathsf{LDA}_X}$ and $\Diamond_i\varphi \in \Gamma$ then $\{\varphi\} \cup \{\psi : \Box_i\psi \in \Gamma\} \in TC_{\mathsf{LDA}_X}$,
- if $\mathbf{D}_{\Box_i} \in X$, $\Gamma \in TC_{\mathsf{LDA}_X}$ and $\Box_i\varphi \in \Gamma$ then $\Gamma \cup \{\Diamond_i\varphi\} \in TC_{\mathsf{LDA}_X}$, and
- if $\mathbf{T}_{\Box_i} \in X$, $\Gamma \in TC_{\mathsf{LDA}_X}$ and $\Box_i\varphi \in \Gamma$ then $\Gamma \cup \{\varphi\} \in TC_{\mathsf{LDA}_X}$.

As the following proposition highlights, tableau consistency is closed under subset operation.

Proposition 3.10. *Let $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$ and let $\Gamma, \Gamma' \subseteq \mathcal{L}_{LDA}$. Then, if $\Gamma' \subseteq \Gamma$ and $\Gamma \in TC_{LDA_X}$ then $\Gamma' \in TC_{LDA_X}$.*

Proof. The right-to-left direction is trivial. The left-to-right direction is provable by checking that if $\Gamma' \subseteq \Gamma$ and $\Gamma \in TC_{\mathsf{LDA}_X}$ then Γ' satisfies all previous conditions for the elements in TC_{LDA_X}. □

The following lemma is essential for proving our main result.

Lemma 3.11. *Let $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. Then, if $\vdash_{LDA_X} \varphi$ then $\{\neg\varphi\} \notin TC_{LDA_X}$.*

Proof. Suppose $\vdash_{\mathsf{LDA}_X} \varphi$. This means that that there exists a proof of φ in LDA_X. By induction on the length of the proof, we show that $\{\neg\varphi\} \notin TC_{\mathsf{LDA}_X}$.

(Base case). Suppose the proof has length 1. This means that φ is an instance of an axiom of the logic LDA_X. Suppose φ is an instance of Axiom $\mathbf{Int}_{\triangle_i,\Box_i}$ of the form $\neg\triangle_i\alpha \vee \Box_i\alpha$. Moreover, suppose that $\{\triangle_i\alpha \wedge \neg\Box_i\alpha\} \in TC_{\mathsf{LDA}_X}$. By definition of TC_{LDA_X}, we have $\{\triangle_i\alpha, \neg\Box_i\alpha, \triangle_i\alpha \wedge \neg\Box_i\alpha\} \in TC_{\mathsf{LDA}_X}$. Moreover, $\{\triangle_i\alpha, \Box_i\alpha, \neg\Box_i\alpha, \triangle_i\alpha \wedge \neg\Box_i\alpha\} \in TC_{\mathsf{LDA}_X}$ and, consequently, $\{\bot, \triangle_i\alpha, \Box_i\alpha, \neg\Box_i\alpha, \triangle_i\alpha \wedge \neg\Box_i\alpha\} \in TC_{\mathsf{LDA}_X}$. The latter contradicts the fact that every element of TC_{LDA_X} does not contain $\bot$. We can also prove that if φ is an instance of a tautology of propositional calculus. Moreover, in a similar way, we can prove that if φ an instance of Axiom $\mathbf{T}_{\Box_i}$ (resp. $\mathbf{D}_{\Box_i}$) and $\mathbf{T}_{\Box_i} \in X$ (resp. $\mathbf{D}_{\Box_i} \in X$) then $\{\neg\varphi\} \notin TC_{\mathsf{LDA}_X}$.

(Induction case). Suppose the proof has length $k+1$. Thus, we have $\varphi = \varphi_{k+1}$, where φ_{k+1} is the last element of the proof $(\varphi_1, \ldots, \varphi_{k+1})$. Moreover, either φ_{k+1} is an instance of one of the axiom schema of LDA_X, or there are formulas $\varphi_{k_1}, \ldots, \varphi_{k_m}$ such that $k_1, \ldots, k_m < k+1$ and $\frac{\varphi_{k_1},\ldots,\varphi_{k_m}}{\varphi_{k+1}}$ is an instance of the inference rule $\mathbf{Trans}_{\Box_i}$. The first case is treated as in the base case of the main proof. Let us prove the second case after assuming that $\{\neg\varphi_{k+1}\} \in TC_{\mathsf{LDA}_X}$. We have that φ_{k+1} is of the form $\Diamond_i\psi \rightarrow (\Diamond_i\psi_1 \vee \ldots \vee \Diamond_i\psi_m)$ since φ_{k+1} is the result of the application of the inference rule $\mathbf{Trans}_{\Box_i}$ on φ_{k+1}. Thus, $\neg\varphi_{k+1}$ is $\Diamond_i\psi \wedge (\Box_i\neg\psi_1 \wedge \ldots \wedge \Box_i\neg\psi_m)$. Hence, by definition of TC_{LDA_X}, $\{\Diamond_i\psi \wedge \Box_i\neg\psi_1 \wedge \ldots \wedge \Box_i\neg\psi_m, \Diamond_i\psi, \Box_i\neg\psi_1, \ldots, \Box_i\neg\psi_m\} \in TC_{\mathsf{LDA}_X}$. Thus, by the definition of TC_{LDA_X}, $\{\psi, \neg\psi_1, \ldots, \neg\psi_m\} \in TC_{\mathsf{LDA}_X}$ and, consequently, $\{\psi, \neg\psi_1, \ldots, \neg\psi_m, \psi \wedge \neg\psi_1 \wedge \ldots \wedge \neg\psi_m\} \in TC_{\mathsf{LDA}_X}$. By Proposition 3.10, the latter implies $\{\psi \wedge \neg\psi_1 \wedge \ldots \wedge \neg\psi_m\} \in TC_{\mathsf{LDA}_X}$. Notice that $\varphi_k = \psi \wedge \neg\psi_1 \wedge \ldots \wedge \neg\psi_m$. Hence, by induction hypothesis, $\{\psi \wedge \neg\psi_1 \wedge \ldots \wedge \neg\psi_m\} \notin TC_{\mathsf{LDA}_X}$ which leads to a contradiction. □

It is easy to adapt the proof of [14, Theorem 6.1] to check that, for every $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$, if $\{\neg\varphi\} \notin TC_{\mathsf{LDA}_X}$ then there exists a LDA_X tableau for $\{\neg\varphi\}$ which is closed. Thus, by Lemma 3.11 and the completeness of each logic LDA_X with respect to its corresponding class of models $\mathbf{M}_{\{g(x):x\in X\}}$, we can state the following theorem.

Theorem 3.12. *Let $\varphi \in \mathcal{L}_{LDA}$ and let $X \subseteq \{$T-rule, D-rule$\}$. Then, if all LDA$_X$-tableaux for $\{\varphi\}$ are open then φ is satisfiable for the class $\mathbf{M}_{\{cf(x):x\in X\}}$.*

The previous Theorems 3.8 and 3.12 allow us to determine whether a formula φ is satisfiable for the class $\mathbf{M}_{\{cf(x):x\in X\}}$ by checking whether all LDA$_X$-tableaux for $\{\varphi\}$ are open. Now, following [16], we can easily construct a satisfability checking procedure running in PSPACE based on this result.

To search for the existence of a closed LDA$_X$-tableau for $\{\varphi\}$, we use a backtracking procedure that will construct a closed LDA$_X$-tableau for $\{\varphi\}$ if there is any.

The length of a branch of a LDA$_X$-tableau whose root vertex carryies a pair $(\{\neg\varphi\}, \rho)$, for some arbitrary rule ρ, is polynomial in the size of φ. Since the backtracking procedure only needs to keep in memory a single branch, the procedure runs in PSPACE.

We conclude this section by the following complexity result.

Theorem 3.13. *Let $X \subseteq \{GC, BC\}$. Then, checking satisfiability of formulas in $\mathcal{L}_{LDA}$ relative to the class $\mathbf{M}_X$ is a PSPACE-complete problem.*

Proof. PSPACE-membership is guaranteed to hold by the previous tableau-based satisfiability checking procedure. PSPACE-hardness follows from known PSPACE-hardness results for modal logics K, KT and KD [16]. □

4 DYNAMIC EXTENSION

In this section, we provide a dynamic extension of the logic LDA by a simple form of action of type 'assignment' which consists in setting to true or to false the truth value of a given propositional variable. We call this extension DLDA, which stands for Dynamic LDA. The notion of action viewed as a propositional assignment has been explored by [3, 21, 39, 40]. It is compatible with the idea of viewing an action as bringing about or effecting (at will) of a change [43].

In DLDA assignments are performed by agents. An assignment changes not only the environment but also the agents' beliefs depending on whether and how they jointly perceive its execution. We call *perceptive context* a description of what the agents can see in a given situation of interaction. We define it formally as a finite set of formulas from the following language $\mathcal{L}_{\text{OBS}}$:

$$\omega \quad ::= \quad \mathsf{s}_{i,j} \mid \mathsf{s}_i\omega$$

where i and j range over Agt. The expression $\mathsf{s}_{i,j}$ describes the fact that "agent i observes agent j" or, more precisely, "agent i sees what agent j does". The more complex expressions of the form $\mathsf{s}_i\omega$ represents the fact that "agent i sees that ω". For example, $\mathsf{s}_i\mathsf{s}_{z,j}$ represents the fact that "agent i sees that agent z sees what agent j does".

Definition 4.1 (Perception precondition). Let $\Delta = \bigcup_{j\in Agt}\{\triangle_j\}$, let $Seq(\Delta)$ be the set of all sequences of elements of Δ and let $i, j \in Agt$. Elements of $Seq(\Delta)$ are denoted by $\sigma, \sigma', \ldots$ We define

$$F_{j,i}(\sigma) = \bigcup_{\sigma'\sqsubseteq\sigma} \{f_{j,i}(\sigma')\}$$

where $f_{j,i} : Seq(\Delta) \longrightarrow \mathcal{L}_{\text{OBS}}$ such that:

$$f_{j,i}(\curlywedge) = \mathsf{s}_{j,i}$$

and for all $z \in Agt$:

$$f_{j,i}(\triangle_z\sigma) = \mathsf{s}_j f_{z,i}(\sigma)$$

with $\curlywedge$ the empty sequence and $\sigma' \sqsubseteq \sigma$ meaning that σ' is a subsequence of σ.

$F_{j,i}(\sigma)$ identifies the perception precondition for agent j to acquire σ-type information from agent i's performing an action, where $\curlywedge$-type information is information about the state of the world and $\triangle_z\sigma$-type information is information about agent z's acquiring σ-type information. For example, if $\sigma = \triangle_z$ then $F_{j,i}(\sigma) = \{\mathsf{s}_{j,i}, \mathsf{s}_j\mathsf{s}_{z,i}\}$. This means that, for agent j to acquire information about z's acquiring information from agent i performing an action, it has to be the case that j sees what i does and j sees that z sees what i does.

The language of the logic DLDA, denoted by $\mathcal{L}_{\text{DLDA}}$, is defined by the following grammar:

$$\varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i\varphi \mid [(i,\tau p,\Omega)]\varphi$$

where p ranges over Atm, i ranges over Agt, α ranges over the language $\mathcal{L}_0$ defined in Section 2, τ ranges over $\{+,-\}$ and Ω is a finite set of formulas from the previous language $\mathcal{L}_{\text{OBS}}$.

The action $+p$ consists in setting the value of the atomic variable p to true, whereas the action $-p$ consists in setting the value of the atomic variable p to false. The formula $[(i,\tau p,\Omega)]\varphi$ has to be read "φ holds after the performance of the action τp by agent i in the perceptive context Ω". For example, suppose $\Omega = \{\mathsf{s}_{i,i}, \mathsf{s}_{j,i}, \mathsf{s}_i\mathsf{s}_{j,i}\}$. Then, $[(i, +p,\Omega)]\varphi$ means that "φ holds after agent i has set the value of p to true, in the situation in which agent i sees what she does, j sees what i does and i sees that j sees what i does".

The following definition extends the satisfaction relation to the dynamic operator $[(i,\tau p,\Omega)]$.

Definition 4.2 (Satisfaction relation (cont.)). Let $B = (B_1, \ldots, B_n, S) \in \mathbf{B}$ and let $(B, Cxt) \in \mathbf{M}$. Then:

$$(B, Cxt) \models [(i,\tau p,\Omega)]\varphi \iff (B^{(i,\tau p,\Omega)}, Cxt) \models \varphi$$

with $B^{(i,\tau p,\Omega)} = (B_1^{(i,\tau p,\Omega)}, \ldots, B_n^{(i,\tau p,\Omega)}, S^{(i,\tau p,\Omega)})$, where:

$$S^{(i,\tau p,\Omega)} = S \cup \{p\} \qquad \text{if } \tau = +$$
$$S^{(i,\tau p,\Omega)} = S \setminus \{p\} \qquad \text{if } \tau = -$$

and for all $j \in Agt$:

$$B_j^{(i,\tau p,\Omega)} = \big(B_j \cup \{\sigma p : \sigma \in Seq(\Delta) \text{ and } F_{j,i}(\sigma) \subseteq \Omega\}\big) \qquad \text{if } \tau = +$$
$$B_j^{(i,\tau p,\Omega)} = \big(B_j \cup \{\sigma\neg p : \sigma \in Seq(\Delta) \text{ and } F_{j,i}(\sigma) \subseteq \Omega\}\big) \qquad \text{if } \tau = -$$

where $F_{j,i}(\sigma)$ is defined as in Definition 4.1.

According to the previous definition, the effects of the action $+p$ (resp. $-p$) performed by i in the perceptive context Ω consist in setting the truth value of p to true (resp. false) and in modifying the belief bases of the agents who perceived the action occurrence, according to the perceptive context Ω. The idea is that the performance of action τp by agent i as well as its effects are perceived by agent j if and only if i does τp and j sees what i does (i.e., $\{\mathsf{s}_{j,i}\} \subseteq \Omega$). Agent j believes that the performance of the action τp by i as well as its effects are perceived by agent z_1 if and only if i does τp, j sees what i does, and j sees that z_1 sees what i does (i.e., $\{\mathsf{s}_{j,i}, \mathsf{s}_j\mathsf{s}_{z_1,i}\} \subseteq \Omega$). Agent j believes that agent z_1 believes that the

performance of the action τp by i as well as its effects are perceived by agent z_2 if and only if i does τp, j sees what i does, j sees that z_1 sees what i does, and j sees that z_1 sees that z_2 sees what i does (i.e., $\{s_{j,i}, s_j s_{z_1,i}, s_j s_{z_1} s_{z_2,i}\} \subseteq \Omega$), and so on so forth. Note that the execution of an action by an agent does not change the agents' common ground. Dynamics of the agents' common ground under public announcements in the sense of [34] are not investigated in this paper.

The following validities capture some interesting properties of action observation. For all $i, j, z \in Agt$ and $p \in Atm$, we have:

$$\models \neg s_{j,i} \rightarrow (\triangle_j \alpha \leftrightarrow [(i,\tau p,\Omega)]\triangle_j \alpha) \quad (1)$$

$$\models [(i,+p,\Omega)]\triangle_j p \quad \text{if } \{s_{j,i}\} \subseteq \Omega \quad (2)$$

$$\models [(i,-p,\Omega)]\triangle_j \neg p \quad \text{if } \{s_{j,i}\} \subseteq \Omega \quad (3)$$

$$\models [(i,+p,\Omega)](\triangle_j p \wedge \triangle_j \triangle_z p) \quad \text{if } \{s_{j,i}, s_j s_{z,i}\} \subseteq \Omega \quad (4)$$

$$\models [(i,-p,\Omega)](\triangle_j \neg p \wedge \triangle_j \triangle_z \neg p) \quad \text{if } \{s_{j,i}, s_j s_{z,i}\} \subseteq \Omega \quad (5)$$

$$\models [(i,+p,\Omega)]\triangle_j \sigma p \quad \text{if } F_{j,i}(\sigma) \subseteq \Omega \quad (6)$$

$$\models [(i,-p,\Omega)]\triangle_j \sigma \neg p \quad \text{if } F_{j,i}(\sigma) \subseteq \Omega \quad (7)$$

According to the first validity, if agent j does not observe agent i, then her explicit beliefs are not affected by what i does. According to the second and third validities, if j observes i and i performs action $+p$ (resp. $-p$), then agent j will explicitly believe that p (resp. $\neg p$) afterwards. According to the fourth and fifth validities, if j observes i and j explicitly believes that z observes i, then after i has performed action $+p$ (resp. $-p$), j will explicitly believe that p (resp. $\neg p$). Moreover, j will explicitly believe that z explicitly believes that p (resp. $\neg p$). The sixth and seventh validities generalize the other validities to explicit beliefs of any depth.

4.1 Reduction axioms and complexity

The following proposition provides reduction principles for the dynamic operators $[(i,\tau p,\Omega)]\varphi$.

Proposition 4.3. *The following formulas are valid for the class* **M***:*

$$[(i,\tau p,\Omega)]q \leftrightarrow \top \text{ if } p = q \text{ and } \tau = +$$
$$[(i,\tau p,\Omega)]q \leftrightarrow \bot \text{ if } p = q \text{ and } \tau = -$$
$$[(i,\tau p,\Omega)]q \leftrightarrow q \text{ if } p \neq q$$
$$[(i,\tau p,\Omega)]\triangle_j \alpha \leftrightarrow \triangle_j \alpha \text{ if } \alpha \notin \{\sigma p : \sigma \in Seq(\Delta)\}$$
$$[(i,\tau p,\Omega)]\triangle_j \sigma p \leftrightarrow \top \text{ if } F_{j,i}(\sigma) \subseteq \Omega$$
$$[(i,\tau p,\Omega)]\triangle_j \sigma p \leftrightarrow \triangle_j \sigma p \text{ if } F_{j,i}(\sigma) \nsubseteq \Omega$$
$$[(i,\tau p,\Omega)]\neg\varphi \leftrightarrow \neg[(i,\tau p,\Omega)]\varphi$$
$$[(i,\tau p,\Omega)](\varphi \wedge \psi) \leftrightarrow ([(i,\tau p,\Omega)]\varphi \wedge [(i,\tau p,\Omega)]\psi)$$
$$[(i,\tau p,\Omega)]\Box_j \varphi \leftrightarrow \Box_j(\bigwedge_{\sigma p : F_{j,i}(\sigma) \subseteq \Omega} \sigma p \rightarrow \varphi)$$

The equivalences of Proposition 4.3 allow to find for every formula of the language $\mathcal{L}_{\text{DLDA}}$ an equivalent formula of the language $\mathcal{L}_{\text{LDA}}$. Call *red* the mapping which iteratively applies the equivalences of Proposition 4.3 from the left to the right, starting from one of the innermost modal operators. *red* pushes the dynamic operators $[(i,\tau p,\Omega)]$ inside the formula, and finally eliminates them when facing an atomic formula. Specifically, the mapping *red* is inductively defined by:

1. $red(p) = p$
2. $red(\triangle_j \alpha) = \triangle_j red(\alpha)$
3. $red(\neg\varphi) = \neg red(\varphi)$
4. $red(\varphi \wedge \psi) = red(\varphi) \wedge red(\psi)$
5. $red(\Box_j \varphi) = \Box_j red(\varphi)$
6. $red([(i,\tau p,\Omega)]q) = red(q)$ if $p = q$ and $\tau = +$
7. $red([(i,\tau p,\Omega)]q) = red(\neg q)$ if $p = q$ and $\tau = -$
8. $red([(i,\tau p,\Omega)]q) = red(q)$ if $p \neq q$
9. $red([(i,\tau p,\Omega)]\triangle_j \alpha) = red(\triangle_j \alpha)$ if $\alpha \notin \{\sigma p : \sigma \in Seq(\Delta)\}$
10. $red([(i,\tau p,\Omega)]\triangle_j \sigma p) = \top$ if $F_{j,i}(\sigma) \subseteq \Omega$
11. $red([(i,\tau p,\Omega)]\triangle_j \sigma p) = red(\triangle_j \sigma p)$ if $F_{j,i}(\sigma) \nsubseteq \Omega$
12. $red([(i,\tau p,\Omega)]\neg\varphi) = red(\neg[(i,\tau p,\Omega)]\varphi)$
13. $red([(i,\tau p,\Omega)](\varphi \wedge \psi)) = red([(i,\tau p,\Omega)]\varphi \wedge [(i,\tau p,\Omega)]\psi)$
14. $red([(i,\tau p,\Omega)]\Box_j \varphi) = red(\Box_j(\bigwedge_{\sigma p : F_{j,i}(\sigma) \subseteq \Omega} \sigma p \rightarrow \varphi))$

We can state the following proposition.

Proposition 4.4. *Let $\varphi \in \mathcal{L}_{DLDA}$. Then, $\varphi \leftrightarrow red(\varphi)$ is valid relative to the class* **M**.

The fact that complexity of DLDA satisfiability checking is in PSPACE follows straightforwardly from the upper bound of complexity for LDA satisfiability checking. Indeed, *red* provides an effective procedure for reducing a formula φ in $\mathcal{L}_{\text{DLDA}}$ into an equivalent formula $red(\varphi)$ in $\mathcal{L}_{\text{LDA}}$ whose size is polynomial in the size of φ. Therefore, in order to verify whether φ is satisfiable for the class **M**, one just needs to check whether $red(\varphi)$ is satisfiable for the class **M** by using the PSPACE tableau-based satisfiability checking procedure for LDA given in Section 3.

As a consequence, we can state the following complexity result.

Theorem 4.5. *Checking satisfiability of formulas in $\mathcal{L}_{DLDA}$ relative to the class* **M** *is a PSPACE-complete problem.*

4.2 Example

In this section, we use the logic DLDA to formalize a simple scenario of human-robot interaction in a dynamic domain inspired the famous Sally-Anne false belief's task from the psychological literature on Theory of Mind [5].

We assume that $Agt = \{h, r\}$ where h denotes the human and r denotes the robot. The scenario is depicted in Figure 2. The human and the robot are standing in front of each other on the opposite sides of a table. The robot has two boxes and two balls in front of him: box 1, box 2, a black ball and a grey ball. In the initial situation the black ball is inside box 1 and the grey ball is inside box 2. The human can perfectly observe her actions as well as the robot's actions. Similarly, the robot can perfectly observe its actions as well as the human's actions. Moreover, the robot can see that the human can see its actions and the human can see that the robot can see her actions. Therefore, the perceptive context is described by the

following set of formula from the language $\mathcal{L}_{\text{OBS}}$:

$$\Omega_1 = \{s_{r,r}, s_{h,h}, s_{r,h}, s_{h,r}, s_r s_{h,r}, s_h s_{r,h}\}.$$

Let the atomic proposition *blackIn1* denote the fact that the black ball is inside box 1 and let *blackIn2* denote the fact that the black ball is inside box 2. Similarly, let *greyIn1* and *greyIn2* denote, respectively, the fact that the grey ball is inside box 1 and the fact that the grey ball is inside box 2.

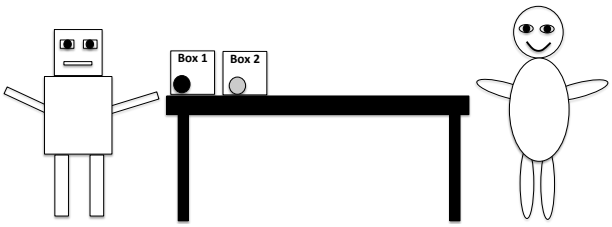

Figure 2: Balls in the boxes scenario

We assume that in the initial situation the human does not explicitly believe that the black ball is inside box 1 and the human does not explicitly believe that the black ball is inside box 2, as she cannot see the box's content. Similarly, the human does not explicitly believe that the grey ball is inside box 1 and the human does not explicitly believe that the grey ball is inside box 2. We also assume that the robot does not explicitly believe that the human explicitly believes that the black ball is inside box 1 (resp. box 2) and that the robot does not explicitly believe that the human explicitly believes that the grey ball is inside box 1 (resp. box 2):

$$\begin{aligned} Hyp_1 \stackrel{\text{def}}{=} & \neg\triangle_h blackIn1 \wedge \neg\triangle_h blackIn2 \wedge \neg\triangle_h greyIn1 \wedge \\ & \neg\triangle_h greyIn2 \wedge \neg\triangle_r\triangle_h blackIn1 \wedge \neg\triangle_r\triangle_h blackIn2 \wedge \\ & \neg\triangle_r\triangle_h greyIn1 \wedge \neg\triangle_r\triangle_h greyIn2 \end{aligned}$$

Moreover, we assume that the robot explicitly believes that if the human explicitly believes that one ball is inside one box then she explicitly believes that the ball cannot be inside the other box:

$$\begin{aligned} Hyp_2 \stackrel{\text{def}}{=} & \triangle_r\big((\triangle_h blackIn1 \rightarrow \triangle_h\neg blackIn2) \wedge \\ & (\triangle_h blackIn2 \rightarrow \triangle_h\neg blackIn1) \wedge \\ & (\triangle_h greyIn1 \rightarrow \triangle_h\neg greyIn2) \wedge \\ & (\triangle_h greyIn2 \rightarrow \triangle_h\neg greyIn1)\big) \end{aligned}$$

We can use the logic DLDA to infer that, in the perceptive context Ω_1, if the robot moves the black ball from box 1 to box 2 then, after the occurrence of the action, both the human and the robot will explicitly believe that the black ball is inside box 2, the robot will explicitly believe that the human explicitly believes that the black ball is inside box 2, and the robot will implicitly believe that the human explicitly believes that the black ball is outside box 1:

$$\begin{aligned} (Hyp1 \wedge Hyp2) \rightarrow [(r, +\, blackIn2, \Omega_1)](\triangle_r blackIn2 \wedge \\ \triangle_h blackIn2 \wedge \\ \triangle_r\triangle_h blackIn2 \wedge \\ \Box_r\triangle_h\neg blackIn1) \end{aligned}$$

Now, suppose the human moves away so that she cannot see anymore what the robot does and the robot knows this. In other words, let us suppose that situation has changed into the following perceptive context Ω_2 in which the robot and the human can see their own actions but cannot see the actions of the other:

$$\Omega_2 = \{s_{r,r}, s_{h,h}\}.$$

In the new perceptive context Ω_2, if the robot moves the grey ball from box 2 to box 1 then, after the occurrence of the robot's action, the human will continue to believe that the black ball is inside box 2, without believing that the grey ball is inside box 1. Moreover, the robot still does not believe that the human believes that the grey ball is inside box 1:

$$\begin{aligned} (Hyp1 \wedge Hyp2) \rightarrow & [(r, +\, blackIn2, \Omega_1)] \\ & [(r, +\, greyIn1, \Omega_2)](\triangle_h blackIn2 \wedge \\ & \neg\triangle_h greyIn1 \wedge \neg\triangle_r\triangle_h greyIn1) \end{aligned}$$

5 CONCLUSION

We have presented a family of multi-agent epistemic logics whose semantics exploit the concept of belief base and provided PSPACE tableau-based satisfiability checking procedures for them. Our logics distinguish the concept of explicit belief from the concept of implicit belief. We have introduced a dynamic extension based on propositional assignments whose executions can be more or less visible by the agents. The proposed extension allows us to model higher-order forms of perception, e.g., the fact that a first agent j perceives that a second agent z perceives that a third agent i performs a certain action, thereby coming to believe that z believes that a certain result has been obtained by i's action. We have instantiated the logic in a concrete scenario of human-robot interaction and illustrated its expressive power in capturing subtle aspects of Theory of Mind, with special emphasis on reasoning about beliefs of others on the basis of what the others perceive.

As we have shown in Section 4, if agent i executes a propositional assignment $+p$ (resp. $-p$) and agent j sees what i does, then agent j will expand her belief base by adding p (resp. $\neg p$) to it. This belief base expansion operation may make agent j's belief base globally inconsistent. This explains why we only considered the satisfiability problem of formulas in $\mathcal{L}_{\text{DLDA}}$ relative to the generic class $\mathbf{M}$ and not to the classes $\mathbf{M}_{GC}$ or $\mathbf{M}_{BC}$. In future research, we plan to study a richer variety of dynamic extensions of the logic LDA by different types of consistency-preserving belief base change operations in the style of [19]. Another direction we intend to explore is the connection between the logic LDA and machine learning. Specifically, we plan to combine the logic LDA with machine learning methods, such as inductive logic programming, in order to acquire information to be added to an agent's belief base through experience. The interesting and novel aspect of our approach is that an agent's belief base may contain not only propositional facts but also a theory of the other agents' minds and, in particular, information about the other agents' explicit beliefs. Learning a theory of mind is a fascinating issue that, we believe, can be adequately modeled in the context of our semantics for epistemic logic exploiting belief bases.

ACKNOWLEDGMENTS

This work was supported by the PEPS-CNRS project RoToM "Robots with a Theory of Mind: from logical formalization to implementation".

REFERENCES

[1] N. Alechina, M. Dastani, and B. Logan. 2016. Verifying existence of resource-bounded coalition uniform strategies. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI 2016)*. AAAI Press, 24–30.

[2] G. Aucher and F. Schwarzentruber. 2013. On the complexity of dynamic epistemic logic. In *Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2013)*. 19–28.

[3] P. Balbiani, A. Herzig, and N. Troquard. 2013. Dynamic Logic of Propositional Assignments: A Well-Behaved Variant of PDL. In *Proceedings of the 28th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2013)*. IEEE Computer Society, 143–152.

[4] A. Baltag, L. Moss, and S. Solecki. 1998. The Logic of Public Announcements, Common Knowledge and Private Suspicions. In *Proceedings of the Seventh International Conference on Theoretical Aspects of Rationality and Knowledge (TARK'98)*, Itzhak Gilboa (Ed.). Morgan Kaufmann, San Francisco, CA, 43–56.

[5] S. Baron-Cohen, A. M. Leslie, and U. Frith. 1985. Does the autistic child have a "theory of mind"? *Cognition* 21, 1 (1985), 37–46.

[6] S. Benferhat, D. Dubois, H. Prade, and M.-A. Williams. 2002. A practical approach to revising prioritized knowledge bases. *Studia Logica* 70, 1 (2002), 105–130.

[7] T. Bolander. 2014. Seeing is believing: Formalising false-belief tasks in dynamic epistemic logic. In *Proceedings of the European conference on Social Intelligence (ECSI-2014)*, A. Herzig and E. Lorini (Eds.). 87–107.

[8] T. Bolander, H. van Ditmarsch, A. Herzig, E. Lorini, P. Pardo, and F. Schwarzentruber. 2015. Announcements to Attentive Agents. *Journal of Logic, Language and Information* 25, 1 (2015), 1–35.

[9] T. Bosse, Z. Memon, and J. Treur. 2011. A recursive BDI-agent model for theory of mind and its applications. *Applied Artificial Intelligence* 25, 1 (2011), 1–44.

[10] C. Breazeal, J. Gray, and M. Berlin. 2009. An Embodied Cognition Approach to Mindreading Skills for Socially Intelligent Robots. *The International Journal of Robotics Research* 20, 4 (2009), 656–680.

[11] T. Charrier, A. Herzig, E. Lorini, F. Maffre, and F. Schwarzentruber. 2016. Building Epistemic Logic from Observations and Public Announcements. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning (KR 2016)*. AAAI Press, 268–277.

[12] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. 1995. *Reasoning about Knowledge*. MIT Press, Cambridge.

[13] R. Fagin and J. Y. Halpern. 1987. Belief, Awareness, and Limited Reasoning. *Artificial Intelligence* 34(1) (1987), 39–76.

[14] M. Fitting. 1983. *Proof Methods for Modal and Intuitionistic Logics*. D. Reidel, Dordrecht.

[15] J. Gerbrandy and W. Groeneveld. 1997. Reasoning about information change. *Journal of Logic, Language, and Information* 6 (1997), 147–196.

[16] J. Y. Halpern and Y. Moses. 1992. A Guide to Completeness and Complexity for Modal Logics of Knowledge and Belief. *Artificial Intelligence* 54, 2 (1992), 319–379.

[17] S. O. Hansson. 1991. *Belief Base Dynamics*. Ph.D. Dissertation. Uppsala University, Sweden.

[18] S. O. Hansson. 1993. Theory contraction and base contraction unified. *Journal of Symbolic Logic* 58, 2 (1993), 602–625.

[19] S. O. Hansson. 1999. *A Textbook of Belief Dynamics: Theory Change and Database Updating*. Kluwer, Dordrecht, Netherland.

[20] M. Harbers, K. van den Bosch, and J.-J.Ch. Meyer. 2012. Modeling agents with a theory of mind : Theory-theory versus simulation theory. *Web Intelligence and Agent Systems* 10, 3 (2012), 331–343.

[21] A. Herzig, E. Lorini, F. Moisan, and N. Troquard. 2011. A Dynamic Logic of Normative Systems. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*. AAAI Press, 228–233.

[22] J. Hintikka. 1962. *Knowledge and Belief*. Cornell University Press, New York.

[23] S. Konieczny and R. Pino Pérez. 2002. Merging information under constraints: a logical framework. *Journal of Logic and Computation* 12, 5 (2002), 773–808.

[24] A. Kratzer. 1981. The Notional Category of Modality. In *Words, Worlds, and Contexts*, H.-J. Eikmeyer and H. Rieser (Eds.). de Gruyter, Berlin / New York.

[25] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami. 2017. Artificial cognition for social human-robot interaction: an implementation. *Artificial Intelligence* 247 (2017), 45–69.

[26] H. J. Levesque. 1984. A logic of implicit and explicit belief. In *Proceedings of the Fourth AAAI Conference on Artificial Intelligence (AAAI'84)*. AAAI Press, 198–202.

[27] A. Lomuscio, H. Qu, and F. Raimondi. 2015. MCMAS: an open-source model checker for the verification of multi-agent systems. *International Journal on Software Tools for Technology Transfer* 19 (2015), 1–22. Issue 1.

[28] E. Lorini. 2018. In Praise of Belief Bases: Doing Epistemic Logic Without Possible Worlds. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. AAAI Press, 1915–1922. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16867

[29] E. Lorini. 2018. In Praise of Belief Bases: Doing Epistemic Logic Without Possible Worlds (Extended Version). In *13th International Conference on Logic and the Foundations of Game and Decision Theory (LOFT 2018), Milan, Italy, July 16-18, 2018*.

[30] C. Lutz. 2006. Complexity and succinctness of public announcement logic. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2006)*. ACM, 137–143.

[31] D. Makinson and L. van der Torre. 2000. Input/output logics. *Journal of Philosophical Logic* 29 (2000), 383–408.

[32] G. Milliez, M. Warnier, A. Clodic, and R. Alami. 2014. A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE Press, 1103–1109.

[33] C. Peters. 2005. Foundations of an agent theory of mind model for conversation initiation in virtual environments. In *Proceedings of the AISB 2005 Symposium on Virtual Social Agents*. 163–170.

[34] J. A. Plaza. 1989. Logics of public communications. In *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, M. Emrich, M. Pfeifer, M. Hadzikadic, and Z. Ras (Eds.). 201–216.

[35] D. V. Pynadath, N. Wang, and S. C. Marsella. 2013. Are You Thinking What I'm Thinking? An Evaluation of a Simplified Theory of Mind. In *Proceedings of the 13th International Conference on Intelligent Virtual Agents (IVA 2013) (LNCS)*, Vol. 8108. Springer, 44–57.

[36] B. Scassellati. 2002. Theory of mind for a humanoid robot. *Autonomous Robots* 12 (2002), 13–24.

[37] Y. Shoham. 2009. Logical Theories of Intention and the Database Perspective. *Journal of Philosophical Logic* 38, 6 (2009), 633–648.

[38] R. Stalnaker. 2002. Common ground. *Linguistics and Philosophy* 25(5-6) (2002), 701–721.

[39] M. L. Tiomkin and J. A. Makowsky. 1985. Propositional dynamic logic with local assignments. *Theoretical Computer Science* 36 (1985), 71–87.

[40] J. van Benthem, J. van Eijck, and B. Kooi. 2006. Logics of communication and change. *Information and Computation* 204, 11 (2006), 1620–1662.

[41] W. van der Hoek, P. Iliev, and M. Wooldridge. 2012. A logic of revelation and concealment. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*. IFAAMAS, 1115–1122.

[42] H. P. van Ditmarsch, W. van der Hoek, and B. Kooi. 2007. *Dynamic Epistemic Logic*. Kluwer Academic Publishers.

[43] G. H. Von Wright. 1963. *Norm and Action*. Routledge and Kegan, London.

[44] A. F. T. Winfield. 2018. Experiments in Artificial Theory of Mind: From Safety to Story-Telling. *Frontiers in Robotics and AI* 5, 75 (2018).

[45] G.-Z. Yang, J. Bellingham, P. E. Dupont, P. Fischer, L. Floridi, R. Full, N. Jacobstein, V. Kumar, M. McNutt, R. Merrifield, B. J. Nelson, B. Scassellati, M. Taddeo, R. Taylor, M. Veloso, Z. L. Wang, and R. Wood. 2018. The grand challenges of Science Robotics. *Science Robotics* 3, 14 (2018).