Rethinking epistemic logic with belief bases

Emiliano Lorini

Rethinking Epistemic Logic with Belief Bases

Emiliano Lorini

IRIT-CNRS, Toulouse University, France

Abstract

We introduce a new semantics for a family of logics of explicit and implicit belief based on the concept of multi-agent belief base. Differently from standard semantics for epistemic logic in which the notions of possible world and doxastic/epistemic alternative are primitive, in our semantics they are non-primitive but are computed from the concept of belief base. We provide complete axiomatizations and prove decidability for our logics via finite model arguments. Furthermore, we provide polynomial embeddings of our logics into Fagin & Halpern's logic of general awareness and establish complexity results via the embeddings. We also present variants of the logics incorporating different forms of epistemic introspection for explicit and/or implicit belief and provide complexity results for some of these variants. Finally, we present a number of dynamic extensions of the static framework by informative actions of both public and private type, including public announcement, belief base expansion and forgetting. We illustrate the application potential of the logical framework with the aid of a concrete example taken from the domain of conversational agents.

1. Introduction

An important distinction in formal epistemology is between *explicit belief* and *implicit belief*. According to Levesque [54], "...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 particularly relevant for the design of resource-bounded agents who spend time to make inferences and do not believe all facts that are deducible from their actual beliefs. It is also acknowledged by Fagin & Halpern (F&H)'s logic of general awareness [29] which defines explicit belief as a formula implicitly believed by an agent and of which the agent is aware.

The concept of explicit belief is tightly connected with the concept of *belief base* [66, 59, 38, 70]. In particular, an agent's belief base, which is not necessarily closed under deduction, includes all facts that are explicitly believed by the agent.

In this article, we present a multi-agent logic, called Logic of Doxastic Attitudes (LDA), that precisely articulates the distinction between explicit belief, as a fact in

Email address: Emiliano.Lorini@irit.fr (Emiliano Lorini)

an agent's belief base, and implicit belief, as a fact that is deducible from the agent's explicit beliefs, given the agents' common ground. The concept of *common ground* [73] corresponds to the body of information that the agents share and that they use to make inferences from their explicit beliefs. The multi-agent aspect of the LDA framework lies in the fact that it supports reasoning about agents' higher-order beliefs, i.e., an agent's explicit (or implicit) belief about the explicit (or implicit) belief of another 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. Specifically, it is assumed that at a given state s agent i considers state s' possible if and only if s' satisfies all formulas that are included in agent i's belief base at s.[1]

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 [45] on the logics of knowledge and belief, it was extended to the multi-agent setting at the end of 80ies [30, 61] and then furtherly developed during the last 20 years, the period of the "dynamic turn", with growing research on dynamic epistemic logic [83]. 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 [38, 40, 14], belief base merging [48], input-output logic [60], as well as more recent work on the so-called "database perspective" to the theory of intention [72] and resource-bounded knowledge and reasoning about strategies [6]. All these approaches defend the idea that the 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.[2]

There is also a practical motivation behind the logic LDA in relation to modelling Theory of Mind (ToM) in social robotics [90, 71, 91] and in the domain of intelligent virtual agents (IVAs) [69, 67, 22, 42]. ToM is the general capacity of ascribing mental attitudes and mental operations to others and of predicting the behavior of others on the basis of this attribution [33]. An important aspect of ToM consists in forming higher-order beliefs about other agents's beliefs. This is essential, among other things, for AI persuasive technologies, since an artificial agent's persuasiveness relies on her capacity of representing the human interlocutor's beliefs in such a way that she can modify them through communication and, consequently, influence the human's behavior.

[1] Other grounded semantics for epistemic logics have been proposed in the AI literature. For instance, a semantics based on the concept of interpreted system is provided in [55], while semantics exploiting the notion of propositional observability are presented in [79, 24].

[2] This approach has also been used in linguistic work on modal expressions. For instance, according to Kratzer [52], conversational common ground can be seen as a 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.

Although existing computational models of ToM used in social robotics and in human-machine interaction (HMI) take this aspect into consideration, they have some limitations. For instance, robotic models of ToM (see, e.g., [53, 62, 23]) only allow to represent higher-order beliefs of depth at most 2. Furthermore, both classes of models do not clearly spell out how the speaker's decision to perform a certain speech act may depend on her higher-order beliefs about the hearer's beliefs, and how the speaker's speech act may affect the hearer's higher-order beliefs.

As shown in [17], the standard epistemic logic (EL) approach [30] and dynamic epistemic logic (DEL) approach [83] overcome these drawbacks by allowing to represent higher-order beliefs of any depth and by offering a framework for formalizing a variety of communication dynamics in a multi-agent setting with the help of so-called action models [12]. Albrecht & Stone [4] include EL and DEL in the category of recursive reasoning methods, in their classification of methods for modelling other agents' minds and predicting other agents' actions. They claim that such methods "...use explicit representations of nested beliefs and "simulate" the reasoning processes of other agents to predict their actions " [4, p. 80].

Unfortunately, standard EL and DEL have other disadvantages. First of all, they do not distinguish between explicit and implicit belief. They only allows to represent what an artificial agent believes a human could *potentially* believe — if she had enough computational resources and time to infer it —, without representing what the artificial agent believes the human *actually* believes. Secondly, modelling complex information dynamics in DEL comes with a price: as shown in [32], in case of private announcements, 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. Thus, the original epistemic model may grow exponentially in the length of the sequence of private announcements. This feature is reflected in the computational aspects of DEL, which exhibits an increase in complexity when moving from public announcements to private forms of communication in a multi-agent setting. In particular, 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., [58, 19]), complexity increases if we move into the realm of full DEL, whose satisfiability problem is known to be NEXPTIME-complete [9]. It is also known that epistemic planning in PAL is decidable, while it becomes undecidable in general DEL, due to the fact that the epistemic model grows as a consequence of a private announcement [18].

The logic LDA we present in this article does not have these disadvantages. First of all, it provides a generalization of the standard EL approach in which the distinction between explicit and implicit belief can be captured: it allows us to represent both what an artificial agent believes a human is explicitly believing in a given situation — which is the essential aspect of ToM — and what an artificial agent believes a human can infer from what she explicitly believes. Secondly, it offers a 'parsimonious' account of private informative actions that — differently from standard DEL — does not require to duplicate epistemic models and to make them exponentially larger. This is due to the fact that private belief change operations are modeled in LDA as set-theoretic operations on the belief base of an agent (the perceiver) which do not affect the belief bases of the non-perceivers. This feature has interesting implications on the

computational level. For instance, we will show that extending LDA by private belief base expansion operators does not increase the PSPACE-complexity of the static logic.

Before discussing related work let us provide some terminological clarifications. We use the term 'private explicit belief change' to refer to change of an agent's explicit beliefs that does not affect the other agents' explicit or implicit beliefs. Symmetrically, we use the term 'private implicit belief change' to refer to change of an agent's implicit beliefs that does not affect the other agents' explicit or implicit beliefs. We use 'private belief change' as a generic term covering both of them.

Related work. The logic LDA belongs to the family of logics for non-omniscient agents. Purely syntactic approaches to the logical omniscience problem have been proposed in which an agent's beliefs are described either by a set of formulas which is not necessarily closed under deduction [28, 64] or by a set of formulas obtained by the application of an incomplete set of deduction rules [49, 46]. Logics of time-bounded reasoning have also been studied [2, 5, 34, 27] in which reasoning is a represented as a process that requires time due to the time-consuming application of inference rules. Other authors [87, 15] have tried to solve the logical omniscience problem by using a non-normal worlds semantics for belief that prevents distributivity of belief operators across implication (so-called Axiom K) from being valid and the rule of necessitation for belief from being admissible.

Justification logic [8] provides a solution to the logical omniscience problem by formalizing a notion of justified belief based on the notion of evidence. The reason why justified belief does not necessarily distribute across implication is that having an immediate evidence in support of φ and having an immediate evidence in support of $\varphi \rightarrow \psi$ does not necessarily imply having an immediate evidence in support ψ, since the agent could draw the conclusion that ψ only through some inference steps.

Finally, logics of (un)awareness have been studied both in AI [29, 80, 3] and economics [63, 43, 36].

As we will show in Section 5, LDA is closely related to Fagin & Halpern (F&H)'s logic of general awareness (LGA), as there exists a polynomial embedding of the former into the latter. Nonetheless, the semantics of the two logics are genuinely different both formally and conceptually. First of all, in the semantics of LGA the notion of doxastic alternative is given as a primitive, while in the LDA semantics it is computed from an agent's belief base. Secondly, the LGA ontology of epistemic attitudes is richer than the LDA ontology, as the former includes the concept of awareness which is not included in the latter. We believe that this is a virtue of LDA compared to LGA. In our opinion, modelling explicit and implicit belief without invoking the notion of awareness is a good thing, as the latter is intrinsically polysemic and ambiguous. This aspect is emphasized by F&H, according to whom the notion of awareness is "...open to a number of interpretations. One of them is that an agent is aware of a formula if he can compute whether or not it is true in a given situation within a certain time or space bound" [29, p. 41].

There are also important differences between LDA and LGA at the level of the belief dynamics. Private explicit belief change in LDA just consists in applying set-theoretic operations on an agent's belief base (e.g., adding a formula to it, removing a formula from it, etc.), without modifying the other agents' belief bases. Moreover, in

LDA private implicit belief change is derivative of private explicit belief change (i.e., an agent's implicit beliefs privately change as a consequence of a change of her belief base). The picture is more convoluted in LGA, since the notions of doxastic alternative and awareness are independent and explicit belief is defined from them. In particular, as shown in [76], in LGA explicit beliefs can change either as a consequence of awareness change, which modifies an agent's awareness function,[3] or as a consequence of implicit belief change, which modifies an agent's set of doxastic alternatives. Hence, modelling private explicit belief change in LGA is at least as complex as modelling private implicit belief change in the standard DEL approach, since the LGA semantics is an extension of the EL semantics by the notion of awareness. As we have emphasized above, the standard DEL semantics for private implicit belief change is intrinsically complex, as it requires to duplicate the original epistemic model by creating one copy of the model for the perceiver and one copy for the non-perceivers. This complication is avoided altogether in our logic LDA.

Another system related to LDA is the logic of local reasoning also presented in [29], in which the distinction between explicit and implicit belief is captured. F&H use a neighborhood semantics for explicit belief: every agent is associated with a set of sets of worlds, called frames of mind. (See also [86, 10] for the use of neighborhood semantics for modelling explicit beliefs.) They define an agent's set of doxastic alternatives as the intersection of the agent's frames of mind. According to F&H's semantics, an agent explicitly believes that φ if and only if she has a frame of mind in which φ is globally true. Moreover, an agent implicitly believes that φ if and only if, φ is true at all her doxastic alternatives. In their semantics, there is no representation of an agent's belief base, corresponding to the set of formulas explicitly believed by the agent. Moreover, differently from the LDA notion of explicit belief, their notion does not completely solve the logical omniscience problem. For instance, while their notion of explicit belief is closed under logical equivalence, the LDA notion is not. Specifically, the following rule of equivalence preserves validity in F&H's logic but not in LDA:

$$\frac{\alpha \leftrightarrow \alpha'}{\triangle_i \alpha \leftrightarrow \triangle_i \alpha'}$$

where $\triangle_i \alpha$ means that agent i has the explicit belief that α. This is a consequence of their use of an extensional semantics for explicit belief. Levesque too provides an extensional semantics for explicit belief with no connection with the notion of belief base [54]. In his logic, explicit beliefs are closed under conjunction, while they are not in our logic LDA.

Plan of the article. The article is organized as follows. In Section 2, we present the language of the family of LDA logics. We talk about the family of LDA logics — or, more shortly, about the LDA logics — instead of a single LDA logic, since we consider different variants of a logic of explicit and implicit belief working under different

[3] Van Benthem & Velasquez define two basic operations of awareness change: the operation of "considering" (or becoming aware of something) and the operation of "dropping" (or becoming unaware of something).

assumptions. For example, we study the LDA logic in which an agent's belief base is assumed to be globally consistent as well as the LDA logic in which an agent's explicit beliefs are assumed to be correct. Then, in Section 3, we introduce a semantics for the LDA language based on the notion of multi-agent belief base. We define two additional Kripke-style semantics in which the notion of doxastic alternative is primitive. These additional semantics will be useful for proving completeness and decidability results for the logics LDA. We show that the three semantics are all equivalent with respect to the formal language under consideration. In Section 4, we provide axiomatizations for the different LDA logics and prove that their satisfiability problems are decidable. In Section 5, we provide polynomial embeddings of the LDA logics into Fagin & Halpern's logic of general awareness (LGA). Thanks to these embeddings and to the known complexity of LGA, we will be able to prove a number of complexity results for the LDA logics. In Section 6, we extend our analysis to logics in the LDA family implementing different forms of introspection for explicit and/or implicit belief. In Section 7, we move from the static to the dynamic setting. We present extensions of LDA by belief change operators both of public and private type. This includes operators for public announcement, private belief base expansion and forgetting. We show how the private belief base expansion operator can be used in the context of an AI application in which a conversational agent is expected to use its persuasive capabilities in its interaction with a human. Section 8 is devoted to the comparison between the LDA approach to private belief change and the DEL approach. We show that the notion of private belief base expansion studied in the LDA setting can be seen as a compact form of private update in the DEL sense, which does not require world duplication. In Section 9, we conclude.[4]

2. A Language for Explicit and Implicit Beliefs

This section presents the language of the logic LDA for representing 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 logical language in two steps.

We first define the language $\mathcal{LANG}_0(Atm, Agt)$ by the following grammar in Backus-Naur Form (BNF):

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

where p ranges over Atm and i ranges over Agt. $\mathcal{LANG}_0(Atm, Agt)$ is the language for representing explicit beliefs of multiple agents. The formula $\triangle_i\alpha$ is read "agent i explicitly (or actually) believes that α is true". In this language, we can represent higher-order explicit beliefs, i.e., an agent's explicit belief about another agent's explicit beliefs.

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

[4] This article is a considerably extended and improved version of [56].

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

where α ranges over $\mathcal{LANG}_0(Atm, Agt)$ and i ranges over Agt. For notational convenience we write $\mathcal{LANG}_0$ instead of $\mathcal{LANG}_0(Atm, Agt)$ and $\mathcal{LANG}_{\mathsf{LDA}}$ instead of $\mathcal{LANG}_{\mathsf{LDA}}(Atm, Agt)$, 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.

For every formula $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$, we write $Atm(\varphi)$ to denote the set of atomic propositions of type p occurring in φ. Moreover, for every set of formulas $X \subseteq \mathcal{LANG}_{\mathsf{LDA}}$, we define $Atm(X) = \bigcup_{\varphi \in X} Atm(\varphi)$.

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:

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

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

Note that the language $\mathcal{LANG}_{\mathsf{LDA}}$ does not allow us to represent explicit beliefs about implicit beliefs. The reason for this syntactic restriction is that the semantics we will present in Section 3 is designed in such a way that an agent's doxastic accessibility relation is computed from the agent's belief base. More generally, the notion of implicit belief is defined in terms of the notion of explicit belief. Therefore, having explicit beliefs about implicit beliefs, would make our definition circular.[5]

3. Formal Semantics

In this section, we present three families of formal semantics for the language of explicit and implicit beliefs defined above. In the first semantics, the notion of doxastic alternative is not primitive but it is defined from the primitive concept of belief base. The second semantics is a Kripke-style semantics, based on the concept of notional model, in which an agent's set of doxastic alternatives coincides with the set of possible worlds in which the agent's explicit beliefs are true. The third semantics is a weaker semantics, based on the concept of *quasi*-notional model. It only requires that an agent's set of doxastic alternatives is included in the set of possible worlds at which the agent's explicit beliefs are true. We will show that the three families of semantics are equivalent with respect to the formal language under consideration.

We consider the first semantics to be the "natural" semantics for the logic $\mathcal{LANG}_{\mathsf{LDA}}$. The reason for introducing the Kripke-style semantics based on notional models is that it is closer to the possible world semantics commonly used in the areas of epistemic logic [30] and modal logic [16] than the belief base semantics. Consequently, we can import methods and techniques from these areas to prove results about axiomatics and complexity for the logic LDA. Another reason is that it eases the tasks of comparing LDA with F&H's logic of general awareness and of defining properties of introspection

[5] A way of avoiding this circularity would be a fixed-point definition of the implicit belief operator which would make our semantics considerably more complex. We leave the study of this alternative semantics for future work.

on beliefs in the LDA framework. These two issues will be investigated in Sections 5 and 6, respectively. The reason for introducing quasi-notional models is purely technical. We can use a standard canonical model argument for proving completeness of LDA relative to this class of models. Then, given the equivalence between the belief base semantics, the notional model semantics and the quasi-notional model semantics for the language $\mathcal{LANG}_{\mathsf{LDA}}$, we have completeness of LDA relative to the three model classes.

The following are some useful properties of binary relations that we will use in different parts of the article. Let S be a set and let $\mathcal{R} \subseteq S \times S$. We say that:

- the relation $\mathcal{R}$ is serial if and only if, for every $s \in S$, there exists $s' \in S$ such that $s\mathcal{R}s'$;
- the relation $\mathcal{R}$ is reflexive if and only if, for every $s \in S$, $s\mathcal{R}s$;
- the relation $\mathcal{R}$ is transitive if and only if, for every $s, s', s'' \in S$, if $s\mathcal{R}s'$ and $s'\mathcal{R}s''$ then $s\mathcal{R}s''$;
- the relation $\mathcal{R}$ is Euclidean if and only if, for every $s, s', s'' \in S$, if $s\mathcal{R}s'$ and $s\mathcal{R}s''$ then $s'\mathcal{R}s''$.

3.1. Multi-agent belief base semantics

We first consider the semantics based on the concept of multi-agent belief base that is defined as follows.

Definition 1 (Multi-agent belief base). *A multi-agent belief base is a tuple* $B = (B_1, \ldots, B_n, V)$ *where:*

- *for every* $i \in Agt$, $B_i \subseteq \mathcal{LANG}_0$ *is agent* i*'s belief base,*
- $V \subseteq Atm$ *is the actual state.*

The set of multi-agent belief bases is denoted by $\mathbf{B}$.

The sublanguage $\mathcal{LANG}_0(Atm, Agt)$ is interpreted with respect to multi-agent belief bases, as follows.

Definition 2 (Satisfaction relation). *Let* $B = (B_1, \ldots, B_n, V) \in \mathbf{B}$. *Then:*

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

Observe in particular the set-theoretic interpretation of the explicit belief operator: agent i explicitly believes that α if and only if α is included in her belief base.

It is worth to consider correct multi-agent belief bases according to which every fact that an agent explicitly believes has to be true.

Definition 3 (Correct multi-agent belief base). *The multi-agent belief base* $B = (B_1, \ldots, B_n, V)$ *is said to be correct if and only if, for every* $i \in Agt$ *and for every* $\alpha \in \mathcal{LAN}\mathcal{G}_0$*, if* $\alpha \in B_i$ *then* $B \models \alpha$.

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 [73], i.e., the body of information that the agents commonly believe to be the case.

Definition 4 (Multi-agent belief model). *A multi-agent belief model (MAB) is a pair* (B, Cxt)*, where* $B \in \mathbf{B}$ *and* $Cxt \subseteq \mathbf{B}$.

Note that in the previous definition we do not require $B \in Cxt$. Let us illustrate the concept of MAB with the aid of an example.

Example *Let* $Agt = \{1, 2\}$ *and* $\{p, q\} \subseteq Atm$*. Moreover, let* (B_1, B_2, V) *be such that:*

$$\begin{aligned} B_1 &= \{p, \triangle_2 p\}, \\ B_2 &= \{p\}, \\ V &= \{p, q\}. \end{aligned}$$

Suppose that the agents have in their common ground the fact $p \rightarrow q$*. In other words, they commonly believe that* p *implies* q*. This means that:*

$$Cxt = \{B' \in \mathbf{B} : B' \models p \rightarrow q\}.$$

The following definition introduces the concept of doxastic alternative.

Definition 5 (Doxastic alternatives). *Let* $i \in Agt$*. Then,* $\mathcal{R}_i$ *is the binary relation on the set of multi-agent belief bases* $\mathbf{B}$ *such that, for all* $B = (B_1, \ldots, B_n, V), B' = (B'_1, \ldots, B'_n, V') \in \mathbf{B}$*:*

$$B\mathcal{R}_i B' \text{ if and only if } \forall \alpha \in B_i : B' \models \alpha.$$

$B\mathcal{R}_i B'$ means that B' is a doxastic alternative for agent i at B (i.e., at B agent i considers B' possible). 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. Observe that $B\mathcal{R}_i B'$ does not imply $B_i \subseteq B'$. In Section 9, we will discuss a variant of the logic LDA for which the implication from $B\mathcal{R}_i B'$ to $B_i \subseteq B'$ holds.

The following definition generalizes Definition 2 of satisfaction relation to the full language $\mathcal{LAN}\mathcal{G}_{\mathsf{LDA}}$. Its formulas are interpreted with respect to MABs. (Boolean cases are omitted, as they are defined in the usual way.)

Definition 6 (Satisfaction relation (cont.)). *Let* (B, Cxt) *be a MAB. Then:*

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

Let us go back to the example.

Example (cont.) *It is easy to check that the following holds:*

$$(B, Cxt) \models \Box_1(p \wedge q) \wedge \Box_2(p \wedge q) \wedge \Box_1\Box_2(p \wedge q).$$

Indeed, we have:

$$\begin{aligned}\big(\mathcal{R}_1(B) \cap Cxt\big) &\subseteq \{B' \in \mathbf{B} : B' \models p \wedge q\},\\ \big(\mathcal{R}_2(B) \cap Cxt\big) &\subseteq \{B' \in \mathbf{B} : B' \models p \wedge q\},\\ \Big(\big(\mathcal{R}_1 \circ \mathcal{R}_2(B)\big) \cap Cxt\Big) &\subseteq \{B' \in \mathbf{B} : B' \models p \wedge q\},\end{aligned}$$

where $\circ$ is the composition operation between binary relations and $\mathcal{R}_i(B) = \{B' \in \mathbf{B} : B\mathcal{R}_iB'\}$.

We focus on subclasses of MABs that guarantee, respectively, global consistency of the agents' belief bases and correctness of the agents' beliefs. For the sake of exposition, when talking about correct (or true) explicit (resp. implicit) belief, we sometimes use the terms explicit (resp. implicit) knowledge. Indeed, we assume that the terms "true belief", "correct belief" and "knowledge" are all synonyms.

Definition 7 (Global consistency for MAB). *The MAB (B, Cxt) satisfies global consistency (GC) if and only if, for every $i \in Agt$ and for every $B' \in (\{B\} \cup Cxt)$, there exists $B'' \in Cxt$ such that $B'\mathcal{R}_iB''$.*

Global consistency means that in every possible situation an agent has at least one doxastic alternative. Saying that (B, Cxt) satisfies GC is the same thing as saying that, for every $i \in Agt$, the relation $\mathcal{R}_i \cap \big((\{B\} \cup Cxt) \times Cxt\big)$ is serial.

We distinguish global consistency of a MAB from local consistency of the agents' belief bases. Local consistency can mean different things. For instance, it could mean that (i) for every $i \in Agt$, for every $B' \in (\{B\} \cup Cxt)$ and for every $\alpha \in \mathcal{LAN}\mathcal{G}_0$, $\alpha \wedge \neg\alpha \notin B'_i$, or that (ii) for every $i \in Agt$, for every $B' \in (\{B\} \cup Cxt)$ and for every $\alpha \in \mathcal{LAN}\mathcal{G}_0$, $\{\alpha, \neg\alpha\} \not\subseteq B'_i$. Global consistency defined above implies local consistency in both senses, but not vice versa. Indeed, an agent's beliefs could be locally consistent in one of the two senses, and they are rendered inconsistent by means of purely deductive inference.

Definition 8 (Belief correctness for MAB). *The MAB (B, Cxt) satisfies belief correctness (BC) if and only if $B \in Cxt$ and, for every $i \in Agt$ and for every $B' \in Cxt$, $B'\mathcal{R}_iB'$.*

Saying that (B, Cxt) satisfies BC is the same thing as saying that $B \in Cxt$ and, for every $i \in Agt$, the relation $\mathcal{R}_i \cap (Cxt \times Cxt)$ is reflexive.

For every $X \subseteq \{GC, BC\}$, we denote by $\mathbf{MAB}_X$ the class of MABs satisfying every condition in X. $\mathbf{MAB}_\emptyset$ is the class of all MABs. For notational convenience, we sometimes write $\mathbf{MAB}$ instead of $\mathbf{MAB}_\emptyset$. We have $\mathbf{MAB}_{\{BC\}} \subseteq \mathbf{MAB}_{\{GC\}}$. In fact, BC implies GC since if $B \in Cxt$ and $B'\mathcal{R}_iB'$ for every $B' \in Cxt$ then,

for every $B' \in (\{B\} \cup Cxt)$, there exists $B'' \in Cxt$ such that $B'\mathcal{R}_iB''$. Therefore, $\mathbf{MAB}_{\{GC,BC\}} = \mathbf{MAB}_{\{BC\}}$.

Note that the condition $B \in Cxt$ in Definition 8 is necessary to make agents' implicit beliefs correct, i.e., to make the formula $\Box_i\varphi \rightarrow \varphi$ valid.

As the following proposition highlights, belief correctness for MABs is completely characterized by the fact that the actual world is included in the agents' common ground and that the agents' explicit beliefs are correct in the sense of Definition 3.

Proposition 1. *Let (B, Cxt) be a MAB. Then, it satisfies BC if and only if $B \in Cxt$ and, for every $B' \in Cxt$, B' is correct.*

PROOF. (B, Cxt) satisfies belief correctness iff $B \in Cxt$ and for every $i \in Agt$ and for every $B' \in Cxt$, $B'\mathcal{R}_iB'$. By Definition 5, the latter is equivalent to stating that $B \in Cxt$ and for every $i \in Agt$, for every $B' \in Cxt$ and for every $\alpha \in B'_i$, if $\alpha \in B'_i$ then $B' \models \alpha$. By Definition 3, the latter means that $B \in Cxt$ and every $B' \in Cxt$ is correct. ■

Let $\varphi \in \mathcal{LAN}\mathcal{G}_{\mathsf{LDA}}$, we say that φ is valid for the class $\mathbf{MAB}_X$ if and only if, for every $(B, Cxt) \in \mathbf{MAB}_X$ we have $(B, Cxt) \models \varphi$. We say that φ is satisfiable for the class $\mathbf{MAB}_X$ if and only if $\neg\varphi$ is not valid for the class $\mathbf{MAB}_X$.

In the article, given a class of models $\mathbf{M}$ and formula φ, we use the symbol $\models_{\mathbf{M}} \varphi$ to mean that the formula φ is valid relative to the class $\mathbf{MAB}$. For instance, $\models_{\mathbf{MAB}_X} \varphi$ means that the formula φ is relative to the class $\mathbf{MAB}_X$.

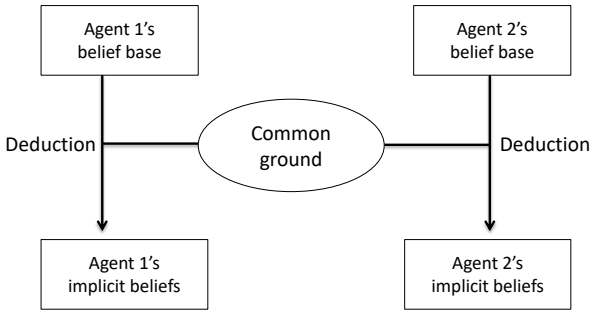


Figure 1: Conceptual framework

Figure 1 resumes the general idea behind the LDA framework, 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. Common ground should be conceived as a sort of implicit common belief. Indeed, as emphasized by Stalnaker, common ground can be described as the "...presumed *background* information shared by participants in a conversation..." [73,

p. 701]. An information in the common ground, being in background, is taken into consideration by the agents when making inferences, but is not necessarily part of an agent's belief base before an inference is made.

3.2. Notional model semantics

Let us now define a new semantics for the language $\mathcal{LANG}_{\mathsf{LDA}}$ which extends the standard multi-relational Kripke semantics of epistemic logic by agents' belief bases.

Definition 9 (Notional doxastic model). *A notional doxastic model (NDM) is a tuple* $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ *where:*

- W *is a set of worlds,*
- $\mathcal{D} : Agt \times W \longrightarrow 2^{\mathcal{LANG}_0}$ *is a doxastic function,*
- $\mathcal{N} : Agt \times W \longrightarrow 2^W$ *is a notional function,*
- $\mathcal{V} : Atm \longrightarrow 2^W$ *is a valuation function,*

and such that, given the following inductive definition of the semantic interpretation of formulas $\mathcal{LANG}_{\mathsf{LDA}}$ *relative to a pair* (M, w) *with* $w \in W$*:*

$$\begin{array}{rcl}
(M,w) \models p & \iff & w \in \mathcal{V}(p), \\
(M,w) \models \neg\varphi & \iff & (M,w) \not\models \varphi, \\
(M,w) \models \varphi \wedge \psi & \iff & (M,w) \models \varphi \text{ and } (M,w) \models \psi, \\
(M,w) \models \triangle_i \alpha & \iff & \alpha \in \mathcal{D}(i,w), \\
(M,w) \models \Box_i \varphi & \iff & \forall v \in \mathcal{N}(i,w) : (M,v) \models \varphi,
\end{array}$$

it satisfies the following condition, for all $i \in Agt$ *and for all* $w \in W$*:*

(C1) $\mathcal{N}(i,w) = \bigcap_{\alpha \in \mathcal{D}(i,w)} ||\alpha||_M$*,*

with $||\alpha||_M = \{v \in W : (M,v) \models \alpha\}$.

We call the pair (M, w) in the previous definition pointed NDM.

For every agent i and for every world w, $\mathcal{D}(i,w)$ denotes agent i's set of explicit beliefs at w.

The set $\mathcal{N}(i,w)$, used in the interpretation of the implicit belief operator $\Box_i$, is called agent i's set of *notional* worlds at world w. The term 'notional' is borrowed from the philosopher D. Dennett [25, 26] (see, also, [49]): an agent's notional world is a world at which all the agent's explicit beliefs are true. This idea is clearly expressed by Condition C1.

In order to relate NDMs with the MABs defined in Section 3.1, we consider specific subclasses in which the accessibility relations

$$\mathcal{N}_i = \{(w,v) \in W \times W : v \in \mathcal{N}(i,w)\}$$

are assumed to be serial and reflexive, respectively.

Definition 10 (Global consistency for NDM). *The NDM $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ satisfies global consistency (GC) if and only if, for every $i \in Agt$ and for every $w \in W$, $\mathcal{N}(i, w) \neq \emptyset$.*

Global consistency for NDMs just means that an agent's set of notional worlds must be non-empty, i.e., there exists at least one situation which is compatible with what an agent explicitly believes.

Definition 11 (Belief correctness for NDM). *The NDM $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ satisfies belief correctness (BC) if and only if, for every $i \in Agt$ and for every $w \in W$, $w \in \mathcal{N}(i, w)$.*

Belief correctness for NDMs just means that an agent's set of notional worlds must include the actual world.

As in Section 3.1, we define different model classes satisfying such properties. For every $X \subseteq \{GC, BC\}$, we denote by $\mathbf{NDM}_X$ the class of NDMs satisfying every condition in X. $\mathbf{NDM}_\emptyset$ is the class of all NDMs. For notational convenience, we sometimes write $\mathbf{NDM}$ instead of $\mathbf{NDM}_\emptyset$. As for MABs, we have $\mathbf{NDM}_{\{BC\}} \subseteq \mathbf{NDM}_{\{GC\}}$.

We say that a NDM $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ is *finite* if and only if W, $\mathcal{D}(i, w)$ and $\mathcal{V}^{\leftarrow}(w)$ are finite sets for every $i \in Agt$ and for every $w \in W$, where

$$\mathcal{V}^{\leftarrow}(w) = \{p \in Atm : w \in \mathcal{V}(p)\}.$$

The class of finite NDMs satisfying every condition in $X \subseteq \{GC, BC\}$ is denoted by finite-$\mathbf{NDM}_X$.

Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$, we say that φ is valid for the class $\mathbf{NDM}_X$ if and only if, for every $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V}) \in \mathbf{NDM}_X$ and for every $w \in W$, we have $(M, w) \models \varphi$. We say that φ is satisfiable for the class $\mathbf{NDM}_X$ if and only if $\neg\varphi$ is not valid for the the class $\mathbf{NDM}_X$.

3.3. Quasi-model semantics

In this section we provide an alternative semantics for the language $\mathcal{LANG}_{\mathsf{LDA}}$ based on a more general class of models, called quasi-notional doxastic models (quasi-NDMs). This semantics will be fundamental for proving completeness of the logics we will define in Section 4.

Definition 12 (Quasi-model). *A quasi-notional doxastic model (quasi-NDM) is a tuple $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ where $W, \mathcal{D}, \mathcal{N}$ and $\mathcal{V}$ are as in Definition 9 except that Condition C1 is replaced by the following weaker condition, for all $i \in Agt$ and for all $w \in W$:*

(C1)* $\mathcal{N}(i, w) \subseteq \bigcap_{\alpha \in \mathcal{D}(i,w)} ||\alpha||_M$.

As for NDMs, for every $X \subseteq \{GC, BC\}$, we denote by $\mathbf{QNDM}_X$ the class of quasi-NDMs satisfying every condition in X. $\mathbf{QNDM}_\emptyset$ is the class of all quasi-NDMs. We sometimes write $\mathbf{QNDM}$ instead of $\mathbf{QNDM}_\emptyset$. As for MABs and NDMs, we have $\mathbf{QNDM}_{\{BC\}} \subseteq \mathbf{QNDM}_{\{GC\}}$.

Truth conditions of formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ relative to quasi-NDMs are the same as truth conditions of formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ relative to the NDMs. Validity and satisfiability of formulas for a class $\mathbf{QNDM}_X$ are defined in the usual way.

As for NDMs, we say that the quasi-NDM $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ is *finite* if and only if W, $\mathcal{D}(i, w)$ and $\mathcal{V}^{\leftarrow}(w) = \{p \in Atm : w \in \mathcal{V}(p)\}$ are finite sets for every $i \in Agt$ and for every $w \in W$. The class of finite quasi-NDMs satisfying every condition in $X \subseteq \{GC, BC\}$ is denoted by finite-$\mathbf{QNDM}_X$.

3.4. Equivalence between semantics

The main technical result of this section is an equivalence result between the five different semantics for the language $\mathcal{LANG}_{\mathsf{LDA}}$ defined in Sections 3.1, 3.2 and 3.3, namely, the multi-agent belief base semantics, the notional model semantics, the finite notional model semantics, the quasi-notional model semantics, and the finite quasi-notional model semantics.

Theorem 1. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$ and let $X \subseteq \{GC, BC\}$. Then, the following five statements are equivalent:*

- *φ is satisfiable for the class $\mathbf{QNDM}_X$,*
- *φ is satisfiable for the class finite-$\mathbf{QNDM}_X$,*
- *φ is satisfiable for the class finite-$\mathbf{NDM}_X$,*
- *φ is satisfiable for the class $\mathbf{NDM}_X$,*
- *φ is satisfiable for the class $\mathbf{MAB}_X$.*

The proof of Theorem 1 is given in the technical annex at the end of the paper.

4. Logic of Doxastic Attitudes

This section is devoted to present a number of logics of explicit and implicit belief that are interpreted relative to the semantics defined in the previous section. They are all extensions of a basic logic called LDA which stands for "Logic of Doxastic Attitudes". We will provide axiomatics and decidability results for this family of logics.

4.1. Definition of the Logic

The following is the definition of the logics in the LDA family, each logic being defined by a set of axioms and rules of inference.

Definition 13 (LDA). *We define LDA to be the extension of classical propositional logic by the following axioms and rule of inference:*

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

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

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

For every $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$, we define LDA_X to be the extension of the logic LDA by each axiom in X, where $\mathbf{D}_{\Box_i}$ and $\mathbf{T}_{\Box_i}$ are the following axioms:

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

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

As usual, for every logic LDA_X with $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$ and for every $\varphi \in \mathcal{LAN}\mathcal{G}_{\mathsf{LDA}}$, we write $\vdash_{\mathsf{LDA}_X} \varphi$ to mean that φ is deducible in LDA_X, that is, there is a sequence of formulas $(\varphi_1, \ldots, \varphi_m)$ such that:

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

We say that the set of formulas Γ from $\mathcal{LAN}\mathcal{G}_{\mathsf{LDA}}$ is LDA_X-consistent if there are no formulas $\varphi_1, \ldots, \varphi_m \in \Gamma$ such that $\vdash_{\mathsf{LDA}_X} (\varphi_1 \wedge \ldots \wedge \varphi_m) \rightarrow \bot$. Moreover, φ is LDA_X-consistent if $\{\varphi\}$ is LDA_X-consistent.

According to the previous definition, LDA (*alias*, $\mathsf{LDA}_\emptyset$) is the most general system for reasoning about explicit and implicit beliefs of multiple agents. It includes the principles of system K for the implicit belief operator $\Box_i$ as well as an Axiom $\mathbf{Int}_{\triangle_i,\Box_i}$ relating explicit belief with implicit belief. All other logics are extensions of LDA by specific principles about consistency of implicit beliefs (Axiom $\mathbf{D}_{\Box_i}$) and correctness of implicit beliefs (Axiom $\mathbf{T}_{\Box_i}$).

The logic $\mathsf{LDA}_{\{\mathbf{D}_{\Box_i}\}}$ includes the principles of system KD for the implicit belief operator $\Box_i$, while the logic $\mathsf{LDA}_{\{\mathbf{T}_{\Box_i}\}}$ includes the principles of system KT for the implicit belief operator $\Box_i$. $\mathsf{LDA}_{\{\mathbf{T}_{\Box_i}\}}$ has to be conceived as a logic of explicit and implicit knowledge, since knowledge is necessarily veridical while a mere belief might be wrong. Clearly, the logics $\mathsf{LDA}_{\{\mathbf{D}_{\Box_i},\mathbf{T}_{\Box_i}\}}$ and $\mathsf{LDA}_{\{\mathbf{T}_{\Box_i}\}}$ coincide since Axiom $\mathbf{D}_{\Box_i}$ is deducible in $\mathsf{LDA}_{\{\mathbf{T}_{\Box_i}\}}$ by means of Axiom $\mathbf{T}_{\Box_i}$. By means of Axioms $\mathbf{T}_{\Box_i}$ and $\mathbf{Int}_{\triangle_i,\Box_i}$, in $\mathsf{LDA}_{\{\mathbf{T}_{\Box_i}\}}$ we can, moreover, derive the following veridicality property for explicit knowledge:

$$\vdash_{\mathsf{LDA}_{\{\mathbf{T}_{\Box_i}\}}} \triangle_i\alpha \rightarrow \alpha. \qquad (1)$$

Note that there is no consensus in the literature about introspection for implicit belief. For instance, in his seminal work on the logics of knowledge and belief [45], Hintikka only assumed positive introspection for belief (Axiom 4) and rejected negative introspection (Axiom 5). Other logicians such as Jones [47] have argued against the use of both positive and negative introspection axioms for belief. Nonetheless, all approaches unanimously assume that a reasonable notion of implicit belief should satisfy Axioms K and D. On this point, see also [13]. In this sense, the logic $\mathsf{LDA}_{\{\mathbf{D}_{\Box_i}\}}$ can be conceived as the minimal logic of explicit and implicit belief. We will go back to the issue of introspection for belief in Section 6.

As we have emphasized above, explicit beliefs are conceivable as an agent's *actual* beliefs, while implicit beliefs are conceivable as *potential* beliefs that the agent can

form through inference from her explicit beliefs and the common ground. Clearly, an actual belief is also a potential belief, but not vice versa. This explains why we have $\vdash_{\mathsf{LDA}_X} \triangle_i\alpha \rightarrow \Box_i\alpha$, for every $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$, but $\nvdash_{\mathsf{LDA}_X} \Box_i\alpha \rightarrow \triangle_i\alpha$. Note that Axiom $\mathbf{Int}_{\triangle_i,\Box_i}$ makes sense not only when α is a propositional formula, but also when it contains epistemic modalities. Let us justify this claim with the aid of an example. Suppose agent i explicitly believes that p (i.e., $\triangle_i p$), that $p \rightarrow q$ (i.e., $\triangle_i(p \rightarrow q)$) and that she does not believe that q explicitly (i.e., $\triangle_i\neg\triangle_i q$). By Axiom $\mathbf{Int}_{\triangle_i,\Box_i}$, it follows that agent i implicitly believes that she does not believe that q explicitly (i.e., $\Box_i\neg\triangle_i q$) and, by Axiom $\mathbf{K}_{\Box_i}$ together with Axiom $\mathbf{Int}_{\triangle_i,\Box_i}$, it follows that agent i implicitly believes that q (i.e., $\Box_i q$). An intuitive explanation of why the following formula

$$\triangle_i p \wedge \triangle_i(p \rightarrow q) \wedge \triangle_i\neg\triangle_i q \wedge \Box_i\neg\triangle_i q \wedge \Box_i q$$

is LDA_X-consistent, for every $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$, is that q is the conclusion of agent i's inferential process but, according to agent i, q is not included in the set of premises from which her inference started.

In the next section, we are going to prove completeness of each logic LDA_X. To this aim, we introduce the correspondence function cf between Axioms $\mathbf{D}_{\Box_i}$ and $\mathbf{T}_{\Box_i}$ and their semantics counterparts of global consistency (GC) and belief correctness (BC) defined in Section 3:

- $cf(\mathbf{D}_{\Box_i}) = GC$,
- $cf(\mathbf{T}_{\Box_i}) = BC$.

4.2. Completeness and Decidability

This section is devoted to prove completeness and decidability results for the logics in the LDA family.

We first prove completeness relative to the quasi-notional model semantics by using a canonical model argument.

We consider maximally LDA_X-consistent sets of formulas in $\mathcal{LAN}\mathcal{G}_{\mathsf{LDA}}$ (LDA_X-MCSs) with $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. The following proposition specifies some usual properties of MCSs.

Proposition 2. *Let Γ be a LDA_X-MCS with $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. Then:*

- *if $\varphi, \varphi \rightarrow \psi \in \Gamma$ then $\psi \in \Gamma$;*
- *$\varphi \in \Gamma$ or $\neg\varphi \in \Gamma$;*
- *$\varphi \vee \psi \in \Gamma$ iff $\varphi \in \Gamma$ or $\psi \in \Gamma$.*

The following is the Lindenbaum's lemma for our logics. Its proof is standard (cf. Lemma 4.17 in [16]) and we omit it.

Lemma 1. *Let Γ be a LDA_X-consistent set of formulas with $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. Then, there exists a LDA_X-MCS Γ' such that $\Gamma \subseteq \Gamma'$.*

Let the LDA_X-canonical model be the tuple $M^{\mathsf{LDA}_X} = (W^{\mathsf{LDA}_X}, \mathcal{D}^{\mathsf{LDA}_X}, \mathcal{N}^{\mathsf{LDA}_X}, \mathcal{V}^{\mathsf{LDA}_X})$ such that:

- W^{LDA_X} is set of all LDA_X-MCSs;
- for all $w \in W^{\mathsf{LDA}_X}$, for all $i \in Agt$ and for all $\alpha \in \mathcal{LANG}_0$, $\alpha \in \mathcal{D}^{\mathsf{LDA}_X}(i, w)$ iff $\triangle_i \alpha \in w$;
- for all $w, v \in W^{\mathsf{LDA}_X}$ and for all $i \in Agt$, $v \in \mathcal{N}^{\mathsf{LDA}_X}(i, w)$ iff, for all $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$, if $\Box_i \varphi \in w$ then $\varphi \in v$;
- for all $w \in W^{\mathsf{LDA}_X}$ and for all $p \in Atm$, $w \in \mathcal{V}^{\mathsf{LDA}_X}(p)$ iff $p \in w$.

The next step in the proof consists in stating the following existence lemma. The proof is again standard (cf. Lemma 4.20 in [16]) and we omit it.

Lemma 2. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$ and let $w \in W^{\mathsf{LDA}_X}$ with $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. Then, if $\Diamond_i \varphi \in w$ then there exists $v \in \mathcal{N}^{\mathsf{LDA}_X}(i, w)$ such that $\varphi \in v$.*

Then, we prove the following truth lemma.

Lemma 3. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$ and let $w \in W^{\mathsf{LDA}_X}$ with $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. Then, $M^{\mathsf{LDA}_X}, w \models \varphi$ iff $\varphi \in w$.*

PROOF. The proof is by induction on the structure of the formula. The cases with φ atomic, Boolean, and of the form $\Box_i \psi$ are provable in the standard way by means of Proposition 2 and Lemma 2 (cf. Lemma 4.21 in [16]). The proof for the case $\varphi = \triangle_i \alpha$ goes as follows: $\triangle_i \alpha \in w$ iff $\alpha \in \mathcal{D}^{\mathsf{LDA}_X}(i, w)$ iff $M^{\mathsf{LDA}_X}, w \models \triangle_i \alpha$. ■

The last step consists in proving that the LDA_X-canonical model belongs to the appropriate model class for the logic LDA_X.

Proposition 3. *Let $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. Then, $M^{\mathsf{LDA}_X} \in \mathbf{QNDM}_{\{cf(x):x \in X\}}$.*

PROOF. We have to prove that M^{LDA_X} satisfies Condition C1* in Definition 12. To this aim, we just need to prove that if $\alpha \in \mathcal{D}^{\mathsf{LDA}_X}(i, w)$ then $\mathcal{N}^{\mathsf{LDA}_X}(i, w) \subseteq ||\alpha||_{M^{\mathsf{LDA}_X}}$. Suppose $\alpha \in \mathcal{D}^{\mathsf{LDA}_X}(i, w)$. Thus, $\triangle_i \alpha \in w$. Hence, by Axiom $\mathbf{Int}_{\triangle_i, \Box_i}$ and Proposition 2, $\Box_i \alpha \in w$. By the definition of M^{LDA_X}, it follows that, for all $v \in \mathcal{N}^{\mathsf{LDA}_X}(i, w)$, $\alpha \in v$. Thus, by Lemma 3, for all $v \in \mathcal{N}^{\mathsf{LDA}_X}(i, w)$, $(M^{\mathsf{LDA}_X}, v) \models \alpha$. The latter means that $\mathcal{N}^{\mathsf{LDA}_X}(i, w) \subseteq ||\alpha||_{M^{\mathsf{LDA}_X}}$.

It is easy to verify that if $\mathbf{D}_{\Box_i} \in X$ then M^{LDA_X} satisfies the condition of global consistency of Definition 10 and that if $\mathbf{T}_{\Box_i} \in X$ then M^{LDA_X} satisfies the condition of belief correctness of Definition 11. ■

The following is the central result of this section.

Theorem 2. *Let $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. Then, the logic LDA_X is sound and complete for the class $\mathbf{QNDM}_{\{cf(x):x \in X\}}$.*

PROOF. As for soundness, it is routine to check that the axioms of LDA_X are all valid for the class $\mathbf{QNDM}_{\{cf(x):x\in X\}}$ and that the rule of inference $\mathbf{Nec}_{\Box_i}$ preserves validity.

As for completeness, suppose φ is a LDA_X-consistent formula in $\mathcal{LANG}_{\mathsf{LDA}}$. By Lemma 1, there exists $w \in W^{\mathsf{LDA}_X}$ such that $\varphi \in w$. Hence, by Lemma 3, there exists $w \in W^{\mathsf{LDA}_X}$ such that $M^{\mathsf{LDA}_X}, w \models \varphi$. Since, by Proposition 3, $M^{\mathsf{LDA}_X} \in \mathbf{QNDM}_{\{cf(x):x\in X\}}$, φ is satisfiable for the class $\mathbf{QNDM}_{\{cf(x):x\in X\}}$. ■

Completeness of each logic LDA_X relative to their corresponding notional model semantics and multi-agent belief base semantics is a corollary of Theorem 2 and Theorem 1.

Corollary 1. *Let $X \subseteq \{\mathbf{D}_{\Box_i}, \mathbf{T}_{\Box_i}\}$. Then,*

- *LDA_X is sound and complete for the class $\mathbf{NDM}_{\{cf(x):x\in X\}}$, and*
- *LDA_X is sound and complete for the class $\mathbf{MAB}_{\{cf(x):x\in X\}}$.*

PROOF. As for the fist item, it is routine exercise to verify that LDA_X is sound for the class $\mathbf{NDM}_{\{cf(x):x\in X\}}$. Now, suppose that formula φ is LDA_X-consistent. Then, by Theorem 1 and Theorem 2, there exists a finite $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V}) \in \mathbf{QNDM}_X$ and $w \in W$ such that $(M, w) \models \varphi$. Hence, again by Theorem 1, there exists a finite $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V}) \in \mathbf{NDM}_X$ and $w \in W$ such that $(M, w) \models \varphi$. Thus, more generally, φ is satisfiable for the class $\mathbf{NDM}_X$. The second item is a direct consequence of the first item and Theorem 1. ■

We conclude our investigation of the mathematical and computational properties of the LDA logics with a decidability result for their satisfiability problems. Like the proof Theorem 1, the proof of Theorem 3 is given in the technical annex at the end of the paper.

Theorem 3. *Let $X \subseteq \{GC, BC\}$. Then, checking satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ relative to the class $\mathbf{MAB}_X$ is decidable.*

5. Relationship with logic of general awareness and complexity

In this section, we explore the connection between the LDA logics and the logic of general awareness (LGA) by Fagin & Halpern (F&H) [29]. In particular, we will provide polynomial embeddings of the former into the latter and, thanks to these embeddings, we will be able to state complexity results for their satisfiability problems.

The language of the logic of general awareness, denoted by $\mathcal{LANG}_{\mathsf{LGA}}$, is defined by the following grammar:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \mathsf{B}_i\varphi \mid \mathsf{A}_i\varphi \mid \mathsf{X}_i\varphi,$$

where p ranges over Atm and i ranges over Agt.

The formula $\mathsf{A}_i\varphi$ has to be read "agent i is aware of φ". The operators B_i and X_i have the same interpretations as the operators $\Box_i$ and $\triangle_i$ of the language $\mathcal{LANG}_{\mathsf{LDA}}$.

Specifically, $\mathsf{B}_i\varphi$ has to be read "agent i has an implicit belief that φ is true", while $\mathsf{X}_i\varphi$ has to be read "agent i has an explicit belief that φ is true".

The previous language is interpreted with respect to so-called awareness structures, that is, tuples of the form $M = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \pi)$ where S is a set of states, every $\Rightarrow_i \subseteq S \times S$ is a doxastic accessibility relation, every $\mathcal{A}_i : S \longrightarrow 2^{\mathcal{LANG}_{\mathsf{LGA}}}$ is an awareness function and $\pi : Atm \longrightarrow 2^S$ is a valuation function for atomic propositions. Let us denote by $\mathbf{AS}$ the class of awareness structures.

In order to relate LGA with each logic LDA_X defined in Section 4, we focus on specific subclasses of awareness structures.

We say that an awareness structure $M = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \pi)$ satisfies global consistency (GC) if every relation $\Rightarrow_i$ with $1 \leq i \leq n$ is serial. We say that it satisfies belief correctness (BC) if every relation $\Rightarrow_i$ with $1 \leq i \leq n$ is reflexive. For every $X \subseteq \{GC, BC\}$, we denote by $\mathbf{AS}_X$ the class of awareness structures satisfying every condition in X.

In the logic of general awareness, the satisfaction relation is between formulas and pointed models (M, s), where $M = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \pi)$ is an awareness structure and $s \in S$ is a state:

$$\begin{array}{rcl}
(M,s) \models p & \iff & s \in \pi(p), \\
(M,s) \models \neg\varphi & \iff & (M,s) \not\models \varphi, \\
(M,s) \models \varphi_1 \wedge \varphi_2 & \iff & (M,s) \models \varphi_1 \text{ and } (M,s) \models \varphi_2, \\
(M,s) \models \mathsf{B}_i\varphi & \iff & \forall s' \in S : \text{ if } s \Rightarrow_i s' \text{ then } (M,s') \models \varphi, \\
(M,s) \models \mathsf{A}_i\varphi & \iff & \varphi \in \mathcal{A}_i(s), \\
(M,s) \models \mathsf{X}_i\varphi & \iff & (M,s) \models \mathsf{B}_i\varphi \text{ and } (M,s) \models \mathsf{A}_i\varphi.
\end{array}$$

Let us define the following translation tr from the language $\mathcal{LANG}_{\mathsf{LDA}}$ of the logic LDA to the language $\mathcal{LANG}_{\mathsf{LGA}}$ of the logic LGA:

$$\begin{array}{rcl}
tr(p) & = & p \text{ for } p \in Atm, \\
tr(\neg\varphi) & = & \neg tr(\varphi), \\
tr(\varphi_1 \wedge \varphi_2) & = & tr(\varphi_1) \wedge tr(\varphi_2), \\
tr(\triangle_i\alpha) & = & \mathsf{X}_i tr(\alpha), \\
tr(\Box_i\varphi) & = & \mathsf{B}_i tr(\varphi).
\end{array}$$

We extend the translation tr to sets of formulas by defining $tr(X) = \{tr(\varphi) : \varphi \in X\}$, for each $X \subseteq \mathcal{LANG}_{\mathsf{LDA}}$.

Although the logics LDA and LGA are fundamentally different at the semantic level, they are very close at the syntactic level. In fact, as the translation tr highlights, the language $\mathcal{LANG}_{\mathsf{LDA}}$ is a notational variant of $\mathcal{LANG}_{\mathsf{LGA}}$.

The previous translation allows us to embed every logic LDA_X into LGA.

Theorem 4. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$ and let $X \subseteq \{GC, BC\}$. Then, φ is satisfiable for the class $\mathbf{NDM}_X$ if and only if $tr(\varphi)$ is satisfiable for the class $\mathbf{AS}_X$.*

PROOF. We first prove the left-to-right direction. Let $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ be a NDM and let $w \in W$ such that $(M, w) \models \varphi$. We build the corresponding structure $M' = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \pi)$ as follows:

- $S = W$,
- for every $i \in Agt$ and for every $w \in W$, $\Rightarrow_i = \{(w, v) \in W \times W : v \in \mathcal{N}(i, w)\}$,
- for every $i \in Agt$ and for every $w \in W$, $\mathcal{A}_i(w) = tr\big(\mathcal{D}(i, w)\big)$,
- for every $p \in Atm$, $\pi(p) = \mathcal{V}(p)$.

It is easy to verify that M' is an awareness structure and that, for every $x \in \{GC, BC\}$, if M satisfies x then M' satisfies it as well.

By induction on the structure of φ, we prove that for all $w \in W$, "$(M, w) \models \varphi$ iff $(M', w) \models tr(\varphi)$".

The case $\varphi = p$ and the boolean cases $\varphi = \neg\psi$ and $\varphi = \psi_1 \wedge \psi_2$ are clear. Let us consider the case $\varphi = \triangle_i \alpha$.

($\Rightarrow$) $(M, w) \models \triangle_i \alpha$ means that $\alpha \in \mathcal{D}(i, w)$. By definition of $\mathcal{A}_i$, the latter implies that $tr(\alpha) \in \mathcal{A}_i(w)$ which is equivalent to $(M', w) \models \mathsf{A}_i tr(\alpha)$. Moreover, $(M, w) \models \triangle_i \alpha$ implies that $\mathcal{N}(i, w) \subseteq ||\alpha||_M$. By induction hypothesis, we have $||\alpha||_M = ||tr(\alpha)||_{M'}$. Thus, by definition of $\Rightarrow_i(w)$, it follows that $\Rightarrow_i(w) \subseteq ||tr(\alpha)||_{M'}$. The latter means that $(M', w) \models \mathsf{B}_i tr(\alpha)$. From the latter and $(M', w) \models \mathsf{A}_i tr(\alpha)$, it follows that $(M', w) \models \mathsf{X}_i tr(\alpha)$.

($\Leftarrow$) $(M', w) \models \mathsf{X}_i tr(\alpha)$ implies $(M', w) \models \mathsf{A}_i tr(\alpha)$ which is equivalent to $tr(\alpha) \in \mathcal{A}_i(w)$. By definition of $\mathcal{A}_i$, the latter implies $\alpha \in \mathcal{D}(i, w)$ which is equivalent to $(M, w) \models \triangle_i \alpha$.

Finally, let us consider the case $\varphi = \Box_i \psi$. By induction hypothesis, we have $||\psi||_M = ||tr(\psi)||_{M'}$. $(M, w) \models \Box_i \psi$ means that $\mathcal{N}(i, w) \subseteq ||\psi||_M$. By definition of $\Rightarrow_i(w)$ and $||\psi||_M = ||tr(\psi)||_{M'}$, the latter is equivalent to $\Rightarrow_i(w) \subseteq ||tr(\psi)||_{M'}$ which in turn is equivalent to $(M', w) \models \mathsf{B}_i tr(\psi)$.

Thus, $(M', w) \models tr(\varphi)$, since $(M, w) \models \varphi$.

In order to prove the right-to-left direction, we first prove a weaker result: if $tr(\varphi)$ is satisfiable for the class $\mathbf{AS}_X$, then it is satisfiable for the class $\mathbf{QNDM}_X$.

Let $M = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \pi)$ be an awareness structure. We build the model $M' = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ as follows:

- $W = S$,
- for every $i \in Agt$ and for every $s \in S$, $\mathcal{N}(i, s) = \Rightarrow_i(s)$,
- for every $i \in Agt$ and for every $s \in S$, $\mathcal{D}(i, s) = \{\alpha \in \mathcal{LAN}\mathcal{G}_{\mathsf{LDA}} : \Rightarrow_i(s) \subseteq ||tr(\alpha)||_M$ and $tr(\alpha) \in \mathcal{A}_i(s)\}$,
- for every $p \in Atm$, $\mathcal{V}(p) = \pi(p)$.

Let us prove that M' is a quasi-NDM by showing that it satisfies Condition C1* in Definition 12. To this aim, we first prove by induction on the structure of α that

$||tr(\alpha)||_M = ||\alpha||_{M'}$. The case $\alpha = p$ is clear as well as the boolean cases. Let us prove the case $\alpha = \triangle_i\alpha'$. We have $M', s \models \triangle_i\alpha'$ iff $\alpha' \in \mathcal{D}(i,s)$. By definition of $\mathcal{D}(i,s)$, we have $\alpha' \in \mathcal{D}(i,s)$ iff $\Rightarrow_i(s) \subseteq ||tr(\alpha')||_M$ and $tr(\alpha') \in \mathcal{A}_i(s)$. The latter is equivalent to $M, s \models \mathsf{X}_i tr(\alpha')$.

Suppose that $\alpha \in \mathcal{D}(i,s)$. By definition of $\mathcal{D}(i,s)$, it follows that $\Rightarrow_i(s) \subseteq ||tr(\alpha)||_M$. Thus, since $||tr(\alpha)||_M = ||\alpha||_{M'}$, we have $\Rightarrow_i(s) \subseteq ||\alpha||_{M'}$. Hence, by definition of $\mathcal{N}(i,s)$, $\mathcal{N}(i,s) \subseteq ||\alpha||_{M'}$. This shows that M' satisfies Condition C1*.

It is easy to show that, for every $x \in \{GC, BC\}$, if M satisfies x then M' satisfies it as well.

In the rest of the proof we show that for all $s \in S$, "$(M,s) \models tr(\varphi)$ iff $(M',s) \models \varphi$". The proof is by induction on the structure of φ.

The case $\varphi = p$ and the boolean cases are clear. Let us now consider the case $\varphi = \triangle_i\alpha$. $(M,w) \models tr(\triangle_i\alpha)$ means that $(M,w) \models \mathsf{X}_i tr(\alpha)$. The latter is equivalent to $\Rightarrow_i(s) \subseteq ||tr(\alpha)||_M$ and $tr(\alpha) \in \mathcal{A}_i(s)$. By definition of $\mathcal{D}(i,s)$, the latter is equivalent to $\alpha \in \mathcal{D}(i,s)$ which in turn is equivalent to $(M',w) \models \triangle_i\alpha$.

Let us finally consider the case $\varphi = \Box_i\psi$. By induction hypothesis, we have $||tr(\psi)||_M = ||\psi||_{M'}$. $(M,w) \models tr(\Box_i\psi)$ means that $(M,w) \models \mathsf{B}_i tr(\psi)$ which in turn means that $\Rightarrow_i(w) \subseteq ||tr(\psi)||_M$. By definition of $\mathcal{N}(i,w)$ and $||\psi||_M = ||tr(\psi)||_{M'}$, the latter is equivalent to $\mathcal{N}(i,w) \subseteq ||\psi||_{M'}$ which in turn is equivalent to $(M',w) \models \Box_i\psi$.

Thus, $(M',s) \models \varphi$, since $(M,s) \models tr(\varphi)$.

We have proved that if $tr(\varphi)$ is satisfiable for the class $\mathbf{AS}_X$, then it is satisfiable for the class $\mathbf{QNDM}_X$. From Theorem 1, it follows that if $tr(\varphi)$ is satisfiable for the class $\mathbf{AS}_X$, then it is satisfiable for the class $\mathbf{NDM}_X$. ■

The following theorem is a direct consequence of Theorems 1 and 4.

Theorem 5. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$ and let $X \subseteq \{GC, BC\}$. Then, φ is satisfiable for the class $\mathbf{MAB}_X$ if and only if $tr(\varphi)$ is satisfiable for the class $\mathbf{AS}_X$.*

In [1], it is proved that, for every $X \subseteq \{\text{reflexivity}, \text{transitivity}, \text{Euclideanity}\}$, the satisfiability problem for the logic of general awareness interpreted over awareness structures whose relations $\Rightarrow_i$ satisfy all properties in X is in PSPACE. The proof is based on tableau-based PSPACE satisfiability checking procedures for these logics. It is easy to adapt Ågotnes & Alechina's tableau-based method to show that the logic of general awareness interpreted over awareness structures whose doxastic accessibility relations $\Rightarrow_i$ are serial is also in PSPACE. As a consequence, we can prove the following result.

Theorem 6. *Let $X \subseteq \{GC, BC\}$. Then, checking satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ relative to the class $\mathbf{NDM}_X$ (resp. $\mathbf{MAB}_X$) is a PSPACE-complete problem.*

PROOF. Theorem 4 (resp. Theorem 5) guarantees that the translation tr provides a polynomial-time reduction of the satisfiability problem for formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ relative to the class $\mathbf{NDM}_X$ (resp. $\mathbf{MAB}_X$) to the satisfiability problem of formulas in $\mathcal{LANG}_{\mathsf{LGA}}$ relative to the class $\mathbf{AS}_X$. Since the latter problem is in PSPACE, it

follows that the former problem is also in PSPACE. PSPACE-hardness follows from known PSPACE-hardness results for modal logics K, KT and KD [16, 35]. ■

6. Introspection

The reason why we take introspection as a separate issue and devote an entire section of the article to it is that its implications are debatable and require a careful examination. Since Hintikka's seminal work on the logics of knowledge and belief [45], the issue of introspection for mental states has been widely debated, namely, the question whether an agent should have knowledge or belief about her own mental states (see, e.g., [20, 89]). We do not pretend to enter this debate by providing further arguments in favor or against introspection. The contribution of this section is more modest: we simply show how principles of introspection for explicit and implicit belief can be incorporated into the LDA framework and then define a number of 'introspective' variants of LDA assuming such principles. For some of these variants, we will study complexity of their satisfiability problems.

In a logic of explicit and implicit belief, introspection can mean at least four different things:

- *Positive introspection closure for explicit belief* (PIE): if an agent explicitly believes that φ, then she explicitly believes that she explicitly believes that φ,
- *Negative introspection closure for explicit belief* (NIE): if an agent does not explicitly believe that φ, then she explicitly believes that she does not explicitly believe that φ,
- *Positive introspection closure for implicit belief* (PII): if an agent implicitly believes that φ, then she implicitly believes that she implicitly believes that φ,
- *Negative introspection closure for implicit belief* (NII): if an agent does not implicitly believe that φ, then she implicitly believes that she does not implicitly believe that φ.

On the one hand, by assuming PIE, we impose that an agent's belief base must be either empty or infinite and, by assuming NIE, we impose that an agent's belief base must be infinite. Indeed, suppose an agent satisfies PIE and she explicitly believes that α is true. Then, she must explicitly believe that she explicitly believes that α and, consequently, she must explicitly believe that she explicitly believes that she explicitly believes that α, and so on ad infinitum. Similarly, suppose the agent satisfies NIE and her belief base is finite. Then, since the language $\mathcal{LANG}_0$ is infinite, there exist infinitely many formulas of $\mathcal{LANG}_0$ that are not included in the agent's belief base. Thus, since the agent satisfies NIE, for each of these formulas she must explicitly believe that she does not explicitly believe it. This means that the agent explicitly believes an infinite number of facts which contradicts the initial hypothesis that her belief base is finite. The requirement imposed by PIE and NIE that an agent's belief base is necessarily empty or infinite is clearly too strong. Indeed, realistic human or

artificial agents — such as robots or conversational agents — have a limited amount of information in their belief bases.

On the other hand, PII and NII lead to an interpretation of the notion of implicit belief that is different from the usual one. According to the usual interpretation (see Figure 1 in Section 1), an agent's set of implicit beliefs includes all information that the agent can obtain through deduction from her explicit beliefs and the common ground, if she had enough time and computational resources to do it. By assuming introspection over implicit beliefs, we impose that the set of implicit beliefs has to be closed not only under deduction but also under introspection. This assumption is debatable since it conflates in the notion of implicit belief both the information accessible through deduction and the information accessible through introspection, thereby not separating mental acts of different nature.

In the next section we define PIE, NIE, PII and NII formally and study their relationships.

6.1. Relationships between introspection properties

We assume PIE, NIE, PII and NII to be specific conditions on notional doxastic models (NDMs) of Definition 9. For ease of exposition and in order to avoid redundancies, we do not define corresponding introspection conditions on multi-agent belief models (MABs) of Definition 4. The reason why we focus on NDMs rather than on MABs is that, in the notional model semantics, conditions PII and NII are formulated as transitivity and Euclideanity on accessibility relations. This is the standard way to define positive and negative introspection in epistemic logic.

Definition 14 (Introspection conditions). *Let $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ be a NDM. Then,*

- *M satisfies positive introspection closure for explicit belief (PIE) if and only if, for every $i \in Agt$, for every $w \in W$ and for every $\alpha \in \mathcal{LAN}\mathcal{G}_0$, if $\alpha \in \mathcal{D}(i, w)$ then $\triangle_i \alpha \in \mathcal{D}(i, w)$;*
- *M satisfies negative introspection closure for explicit belief (NIE) if and only if, for every $i \in Agt$, for every $w \in W$ and for every $\alpha \in \mathcal{LAN}\mathcal{G}_0$, if $\alpha \notin \mathcal{D}(i, w)$ then $\neg\triangle_i \alpha \in \mathcal{D}(i, w)$;*
- *M satisfies positive introspection closure for implicit belief (PII) if and only if, for every $i \in Agt$, the relation $\mathcal{N}_i = \{(w, v) \in W \times W : v \in \mathcal{N}(i, w)\}$ is transitive;*
- *M satisfies negative introspection closure for implicit belief (NII) if and only if, for every $i \in Agt$, the relation $\mathcal{N}_i$ is Euclidean.*

We extend the definitions of NDM classes by considering NDMs satisfying some of the previous introspection conditions. For example, $\mathbf{NDM}_{\{GC,PIE,NIE\}}$ is the class of NDMs satisfying global consistency, positive and negative introspection closure for explicit belief.

We have the following validities relative to the different introspection properties:

$$\models_{\mathbf{NDM}_X} \triangle_i\alpha \rightarrow \triangle_i\triangle_i\alpha \text{ if } PIE \in X, \tag{2}$$

$$\models_{\mathbf{NDM}_X} \neg\triangle_i\alpha \rightarrow \triangle_i\neg\triangle_i\alpha \text{ if } NIE \in X, \tag{3}$$

$$\models_{\mathbf{NDM}_X} \Box_i\varphi \rightarrow \Box_i\Box_i\varphi \text{ if } PII \in X, \tag{4}$$

$$\models_{\mathbf{NDM}_X} \neg\Box_i\varphi \rightarrow \Box_i\neg\Box_i\varphi \text{ if } NII \in X. \tag{5}$$

As the following Proposition 4 highlights, PIE is at least as strong as PII and NIE is at least as strong as NII. In other words, positive explicit belief introspection implies positive implicit belief introspection, and negative explicit belief introspection implies negative implicit belief introspection.

Proposition 4. *We have the following relations between classes of NDMs:*

- $\mathbf{NDM}_{\{PIE\}} \subseteq \mathbf{NDM}_{\{PII\}}$,
- $\mathbf{NDM}_{\{NIE\}} \subseteq \mathbf{NDM}_{\{NII\}}$.

PROOF. Let $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ be a NDM.

We prove the first item "PIE implies PII".

Let $w, v, u \in W$. Suppose M satisfies PIE, $w\mathcal{N}_i v$ and $v\mathcal{N}_i u$. We are going to show that $w\mathcal{N}_i u$.

By Condition C1 in Definition 9, $w\mathcal{N}_i v$ means that, for all $\alpha \in \mathcal{D}(i, w)$, $(M, v) \models \alpha$. Thus, since M satisfies PIE, for all $\alpha \in \mathcal{D}(i, w)$, $(M, v) \models \alpha$ and $\triangle_i\alpha \in \mathcal{D}(i, w)$. Hence, for all $\alpha \in \mathcal{D}(i, w)$, $(M, v) \models \alpha$ and $(M, v) \models \triangle_i\alpha$. The latter implies that, for all $\alpha \in \mathcal{D}(i, w)$, $\alpha \in \mathcal{D}(i, v)$. Since $v\mathcal{N}_i u$, we have that, for all $\alpha \in \mathcal{D}(i, v)$, $(M, u) \models \alpha$. Thus, we have $(M, u) \models \alpha$, for all $\alpha \in \mathcal{D}(i, w)$. By Condition C1 in Definition 9, the latter means that $w\mathcal{N}_i u$.

Let us prove the second item "NIE implies NII".

Let $w, v, u \in W$. Suppose M satisfies NIE, $w\mathcal{N}_i v$ and $w\mathcal{N}_i u$. We are going to show that $v\mathcal{N}_i u$.

By Condition C1 in Definition 9, $w\mathcal{N}_i v$ means that, for all $\alpha \in \mathcal{D}(i, w)$, $(M, v) \models \alpha$. Thus, since M satisfies NIE, for all $\alpha \in \mathcal{D}(i, w)$, $(M, v) \models \alpha$ and, for all $\beta \notin \mathcal{D}(i, w)$, $\neg\triangle_i\beta \in \mathcal{D}(i, w)$. Thus, for all $\beta \notin \mathcal{D}(i, w)$, $(M, v) \models \neg\triangle_i\beta$. The latter implies that, for all β, if $\beta \notin \mathcal{D}(i, w)$ then $\beta \notin \mathcal{D}(i, v)$, which is equivalent to $\mathcal{D}(i, v) \subseteq \mathcal{D}(i, w)$. Since $w\mathcal{N}_i u$, by Condition C1 in Definition 9, the latter implies $v\mathcal{N}_i u$. ■

PII and NII together do not necessarily imply PIE and NIE. To see this, consider the single agent NDM $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ such that

$$\begin{aligned}
&W = \{w, v, u\},\\
&\mathcal{D}(1, w) = \{p\}, \mathcal{D}(1, v) = \{q\}, \mathcal{D}(1, u) = \{p, q\},\\
&\mathcal{N}(1, w) = \{v, u\}, \mathcal{N}(1, v) = \{u\}, \mathcal{N}(1, u) = \{u\},\\
&\mathcal{V}^{\leftarrow}(w) = \emptyset, \mathcal{V}^{\leftarrow}(v) = \{p\} \text{ and } \mathcal{V}^{\leftarrow}(u) = \{p, q\}.
\end{aligned}$$

Clearly, M satisfies PII and NII since the relation $\mathcal{N}_1$ is transitive and Euclidean, but it does not satisfy PIE or NIE since agent 1's belief bases at w, v and u are not closed under positive introspection or negative introspection.

This point is highlighted by the following proposition.

Proposition 5. *There exists a NDM M such that $M \in \mathbf{NDM}_{\{PII,NII\}}$ and $M \notin \big(\mathbf{NDM}_{\{PIE\}} \cup \mathbf{NDM}_{\{NIE\}}\big)$.*

Therefore, by Propositions 4 and 5, we have that PIE is stronger than PII and that NIE is stronger than NII.

6.2. Complexity of introspective extensions

In this section, we study the complexity of the variants of LDA under the assumption of introspection closure for implicit belief (PII and NII). The next theorem highlights that adding this assumption to the single-agent variant of LDA reduces its complexity by making it NP-complete. It parallels the complexity results for the modal logics K45, KD45 and S5.

Theorem 7. *Let $X \subseteq \{GC, BC, PII, NII\}$ such that $\{PII, NII\} \subseteq X$ and let $|Agt| = 1$. Then, checking satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ relative to the class $\mathbf{NDM}_X$ is a NP-complete problem.*

PROOF. The satisfiability problem is clearly NP-hard since there exists a polynomial-time reduction of SAT to LDA-satisfiability checking.

In order to prove NP-membership, we first show that if $\{PII, NII\} \subseteq X$, $|Agt| = 1$ and φ is satisfiable for the class $\mathbf{NDM}_X$, then there exists a NDM M in $\mathbf{NDM}_X$ such that M satisfies φ and M has at most $|\varphi|$ worlds, where $|\varphi|$ is the *size* of φ, i.e., the number of symbols used to write it.

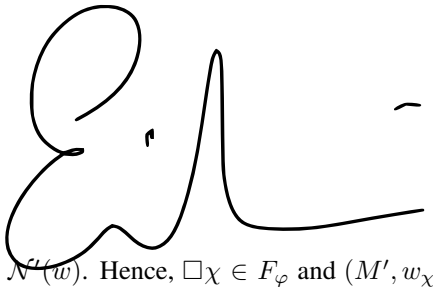
Suppose $|Agt| = 1$ and $\{PII, NII\} \subseteq X$. Let $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V}) \in \mathbf{NDM}_X$ and $w \in W$ such that $(M, w) \models \varphi$. Since $|Agt| = 1$, we can omit the agent argument from the functions $\mathcal{D}$ and $\mathcal{N}$ and from the explicit and implicit belief operators and simply write $\mathcal{D}(w), \mathcal{N}(w), \triangle\alpha$ and $\Box\psi$.

Let $M' = (W', \mathcal{D}', \mathcal{N}', \mathcal{V}')$ such that $W' = \{w\} \cup \mathcal{N}(w)$, $\mathcal{D}' = \mathcal{D}|_{W'}$, $\mathcal{N}' = \mathcal{N}|_{W'}$ and $\mathcal{V}'(p) = \mathcal{V}(p) \cap W'$, for all $p \in Atm$. It is easy to check that $M' \in \mathbf{NDM}_X$. By $\{PII, NII\} \subseteq X$, we have that, for all $v \in W$, $\mathcal{N}(v) = \mathcal{N}(w)$. Therefore, by induction on the structure of φ and the fact that $(M, w) \models \varphi$, it is easy to check that $(M', w) \models \varphi$.

We now distinguish two cases.

Case 1: $\mathcal{N}'(w) = \emptyset$. Clearly, M' has at most one world. Hence, it has at most $|\varphi|$ worlds.

Case 2: $\mathcal{N}'(w) \neq \emptyset$. We adapt the proof of [35, Proposition 6.2] to our case. Let F_φ be the set of subformulas of φ of the form $\Box\psi$ for which $(M', u) \models \neg\Box\psi$ for all $u \in \mathcal{N}'(w)$. Since $\mathcal{N}'(w) \neq \emptyset$, for each formula $\Box\psi \in F_\varphi$, there should be w_ψ in $\mathcal{N}'(w)$ such that $(M', w_\psi) \models \neg\psi$. Let $M'' = (W'', \mathcal{D}'', \mathcal{N}'', \mathcal{V}'')$ such that $W'' = \{w\} \cup \{w_\psi : \Box\psi \in F_\varphi\}$, $\mathcal{D}' = \mathcal{D}|_{W''}$, $\mathcal{N}' = \mathcal{N}|_{W''}$ and $\mathcal{V}'(p) = \mathcal{V}(p) \cap W''$, for all $p \in Atm$. It is easy to check that $M'' \in \mathbf{NDM}_X$. Moreover, M'' has at most $|\varphi|$ worlds since $|F_\varphi| \leq |sub(\varphi)| \leq |\varphi|$. By induction on the structure of φ, it is routine exercise to prove that, for all $v \in W''$ and for all $\psi \in sub(\varphi)$, $(M', v) \models \psi$ iff $(M'', v) \models \psi$. The only non-trivial case is when ψ is of the form $\Box\chi$. We only prove the right-to-left direction, as the left-to-right one is straightforward. Suppose $(M', v) \models \neg\Box\chi$. Since $NII \in X$, we have $(M', v) \models \Box\neg\Box\chi$. Thus, by construction



of M', $(M', u) \models \neg\Box\chi$ for all $u \in \mathcal{N}''(w)$. Hence, $\Box\chi \in F_\varphi$ and $(M', w_\chi) \models \neg\chi$. By construction, $w_\chi \in W''$ and, by induction hypothesis, $(M'', w_\chi) \models \neg\chi$. Since $w_\chi \in \mathcal{N}''(w)$, it follows that $(M'', w) \models \neg\Box\chi$.

Thus, we have $(M'', w) \models \varphi$, since $(M', w) \models \varphi$.

Now, let us prove that if $|Agt| = 1$ and $\{PII, NII\} \subseteq X$ then the satisfiability problem relative to the class $\mathbf{NDM}_X$ is in NP. We have shown that if $|Agt| = 1$ and $\{PII, NII\} \subseteq X$ then every satisfiable formula is satisfiable in a model which is polysize in $|\varphi|$.[6] Here is an non-deterministic algorithm to check if a given formula φ is satisfiable for the class $\mathbf{NDM}_X$:

- guess non-deterministically a NDM $M \in \mathbf{NDM}_X$ whose size is bounded by $|\varphi|$,
- guess non-deterministically a world w of M,
- check whether $(M, w) \models \varphi$.

This algorithm non-deterministically runs in polynomial time. So, if $|Agt| = 1$, then checking satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ relative to a class $\mathbf{NDM}_X$ such that $\{PII, NII\} \subseteq X$ is in NP. ■

We conclude this section with a complexity result for all multi-agent variants of the LDA framework, with and without introspection properties for implicit beliefs. It complements Theorem 6 in Section 5 and the previous Theorem 7.

Theorem 8. *Let $X \subseteq \{GC, BC, PII, NII\}$ and let $|Agt| > 1$. Then, checking satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ relative to the class $\mathbf{NDM}_X$ is a PSPACE-complete problem.*

PROOF. We generalize Theorem 4 by showing that, for every $X \subseteq \{GC, BC, PII, NII\}$, the translation tr given in Section 5 provides a polynomial-time reduction of satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ relative to the class $\mathbf{NDM}_X$ to satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LGA}}$ relative to the class $\mathbf{AS}_X$. From Ågotnes & Alechina [1] we know that the latter problem is in PSPACE if $|Agt| > 1$. PSPACE-hardness follows from known PSPACE-hardness results for multi-agent epistemic logics under the assumption that there is more than one agent in the system [35]. ■

Figure 2 summarizes the complexity results of Theorems 6, 7 and 8. Specifically, it highlights the complexity of checking satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ for each class $\mathbf{NDM}_X$ with $X \subseteq \{GC, BC, PII, NII\}$. Note that the PSPACE-completeness result for $|Agt| = 1$ and $X \subseteq \{GC, BC\}$ is a consequence of Theorem 6 which applies to both cases $|Agt| = 1$ and $|Agt| > 1$.

We leave for future work the analysis of the complexity of the satisfiability problems for the single-agent variants of LDA missing in the table such as, e.g., checking satisfiability relative to the classes $\mathbf{NDM}_{GC,PII}$, $\mathbf{NDM}_{GC,NII}$, $\mathbf{NDM}_{BC,PII}$ and

[6] A model M is said to be polysize in $k \in \mathbb{N}$ if there is polynomial λ such that M contains at most $\lambda(k)$ worlds [16, Definition 6.6].

$\mathbf{NDM}_{BC,NII}$ under the assumption that $|Agt| = 1$. Our conjecture is that all these problems are PSPACE-complete.

		Complexity class	
		NP	**PSPACE**
Number of agents	$\|Agt\| = \mathbf{1}$	If $\{PII, NII\} \subseteq X$	If $X \subseteq \{GC, BC\}$
	$\|Agt\| > \mathbf{1}$	No X	Every X

Figure 2: Complexity results for classes of models $\mathbf{NDM}_X$ such that $X \subseteq \{GC, BC, PII, NII\}$.

7. Dynamic extensions

The aim of this section is to study some extensions of the logic LDA capturing different types of belief dynamics in a multi-agent setting.

The logics in the LDA family allow us to model a rich taxonomy of belief change operations affecting either the agents' common ground or their belief bases. On the one hand, common ground change is typically the result of a public announcement in the sense of public announcement logic (PAL) [68]. On the other hand, belief base change is the result of an agent *privately* perceiving that a certain fact is true or receiving a piece of information from a certain source. As we have emphasized in the introduction, modelling private belief change in standard dynamic epistemic logic (DEL) [32, 83] has a limitation. 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. Thus, the original epistemic model grows exponentially in the length of the sequence of private announcements. As we show in this section, this limitation can be overcome in the context of the LDA framework, by modelling private announcements as operations modifying the belief bases of some agents but not of all agents. This leads to a "parsimonious" account of private informative actions in LDA since, differently from traditional DEL, it does not require to duplicate epistemic models and to make them exponentially larger.

In the rest of this section, we first present the extension of LDA by public announcements. Then, we move to private belief change by presenting an extension of LDA by belief base expansion operators. We illustrate it with the aid of a concrete example from human-machine interaction (HMI). We finally discuss the semantics of further types of private belief change operations including forgetting and belief base contraction.

7.1. Public announcements

In order to represent the effects of a public announcement on the agents' common ground, we extend the language $\mathcal{LANG}_{\mathsf{LDA}}$ by modal operators of the form $[\varphi!]$, thereby obtaining the following language $\mathcal{LANG}_{\mathsf{LDA\text{-}PA}}$:

$$\varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i\varphi \mid [\varphi!]\psi,$$

where α ranges over $\mathcal{LANG}_0$, i ranges over Agt and LDA-PA stands for "Logic of Doxastic Attitudes and Public Announcements".

The formula $[\varphi!]\psi$ has to be read "ψ holds after the public announcement of φ". Following [68], we assume that public announcements are truthful, i.e., an announcement is executable if and only if the formula to be announced is true. Consequently, $\langle\varphi!\rangle\psi$ which abbreviates $\neg[\varphi!]\neg\psi$, has to read "the public announcement of φ is executable and ψ will hold after its occurrence".

We generalize the satisfaction relation between MABs and formulas of Definition 6 to this new type of formulas, as follows:

Definition 15 (Satisfaction relation (cont.)). *Let (B, Cxt) be a MAB. Then:*

$$(B, Cxt) \models [\varphi!]\psi \iff \textit{if } (B, Cxt) \models \varphi \textit{ then } (B, Cxt^{\varphi!}) \models \psi,$$

where

$$Cxt^{\varphi!} = \{B' \in Cxt : (B', Cxt) \models \varphi\}.$$

The idea is that ψ is the consequence of the public announcement of φ if and only if, if φ is true, then ψ is going to be true after restricting the agents' common ground to the states satisfying the announced formula φ. Note that the previous definition would perfectly work, if the formula $[\varphi!]\psi$ was interpreted with respect to the class $\mathbf{MAB}_{\{BC\}}$, as public announcement so defined preserves the necessary and sufficient conditions for the class of multi-agent belief bases satisfying belief correctness given in Proposition 1. In particular, we have that if $(B, Cxt) \models \varphi$ and $(B, Cxt) \in \mathbf{MAB}_{\{BC\}}$ then $B \in Cxt^{\varphi!}$ and $B'\mathcal{R}_iB'$ for every $B' \in Cxt^{\varphi!}$. Therefore, if $(B, Cxt) \models \varphi$ and $(B, Cxt) \in \mathbf{MAB}_{\{BC\}}$ then $(B, Cxt^{\varphi!}) \in \mathbf{MAB}_{\{BC\}}$.

On the contrary, it would not work if the formula $[\varphi!]\psi$ was interpreted with respect to the class $\mathbf{MAB}_{\{GC\}}$. Indeed, restricting the common ground to the φ-situations could empty an agent's set of doxastic alternatives, i.e., it might be the case that $(B, Cxt) \in \mathbf{MAB}_{\{GC\}}$ and $\mathcal{R}_i(B') = \emptyset$ for some $B' \in (\{B\} \cup Cxt^{\varphi!})$. Therefore, $(B, Cxt) \models \varphi$ and $(B, Cxt) \in \mathbf{MAB}_{\{GC\}}$ do not together guarantee that $(B, Cxt^{\varphi!}) \in \mathbf{MAB}_{\{GC\}}$. Consequently, the previous semantics for public announcements is compatible with the logics LDA and $\mathsf{LDA}_{\{\mathbf{T}\Box_i\}}$, while being incompatible with the logic $\mathsf{LDA}_{\{\mathbf{D}\Box_i\}}$. Thus, more generally, our semantics for public announcements is compatible with the model classes $\mathbf{MAB}$ and $\mathbf{MAB}_{\{BC\}}$.

The following proposition provides reduction principles for the dynamic operators $[\varphi!]$.

Proposition 6. *The following formulas are valid relative to every class $\mathbf{MAB}_X$ such that $X \subseteq \{BC\}$.*

$$[\varphi!]p \leftrightarrow (\varphi \rightarrow p)$$
$$[\varphi!]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi!]\psi)$$
$$[\varphi!](\psi_1 \wedge \psi_2) \leftrightarrow ([\varphi!]\psi_1 \wedge [\varphi!]\psi_2)$$
$$[\varphi!]\Box_i\psi \leftrightarrow (\varphi \rightarrow \Box_i[\varphi!]\psi)$$
$$[\varphi!]\triangle_i\alpha \leftrightarrow (\varphi \rightarrow \triangle_i\alpha)$$

PROOF. The first three validities are proved in the same way as in PAL. The last validity is clear since public announcements can modify the common ground but cannot modify the agents' belief bases. Let us prove the fourth validity:

$$
\begin{aligned}
(B, Cxt) \models [\varphi!]\Box_i\psi \quad &\iff \quad \text{if } (B, Cxt) \models \varphi \text{ then } (B, Cxt^{\varphi!}) \models \Box_i\psi, \\
&\iff \quad \text{if } (B, Cxt) \models \varphi \text{ then } \big(\forall B' \in Cxt^{\varphi!} : \text{ if } B\mathcal{R}_iB' \\
&\qquad\quad \text{then } (B', Cxt^{\varphi!}) \models \psi\big), \\
&\iff \quad \text{if } (B, Cxt) \models \varphi \text{ then } \Big(\forall B' \in Cxt : \text{ if } \big(B\mathcal{R}_iB' \text{ and} \\
&\qquad\quad (B', Cxt) \models \varphi\big) \text{ then } (B', Cxt^{\varphi!}) \models \psi\Big), \\
&\iff \quad \text{if } (B, Cxt) \models \varphi \text{ then } \Big(\forall B' \in Cxt : \text{ if } B\mathcal{R}_iB' \text{ then} \\
&\qquad\quad \big(\text{ if } (B', Cxt) \models \varphi \text{ then } (B', Cxt^{\varphi!}) \models \psi\big)\Big), \\
&\iff \quad \text{if } (B, Cxt) \models \varphi \text{ then } \big(\forall B' \in Cxt : \text{ if } B\mathcal{R}_iB' \text{ then} \\
&\qquad\quad (B', Cxt) \models [\varphi!]\psi\big), \\
&\iff \quad \text{if } (B, Cxt) \models \varphi \text{ then } (B, Cxt) \models \Box_i[\varphi!]\psi, \\
&\iff \quad (B, Cxt) \models \varphi \to \Box_i[\varphi!]\psi.
\end{aligned}
$$

■

The first four validities are the standard reduction principles of PAL. The fifth one is the reduction principle for explicit belief. It highlights that an agent's explicit beliefs are not affected by public announcements. Public announcements only operate on the agents' common ground.

The equivalences of Proposition 6 allow to find for every formula of the language $\mathcal{LANG}_{\mathsf{LDA\text{-}PA}}$ an equivalent formula of the language $\mathcal{LANG}_{\mathsf{LDA}}$. Call $red_{\mathsf{LDA\text{-}PA}}$ the mapping which iteratively applies the equivalences of Proposition 6 from the left to the right. It pushes the dynamic operators $[\varphi!]$ inside the formula, and finally eliminates them when facing an atomic formula. Specifically, the mapping $red_{\mathsf{LDA\text{-}PA}}$ is inductively defined as follows:

$$
\begin{aligned}
&1. red_{\mathsf{LDA\text{-}PA}}(p) = p \\
&2. red_{\mathsf{LDA\text{-}PA}}(\triangle_i\alpha) = \triangle_i red_{\mathsf{LDA\text{-}PA}}(\alpha) \\
&3. red_{\mathsf{LDA\text{-}PA}}(\neg\varphi) = \neg red_{\mathsf{LDA\text{-}PA}}(\varphi) \\
&4. red_{\mathsf{LDA\text{-}PA}}(\varphi \wedge \psi) = red_{\mathsf{LDA\text{-}PA}}(\varphi) \wedge red_{\mathsf{LDA\text{-}PA}}(\psi) \\
&5. red_{\mathsf{LDA\text{-}PA}}(\Box_i\varphi) = \Box_i red_{\mathsf{LDA\text{-}PA}}(\varphi) \\
&6. red_{\mathsf{LDA\text{-}PA}}([\varphi!]p) = red_{\mathsf{LDA\text{-}PA}}(\varphi \to p) \\
&7. red_{\mathsf{LDA\text{-}PA}}([\varphi!]\neg\psi) = red_{\mathsf{LDA\text{-}PA}}(\varphi \to \neg[\varphi!]\psi) \\
&8. red_{\mathsf{LDA\text{-}PA}}\big([\varphi!](\psi_1 \wedge \psi_2)\big) = red_{\mathsf{LDA\text{-}PA}}([\varphi!]\psi_1 \wedge [\varphi!]\psi_2) \\
&9. red_{\mathsf{LDA\text{-}PA}}([\varphi!]\Box_i\psi) = red_{\mathsf{LDA\text{-}PA}}(\varphi \to \Box_i[\varphi!]\psi) \\
&10. red_{\mathsf{LDA\text{-}PA}}([\varphi!]\triangle_i\alpha) = red_{\mathsf{LDA\text{-}PA}}(\varphi \to \triangle_i\alpha)
\end{aligned}
$$

We can state the following proposition.

Proposition 7. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA\text{-}PA}}$ and let $X \subseteq \{BC\}$. Then, $\varphi \leftrightarrow red_{\mathsf{LDA\text{-}PA}}(\varphi)$ is valid relative to the class $\mathbf{MAB}_X$.*

PROOF. The proposition is provable by induction on the structure of φ. The cases corresponding to the items 5-10 in the definition of $red_{\mathsf{LDA\text{-}PA}}$ are direct consequences of the validities in Proposition 6. Proving the cases corresponding to the items 1, 3 and 4 in the definition of $red_{\mathsf{LDA\text{-}PA}}$ is a routine exercise. The case corresponding to the item 2 in the definition of $red_{\mathsf{LDA\text{-}PA}}$ relies on the fact that $red_{\mathsf{LDA\text{-}PA}}(\alpha) = \alpha$ for all $\alpha \in \mathcal{LANG}_0$. ■

The fact that checking satisfiability for formulas in $\mathcal{LANG}_{\mathsf{LDA\text{-}PA}}$ is decidable follows from the decidability results for the logics LDA (Theorem 3) and the fact that $red_{\mathsf{LDA\text{-}PA}}$ provides an effective procedure for reducing a formula φ in $\mathcal{LANG}_{\mathsf{LDA\text{-}PA}}$ into an equivalent formula $red_{\mathsf{LDA\text{-}PA}}(\varphi)$ in $\mathcal{LANG}_{\mathsf{LDA}}$.

Theorem 9. *Let $X \subseteq \{BC\}$. Then, checking satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LDA\text{-}PA}}$ relative to the class $\mathbf{MAB}_X$ is decidable.*

Note that the size of the formula $red_{\mathsf{LDA\text{-}PA}}(\varphi)$ may be exponential in the size of φ. We leave for future work the quest for a polynomial-time reduction of the satisfiability problem for formulas in $\mathcal{LANG}_{\mathsf{LDA\text{-}PA}}$ to the satisfiability problem for formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ through an adaptation of the reduction techniques presented in [58]. We also leave for future work the proof-theoretic analysis of the family of logics LDA-PA. We only mention here an interesting difference between standard PAL and LDA-PA. While the following rule of replacement of equivalents (REE) is admissible in PAL

$$\frac{\psi_1 \leftrightarrow \psi_2}{\varphi \leftrightarrow \varphi[\psi_1/\psi_2]}$$

it is not admissible in LDA-PA. To see this, it suffice to observe that p is logically equivalent to $p \wedge (q \vee \neg q)$, but $\triangle_i p$ is not logically equivalent to $\triangle_i\big(p \wedge (q \vee \neg q)\big)$. Note that this is already a difference between standard epistemic logic, in which rule REE is admissible, and the static logic LDA, in which it is not.

7.2. Private belief base expansion

In this section, we present a second dynamic extension of the LDA framework by informative events of private nature that modify an agent's explicit beliefs without modifying the beliefs of the others. Specifically, we extend the language $\mathcal{LANG}_{\mathsf{LDA}}$ by operators of the form $[+_i\alpha]$, where $[+_i\alpha]\varphi$ has to be read "φ holds after agent i has expanded her belief base with α". We obtain the following language $\mathcal{LANG}_{\mathsf{LDA\text{-}PBE}}$:

$$\varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i\varphi \mid [+_i\alpha]\varphi,$$

where α ranges over $\mathcal{LANG}_0$, i ranges over Agt and LDA-PBE stands for "Logic of Doxastic Attitudes and Privated Belief Expansion".

The following definition generalizes the satisfaction relation of Definition 6 to this new family of dynamic operators.

Definition 16 (Satisfaction relation (cont.)). *Let* $B = (B_1, \ldots, B_n, V) \in \mathbf{B}$ *and let* $(B, Cxt) \in \mathbf{MAB}$*. Then:*

$$(B, Cxt) \models [+_i\alpha]\varphi \iff (B^{+_i\alpha}, Cxt) \models \varphi$$

with $B^{+_i\alpha} = (B_1^{+_i\alpha}, \ldots, B_n^{+_i\alpha}, V^{+_i\alpha})$*, where:*

$$V^{+_i\alpha} = V,$$

and for all $j \in Agt$*:*

$$\begin{aligned} B_j^{+_i\alpha} &= B_j \cup \{\alpha\} && \text{if } i = j, \\ B_j^{+_i\alpha} &= B_j && \text{otherwise.} \end{aligned}$$

As the previous definition highlights, the informative event $+_i\alpha$ merely consists in agent i learning that α, thereby expanding her belief base accordingly, while the other agents' belief bases remain unchanged. This is private belief change in the strong sense of "all agents different from i do not know that i has just learnt that α", which is different from the weak sense of "all agents different from i know that i has just learnt something, but they do not know what". Agent i's private belief base expansion indirectly changes agent i's implicit beliefs since, after the operation takes place, agent i's set of doxastic alternatives is recomputed.[7]

Note that the operator $[+_i\alpha]$ cannot be interpreted relative to a class $\mathbf{MAB}_X$ such that $X \cap \{GC, BC\} \neq \emptyset$. Indeed, none of the properties in $\{GC, BC\}$ is necessarily preserved under the operation $+_i\alpha$.

A sufficient condition for preservation of BC under the belief base expansion operation $+_i\alpha$ is that the state resulting from the occurrence of the operation $+_i\alpha$ is compatible with the agents' common ground and that α is deducible from agent i's explicit beliefs before belief base expansion. As for preservation of GC, we only need to assume that α is deducible from agent i's explicit beliefs:

if (B, Cxt) satisfies GC and $(B, Cxt) \models \Box_i\alpha$,
then $(B^{+_i\alpha}, Cxt)$ satisfies GC as well;
if (B, Cxt) satisfies BC, $B^{+_i\alpha} \in Cxt$ and $(B, Cxt) \models \Box_i\alpha$,
then $(B^{+_i\alpha}, Cxt)$ satisfies BC as well.

It is also worth noting that the belief base expansion operation $+_i\alpha$ does not necessarily preserve positive introspection on implicit beliefs, we discussed in Section 6. Indeed, after having expanded her belief base, an agent may start to believe that φ

[7] Private belief base expansion is closely connected to the "consider" operation, the awareness change operation studied by van Benthem & Velázquez-Quesada [76], which consists in extending the set of formulas that an agent is aware of. Given the embedding of LDA into the logic of general awareness, according to which "explicitly believing that α" means "being aware that α and implicitly believing that α", agent i's private belief base change with α is conceivable as the joint execution of the private announcement that α to agent i and agent i's mental act of considering α.

implicitly, without implicitly believing that she implicitly believes that φ. To see this, suppose there is only one agent. Let $B = (B_1, V)$, $B' = (B'_1, V')$, $B'' = (B''_1, V'')$, $Cxt = \{B, B', B''\}$ with $B_1 = B'_1 = B''_1 = \{p\}$, $V = \emptyset$, $V' = \{p\}$ and $V'' = \{p, q\}$. It is easy to check that, in the initial situation, agent 1 has positive introspection on all her implicit beliefs, i.e., for every formula φ, if she implicitly believes that φ, then she implicitly believes that she implicitly believes that φ. Indeed, for every $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$, we have:

$$\text{if } (B, Cxt) \models \Box_1\varphi \text{ then } (B, Cxt) \models \Box_1\Box_1\varphi.$$

This is is due to the fact that $\big(\mathcal{R}_1(B) \cap Cxt\big) = \{B', B''\}$, $\big(\mathcal{R}_1(B') \cap Cxt\big) = \{B', B''\}$, and $\big(\mathcal{R}_1(B'') \cap Cxt\big) = \{B', B''\}$, so that every doxastic alternative from a doxastic alternative from B is also a doxastic alternative from B. Nonetheless, after having expanded her belief base with $\neg q$, agent 1 will implicitly believe that $p \wedge \neg q$ without implicitly believing that she implicitly believes that $p \wedge \neg q$, that is:

$$(B^{+_1\neg q}, Cxt) \models \Box_1(p \wedge \neg q) \wedge \neg\Box_1\Box_1(p \wedge \neg q).$$

In the rest of this section, we focus on the generic class $\mathbf{MAB}$.

The following proposition provides reduction principles for the private belief expansion operators.

Proposition 8. *The following formulas are valid relative to the class* $\mathbf{MAB}$.

$$\begin{array}{l}
[+_i\alpha]p \leftrightarrow p \\
[+_i\alpha]\neg\psi \leftrightarrow \neg[+_i\alpha]\psi \\
[+_i\alpha](\psi_1 \wedge \psi_2) \leftrightarrow ([+_i\alpha]\psi_1 \wedge [+_i\alpha]\psi_2) \\
[+_i\alpha]\Box_j\varphi \leftrightarrow \Box_j\varphi \textit{ if } i \neq j \\
[+_i\alpha]\Box_i\varphi \leftrightarrow \Box_i(\alpha \rightarrow \varphi) \\
[+_i\alpha]\triangle_j\beta \leftrightarrow \triangle_j\beta \textit{ if } i \neq j \textit{ or } \alpha \neq \beta \\
[+_i\alpha]\triangle_i\alpha \leftrightarrow \top
\end{array}$$

Thanks to the equivalences of Proposition 8, for every formula of the language $\mathcal{LANG}_{\mathsf{LDA\text{-}PBE}}$ we can find an equivalent formula of the language $\mathcal{LANG}_{\mathsf{LDA}}$. We define $red_{\mathsf{LDA\text{-}PBE}}$ to be the mapping which iteratively applies the equivalences of Proposition 8 from the left to the right in order to eliminate the dynamic operators $[+_i\alpha]$ from

the formula. Specifically, we define $red_{\mathsf{LDA\text{-}PBE}}$ in an inductive way as follows:

$$
\begin{aligned}
&1.\, red_{\mathsf{LDA\text{-}PBE}}(p) = p\\
&2.\, red_{\mathsf{LDA\text{-}PBE}}(\triangle_j \alpha) = \triangle_j red_{\mathsf{LDA\text{-}PBE}}(\alpha)\\
&3.\, red_{\mathsf{LDA\text{-}PBE}}(\neg\varphi) = \neg red_{\mathsf{LDA\text{-}PBE}}(\varphi)\\
&4.\, red_{\mathsf{LDA\text{-}PBE}}(\varphi \wedge \psi) = red_{\mathsf{LDA\text{-}PBE}}(\varphi) \wedge red_{\mathsf{LDA\text{-}PBE}}(\psi)\\
&5.\, red_{\mathsf{LDA\text{-}PBE}}(\Box_j \varphi) = \Box_j red_{\mathsf{LDA\text{-}PBE}}(\varphi)\\
&6.\, red_{\mathsf{LDA\text{-}PBE}}([+_i\alpha]p) = red_{\mathsf{LDA\text{-}PBE}}(p)\\
&7.\, red_{\mathsf{LDA\text{-}PBE}}([+_i\alpha]\neg\psi) = red_{\mathsf{LDA\text{-}PBE}}(\neg[+_i\alpha]\psi)\\
&8.\, red_{\mathsf{LDA\text{-}PBE}}\big([+_i\alpha](\psi_1 \wedge \psi_2)\big) = red_{\mathsf{LDA\text{-}PBE}}([+_i\alpha]\psi_1 \wedge [+_i\alpha]\psi_2)\\
&9.\, red_{\mathsf{LDA\text{-}PBE}}([+_i\alpha]\Box_j\varphi) = red_{\mathsf{LDA\text{-}PBE}}(\Box_j\varphi) \text{ if } i \neq j\\
&10.\, red_{\mathsf{LDA\text{-}PBE}}([+_i\alpha]\Box_i\varphi) = red_{\mathsf{LDA\text{-}PBE}}\big(\Box_i(\alpha \to \varphi)\big)\\
&11.\, red_{\mathsf{LDA\text{-}PBE}}([+_i\alpha]\triangle_j\beta) = red_{\mathsf{LDA\text{-}PBE}}(\triangle_j\beta) \text{ if } i \neq j \text{ or } \alpha \neq \beta\\
&12.\, red_{\mathsf{LDA\text{-}PBE}}([+_i\alpha]\triangle_i\alpha) = red_{\mathsf{LDA\text{-}PBE}}(\top)
\end{aligned}
$$

The following proposition is provable by induction on the structure of φ in a way similar to Proposition 7.

Proposition 9. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA\text{-}PBE}}$. Then, $\varphi \leftrightarrow red_{\mathsf{LDA\text{-}PBE}}(\varphi)$ is valid relative to the class* $\mathbf{MAB}$.

It is easy to check that the size of the formula $red_{\mathsf{LDA\text{-}PBE}}(\varphi)$ is linear in the size of φ. Therefore, thanks to Theorem 6 in Section 5, we obtain the following complexity result.

Theorem 10. *Checking satisfiability of formulas in $\mathcal{LANG}_{\mathsf{LDA\text{-}PBE}}$ relative to the class* $\mathbf{MAB}$ *is a PSPACE-complete problem.*

Like for the logic LDA-PA, we leave the proof-theoretic analysis of the logic LDA-PBE to future work.

We conclude this section by commenting two modeling issues:

- how the consequences of an agent's inference process can be represented in the logic LDA-PBE, and
- how LDA-PBE can handle semi-private forms of belief change in a multi-agent setting.

As for the first issue, it is interesting to remark that LDA-PBE allows us to elucidate how an agent's inferential action retroacts on the agent's belief base by expanding it with a new piece of information. Specifically, a new family of dynamic operators of type $[infer(i,\alpha)]$ are definable in LDA-PBE, as abbreviations:

$$[infer(i,\alpha)]\varphi \stackrel{\mathtt{def}}{=} \Box_i\alpha \to [+_i\alpha]\varphi.$$

Formula $[infer(i,\alpha)]\varphi$ has to be read "φ holds, after agent i has inferred that α from her belief base". The intuition behind this definition is that inferring α consists in

"actualizing" the implicit belief that α, by expanding the belief base with α. Formula $\Box_i\alpha$ should be conceived as the precondition of agent i's mental action of inferring α, since agent i can infer α if and only if she implicitly believes that α. The following two validities deserve to be mentioned:

$$\models_{\mathbf{MAB}} \Box_i\varphi \rightarrow [\mathit{infer}(i,\alpha)]\Box_i\varphi, \tag{6}$$

$$\models_{\mathbf{MAB}} \neg\Box_i\varphi \rightarrow [\mathit{infer}(i,\alpha)]\neg\Box_i\varphi. \tag{7}$$

They state that, inferring α does not have any influence on an agent's implicit beliefs: if agent i implicitly believes that φ before inferring α, then she will implicitly believe that φ after the inference, and if agent i does not believe that φ implicitly before inferring α, then she will not believe that φ implicitly after the inference. Note that the previous two validities can be recasted as a single validity:

$$\models_{\mathbf{MAB}} \Box_i\alpha \rightarrow (\Box_i\varphi \leftrightarrow [+_i\alpha]\Box_i\varphi). \tag{8}$$

As for the second issue, the idea is rather simple. We model semi-private belief change in LDA-PBE through the joint (or parallel) execution of private expansion operations on the belief bases of multiple agents affecting both the agents' first-order and higher orders beliefs. For example, consider the situation in which two agents 1 and 2 perceives that p is true, 2 observes 1, but 1 does not observe 2. Then, both 1 and 2 will come to explicitly believe that p, 2 will come to explicitly believe that 1 explicitly believes that p, but not vice versa. This situation can be described in LDA-PBE as follows: (i) agent 1 expands her belief base just with p and, in parallel, (ii) agent 2 expands his belief base with both p and the fact that agent 1 explicitly believes that p. Now the question is, how can the joint execution of private belief base expansion operations be modeled? It turns out that in LDA-PBE joint execution can be "simulated" through sequential execution. We are going to show how this can be done.

Let agent i's set of belief base expansion operations (a.k.a. actions) be defined as follows:

$$Act_i = \{+_i\alpha : \alpha \in \mathcal{LAN}\mathcal{G}_0\}.$$

Moreover, let $Act = \bigcup_{i\in Agt} Act_i$. We have the following validity, for all $e_1, \ldots, e_k \in Act$:

$$\models_{\mathbf{MAB}} [e_1]\ldots[e_k]\varphi \leftrightarrow [\sigma(e_1)]\ldots[\sigma(e_k)]\varphi \tag{9}$$

where σ is any permutation of the set $\{e_1, \ldots, e_k\}$. This means that the result of a sequence of private belief base expansion operations is order-independent (i.e., it does not depend on the position of each operation in the sequence).[8] Thus, for every non-empty coalition of agents $J = \{i_1, \ldots, i_k\} \in 2^{Agt*} = \left(2^{Agt} \setminus \{\emptyset\}\right)$, we can safely

[8]Note that this does not hold in general in the context of DEL. For instance, in the logic of private announcements by Gerbrandy & Groeneveld [32], the order of the private announcements in the sequence matters. The reason why the private belief base expansion operations of LDA-PBE commute is due to the fact that they are set-theoretic operations on the agents' belief bases, while private announcements of DEL are operations on the agents' epistemic accessibility relations. More details about the connection between LDA-PBE and DEL will be given in Section 8.

define joint belief base expansion operators of type $[\delta_J]$, as follows:

$$[\delta_J]\varphi \stackrel{\texttt{def}}{=} [\delta_J(i_1)] \ldots [\delta_J(i_k)]\varphi$$

where δ_J is a total function with domain J and codomain Act such that, for every $i \in J$, $\delta(i) \in Act_i$. We call δ_J a *joint belief base expansion operation* for the coalition J. It can also be represented as the tuple $\big(\delta_J(i_1), \ldots, \delta_J(i_k)\big)$, and the abbreviation $[\delta_J]\varphi$ has to read "φ holds after every agent i in J has executed her belief base expansion operation $\delta_J(i)$".

The semi-private form of belief change for agents 1 and 2 discussed above is representable by the joint belief base expansion operation $\delta_{\{1,2\}} = \big(+_1 p, +_2(p \wedge \triangle_1 p)\big)$. We clearly have the following validity:

$$\models_{\mathbf{MAB}} \big[\big(+_1 p, +_2(p \wedge \triangle_1 p)\big)\big]\big(\triangle_1 p \wedge \triangle_2(p \wedge \triangle_1 p)\big). \tag{10}$$

In the next section, we illustrate the expressive power of the logical language $\mathcal{LANG}_{\mathsf{LDA\text{-}PBE}}$ with the aid of a concrete example from human-machine interaction (HMI).

7.3. Example: persuasive agent

We consider a conversational agent that has to interact with a human user in order to support her activity and to take care of her well-being. Specifically, HAL is an artificial companion which takes care of an elderly person called Bob and keeps him company. Bob has to do regular physical activity to be in good health. The problem is that Bob prefers to stay at home watching TV or reading a book rather than to go out for a walk. In this situation, HAL has to play a tutor role: it has to ensure that Bob will do regular physical activity in his interest. To this aim, HAL needs to use its persuasive capabilities in order to induce Bob to adopt a healthy lifestyle. This requires a proper understanding of Bob's beliefs by HAL and, in particular, of the relationship between his beliefs and his actions (i.e., the way Bob's beliefs determine his actions).

It is 3:00 pm and it is less than two hours before the sunset of a winter day. HAL knows that Bob has done no physical activity during the last two days. It decides to recommend to Bob to go out for a walk:

> "Hey Bob! It is a great sunny day. You should take advantage of it and go out for a walk before the end of the day."

Bob replies as follows by expressing discontent:

> "The last time I went out for a walk it was so cold. I did not like it at all. If I am sure that it is not cold outside, then I will follow your advice, otherwise I will not!"

Let us assume that (i) the communication channel between HAL and Bob works well, (ii) HAL (resp. Bob) trusts what Bob (resp. HAL) says, that is, HAL (resp. Bob) believes that if Bob (resp. HAL) provides some information, then this information has to be true, and (iii) both HAL and Bob believe that (i) and (ii). Under the assumptions (i), (ii) and (iii), "agent i's speech act of informing agent j that α", with $i, j \in \{\text{HAL, Bob}\}$

and $i \neq j$, amounts to the operation of making the hearer (agent j) explicitly believe that α and of making the speaker (agent i) explicitly believe that the hearer explicitly believes that α. In formal terms, the latter corresponds to the joint belief base expansion operation $(+_i \triangle_j \alpha, +_j \alpha)$ for the coalition $\{i, j\}$, that we abbreviate as follows:

$$tell_{i,j,\alpha} \stackrel{\texttt{def}}{=} (+_i \triangle_j \alpha, +_j \alpha).$$

Intuitively, $+_i \triangle_j \alpha$ captures the effect of the speech act on the speaker's mind, i.e., what i learns through the performance of the speech act, whereas $+_j \alpha$ represents the effect of the speech act on the hearer's mind, i.e., what j learns through the perception of the speech act performed by i.

We can use the joint belief base expansion operator defined at the end of Section 7.2 to infer that, after Bob's initial speech act, HAL explicitly believes that Bob will intend to go out, if he explicitly believes that it is not cold outside:

$$\models_{\mathbf{MAB}} [tell_{Bob,HAL,(\triangle_{Bob} \neg cold \rightarrow int_{Bob,out})}] \triangle_{HAL} (\triangle_{Bob} \neg cold \rightarrow int_{Bob,out}) \quad (11)$$

where the atomic formulas $cold$ and $int_{Bob,out}$ have to be read, respectively, "it is cold outside" and " Bob has the intention to go out".

Let us further assume that HAL explicitly believes that if Bob explicitly believes that the outside temperature is above 10°C, then he explicitly believes that it is not cold outside. This fact is expressed by the following formula α_0, where the atomic formula $temp_{>10}$ has to be read "the outside temperature is above 10°C":

$$\alpha_0 \stackrel{\texttt{def}}{=} \triangle_{HAL} (\triangle_{Bob} temp_{>10} \rightarrow \triangle_{Bob} \neg cold).$$

On the basis of what it believes about Bob's beliefs, HAL decides to inform him that the outside temperature is acceptable:

> "Bob, you shouldn't worry so much. If you go out, you won't feel cold: the outside temperature is above 10°C."

Again, we can use our language $\mathcal{LANG}_{\mathsf{LDA\text{-}PBE}}$ to infer that, after the occurrence of its last speech act, HAL can conclude that Bob will intend to go out for a walk:

$$\begin{aligned} \models_{\mathbf{MAB}} \alpha_0 \rightarrow & [tell_{Bob,HAL,(\triangle_{Bob} \neg cold \rightarrow int_{Bob,out})}] \\ & [tell_{HAL,Bob,temp_{>10}}] \Box_{HAL} int_{Bob,out}. \end{aligned} \quad (12)$$

Before concluding this section, we would like to clarify the role played by explicit and implicit belief in the example as well as the modeling perspective of LDA-PBE in comparison with the one of DEL.

In our formalization of HAL & Bob's scenario, there is a crucial difference between what HAL explicitly believes about Bob's beliefs and what HAL only implicitly believes about Bob's beliefs and intentions given its explicit beliefs about Bob's beliefs. For instance, after HAL tells to Bob that the outside temperature is above 10°C, HAL explicitly believes that Bob believes that the outside temperature is above 10°C. In other words, the latter belief of HAL is a direct consequence of HAL's speech act,

since HAL does not need to make any inference to form it. On the contrary, in order to form the belief that Bob intends to go out for a walk, HAL has to make an inference from its explicit beliefs about Bob's beliefs. This explains why HAL's belief that Bob intends to go out for a walk is implicit but not explicit. This distinction is relevant for AI modeling, since it allows us to clearly separate the information contained in the artificial agent's database from the information that the artificial agent can infer from it.

LDA and LDA-PBE are specifically designed to account for the relationship between explicit and implicit belief and for belief dynamics induced by belief base change. In dynamic epistemic logic (DEL) there is no such a distinction between explicit and implicit belief, as it only deals with the implicit beliefs of logical omniscient agents. We think that this is a limitation of the DEL approach, compared to LDA-PBE, since in HMI scenarios such as the previous one, we would like to represent the limited reasoning of the human agent, whose explicit beliefs are not necessarily closed under deduction, as well as the unbounded inferential capability of an artificial agent, which is always capable of verifying whether a formula is deducible from its beliefs. Note that the latter capability can be concretely implemented in an artificial agent by exploiting automated reasoning procedures for the logics LDA and LDA-PBE. An initial work in this direction is reported in [57].

Once the inferential capability is implemented in the artificial agent, it can be exploited for plan verification in the domain of epistemic planning [18]. For instance, the artificial agent will be able to verify whether a sequence of speech acts will necessarily induce the human to form a certain belief or intention, like in HAL & Bob's scenario in which HAL can verify whether, after telling to Bob that the outside temperature is above 10°C, Bob intends to go out for a walk.

7.4. Discussion: forgetting and belief base contraction

In the preceding sections we have studied two types of information dynamics at a multi-agent level: common ground change determined by public announcements and private belief base expansion. We here present the semantics of a further type of private belief change operation called *forgetting*.[9]

The idea is simple; we assume that an agent i forgets that α is true if and only if α is removed from her belief base, while all other agents keep their belief bases unchanged. In formal terms, we introduce modal operators of the form $[-_i\alpha]$ where the formula $[-_i\alpha]\varphi$ has to be read "φ holds after agent i has forgotten that α is true". The following is the truth condition of the formula $[-_i\alpha]\varphi$ which is evaluated with respect to a generic MAB (B, Cxt):

$$(B, Cxt) \models [-_i\alpha]\varphi \iff (B^{-_i\alpha}, Cxt) \models \varphi$$

[9] See [85, 31] for logical accounts of forgetting in the standard DEL setting and [10] for a formalization of forgetting in a neighborhood semantics for explicit beliefs.

with $B^{-_i\alpha} = (B_1^{-_i\alpha}, \ldots, B_n^{-_i\alpha}, V)$, and for all $j \in Agt$:

$$B_j^{-_i\alpha} = B_j \setminus \{\alpha\} \qquad \text{if } i = j,$$
$$B_j^{-_i\alpha} = B_j \qquad \text{otherwise.}$$

Note that this semantics for forgetting is compatible with logic $\mathsf{LDA}_{\{\mathbf{D}_{\Box_i}\}}$. In fact, erasing a piece of information from an agent's belief base only increases the agent's uncertainty without making her explicit beliefs inconsistent. More precisely, we have that if $(B, Cxt) \in \mathbf{MAB}_{\{GC\}}$ then $(B^{-_i\alpha}, Cxt) \in \mathbf{MAB}_{\{GC\}}$. On the contrary it is not compatible with the logic $\mathsf{LDA}_{\{\mathbf{T}_{\Box_i}\}}$, as it might be the case that $(B, Cxt) \in \mathbf{MAB}_{\{BC\}}$ and $(B^{-_i\alpha}, Cxt) \notin \mathbf{MAB}_{\{BC\}}$. To see why the operation $-_i\alpha$ does not preserve the property of belief correctness BC suppose that, before the occurrence of $-_1 p$, agent 2 explicitly and correctly believes that agent 1 explicitly believes that p. After the occurrence of $-_1 p$, 1 does not believe anymore that p, but 2 still believes that 1 believes that p, since 1's forgetting is private and does not affect 2's belief base.

We have the following two validities relative to every class of models $\mathbf{MAB}_X$ such that $X \subseteq \{GC\}$:

$$\models_{\mathbf{MAB}_X} [-_i\alpha]\neg\triangle_i\alpha \text{ if } X \subseteq \{GC\}, \tag{13}$$
$$\models_{\mathbf{MAB}_X} \neg\Box_i\varphi \rightarrow [-_i\alpha]\neg\Box_i\varphi \text{ if } X \subseteq \{GC\}. \tag{14}$$

According to the first validity, after having forgot something, an agent does not explicitly believe it anymore. According to the second validity, if an agent forgets something, then she cannot infer anything new that she could not infer before forgetting. We also have the following two validities relative to every class $\mathbf{MAB}_X$ such that $X \subseteq \{GC\}$:

$$\models_{\mathbf{MAB}_X} \triangle_j\beta \leftrightarrow [-_i\alpha]\triangle_j\beta \text{ if } i \neq j \text{ and } X \subseteq \{GC\}, \tag{15}$$
$$\models_{\mathbf{MAB}_X} \Box_j\varphi \leftrightarrow [-_i\alpha]\Box_j\varphi \text{ if } i \neq j \text{ and } X \subseteq \{GC\}. \tag{16}$$

They highlight the private aspect of forgetting: if i and j are different agents then j does not forget or learn anything from i's (privately) forgetting something. The reason why an agent's private forgetting does not affect the implicit beliefs of the others is that, just like private belief base expansion, it changes the belief base of one agent, while keeping the belief bases of the others unchanged. Since the agents' doxastic accessibility relations are built from their belief bases, if the latter do not change, the agents' implicit beliefs do not change either.

As we have emphasized in Section 7.2, the belief base expansion operation $+_i\alpha$ may destroy the global consistency of agent i's belief base. The forgetting operation $-_i\neg\alpha$ is not sufficient to prevent this. To see this, let us extend the language $\mathcal{LANG}_{\mathsf{LDA\text{-}PBE}}$ of Section 7.2 by the forgetting operators to obtain the following language $\mathcal{LANG}_{\mathsf{LDA\text{-}PBEF}}$:

$$\varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i\varphi \mid [+_i\alpha]\varphi \mid [-_i\alpha]\varphi.$$

where LDA-PBEF stands for "Logic of Doxastic Attitudes, Privated Belief Expansion and Forgetting". It is easy to verify that the following formula of the language $\mathcal{LANG}_{\mathsf{LDA\text{-}PBEF}}$ is satisfiable relative to the class of models $\mathbf{MAB}$:

$$\neg\Box_i\bot \wedge [-_i\neg\alpha][+_i\alpha]\Box_i\bot.$$

This means that it is not necessarily the case that if agent i's belief base is globally consistent and i forgets $\neg\alpha$ then, i's belief base will remain globally consistent, after she has expanded her belief base with α. The intuitive explanation of this property is that, after agent i has erased $\neg\alpha$ from her belief base, $\neg\alpha$ could still be deducible from her explicit beliefs. Consequently, if she adds α to her belief base, it could become globally inconsistent.

A way of avoiding this problem consists in introducing a new family of *partial meet contraction* operators in the Hansson's style [37, 40] that guarantee the non-derivability of $\neg\alpha$ from agent i's explicit beliefs, after $\neg\alpha$ was removed from the agent's belief base. The general idea of partial meet contraction is that the belief base resulting from the contraction by $\neg\alpha$ should be equal to the intersection of a selection of the maximal subsets of the initial belief base not entailing $\neg\alpha$.[10] It is worth noting that, by means of such operators and the so-called Levi's identity, we could define a *belief base revision* operator as the composition of partial meet contraction with $\neg\alpha$ followed by belief base expansion with α, as defined in Section 7.2. Again following Hansson [37, 40], we could define *belief base consolidation* — a basic operation aimed at restoring global consistency of an agent's belief base — as *belief base contraction* with $\bot$.

We leave for future work the extension of the LDA framework by operators for belief base contraction and the study of belief base revision and belief base consolidation in the context of this framework. We believe that LDA offers the basis for a minimalistic account of belief revision in a multi-agent setting, inspired by theory of single-agent belief base change [40]. Indeed, modelling belief base revision in LDA does not require a notion of epistemic entrenchment or plausibility ordering over possible worlds as in the traditional DEL account of belief revision [11, 74].

We also leave for future work an analysis in the LDA framework of resource-bounded belief change operations such as local partial meet contraction in which the contraction operation is restricted to a compartment of an agent's belief base [88, 41]. We believe that resource-bounded belief change is relevant for AI applications such as the one discussed in Section 7.3 in which a conversational agent has to interact with a human user who is, by definition, resource-bounded.

8. Comparison with DEL: a closer inspection

In Section 7, we studied some dynamic extensions of LDA including extensions by public announcement, private belief base expansion and contraction. The most widely used logical tool for modeling belief change in a multi-agent setting is DEL. We have already discussed some conceptual differences between the DEL approach to multi-agent belief change and our approach based on LDA. In this section, we make the comparative analysis between the two approaches more precise. We explain in detail how private belief base expansion, as defined in Section 7.2, can be translated into the

[10] Partial meet contraction for belief bases is traditionally opposed to *kernel contraction* as defined in [39] (see, e.g., [21]). In kernel contraction, one has to compute all minimal subsets of the initial belief base that entail $\neg\alpha$ and then remove (in some fashion) an element of each such set, so as to obtain a new belief base that does not entail $\neg\alpha$.

DEL semantics, the latter consisting in update operations on multi-relational Kripke models. Specifically, after having shown how multi-agent belief models (MABs) can be mapped to multi-relational Kripke models (Section 8.1), we introduce in Section 8.2 the general update semantics for DEL based on the notion of arrow update model [50, 51], whereby Kripke models are updated through arrow elimination. The latter slightly differs from the update semantics for DEL based on action (or event) models [12, 77, 81], whereby Kripke models are updated through state elimination.

Finally, in Section 8.3, we show that private belief base expansion of Section 7.2 can be mapped to a specific kind of DEL update of private type based on arrow update models à la Kooi & Renne. Differently from public update by which the beliefs of all agents are changed, private update modifies the beliefs of a single agent, while keeping the beliefs of all other agents unchanged. In particular, we show that the model obtained through private belief base expansion and the model obtained through private update are bisimilar, although the former is more compact than the latter. Indeed, while private update requires world duplication, private belief base expansion does not require it.

8.1. From multi-agent belief models to Kripke models

The standard semantics for epistemic logic (EL) and dynamic epistemic logic (DEL) builds on the class of multi-relational Kripke models or, simply, Kripke models. A Kripke model is a structure in which every agent is identified with her epistemic accessibility relation over states and a valuation function specifies the set of states in which an atomic formula is true. DEL update consists in modifying a Kripke model in different ways, e.g., by eliminating states and/or arrows, changing truth values of atomic formulas, duplicating states.

It is straightforward to provide a semantic interpretation of formulas in the language $\mathcal{LANG}_{\mathsf{LDA}}$ relative to Kripke models of this kind. The idea is to conceive explicit belief formulas of type $\triangle_i\alpha$ as atomic formulas, in the same way as the atomic propositions of type p. Let us show how this can be done.

Let

$$Atm^+ = Atm \cup \{\triangle_i\alpha : i \in Agt \text{ and } \alpha \in \mathcal{LANG}_0\}$$

be the set of atomic formulas. For notational convenience elements of Atm^+ are denoted by $x, y, x', y', \ldots$ We consider Kripke models of the form $K = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \pi)$ where S is a set of states, $\Rightarrow_i \subseteq S \times S$ with $i \in Agt$ is agent i's epistemic accessibility relation and $\pi : Atm^+ \longrightarrow 2^S$ maps atomic formulas to sets of states. A pointed Kripke model is a pair (K, s), where K is a Kripke model and $s \in S$. The class of pointed Kripke models is denoted by $\mathbf{PK}$.

Formulas of the language $\mathcal{LANG}_{\mathsf{LDA}}$ are interpreted relative to pointed Kripke models, as follows:

$$\begin{aligned}
(K,s) \models x &\iff s \in \pi(x) \text{ for } x \in Atm^+,\\
(K,s) \models \neg\varphi &\iff (K,s) \not\models \varphi,\\
(K,s) \models \varphi \wedge \psi &\iff (K,s) \models \varphi \text{ and } (K,s) \models \psi,\\
(K,s) \models \Box_i\varphi &\iff \forall s' \in S: \text{ if } s \Rightarrow_i s' \text{ then } (K,s') \models \varphi,
\end{aligned}$$

where $(K, s) \models \varphi$ means that φ is true at s in K.

There is a natural way to define a functor

$$\mathit{transf} : \mathbf{MAB} \longrightarrow \mathbf{PK}$$

which transforms a multi-agent belief model into a corresponding pointed Kripke model satisfying the same formulas as the original model.[11] A similar transformation is used in the proof of Lemma 7 in the technical annex at the end of the paper, to construct the quasi-NDM corresponding to a given multi-agent belief model. It is formally defined as follows. For all $(B, \mathit{Cxt}) \in \mathbf{MAB}$ and $(K, s) \in \mathbf{PK}$ we have $\mathit{transf}\big((B, \mathit{Cxt})\big) = (K, s)$ with $K = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \pi)$ if and only if:

- $S = \big\{s_{B'} : B' \in \{B\} \cup \mathit{Cxt}\big\}$,
- for every $i \in \mathit{Agt}$,

$$\Rightarrow_i = \big\{(s_{B'}, s_{B''}) \in S \times S : B' \in \mathit{Cxt} \cup \{B\}, B'' \in \mathit{Cxt} \text{ and } B'\mathcal{R}_i B''\big\},$$

- for every $p \in \mathit{Atm}$, $\pi(p) = \{s_{B'} \in S : p \in V'\}$,
- for every $\triangle_i \alpha \in \mathcal{LANG}_0$, $\pi(\triangle_i \alpha) = \{s_{B'} \in S : \alpha \in B'_i\}$, and
- $s = s_B$,

where the binary relation $\mathcal{R}_i$ is defined as in Definition 5.

The idea of connecting different classes of models used to represent epistemic information through functors is borrowed from van Benthem [75], in which the idea of tracking epistemic information from one type of model to another is formally investigated.[12]

It is easy to show that (B, Cxt) and $\mathit{transf}\big((B, \mathit{Cxt})\big)$ satisfy the same formulas of the language $\mathcal{LANG}_{\mathsf{LDA}}$. In other words, information provided at the level of the multi-agent belief model is tracked at the level of the Kripke model.

Proposition 10. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$. Then,*

$$(B, \mathit{Cxt}) \models \varphi \textit{ if and only if } \mathit{transf}\big((B, \mathit{Cxt})\big) \models \varphi.$$

In the next section, we provide a concise presentation of the arrow-eliminating version of the update semantics for DEL.

[11] The term "functor" is used here in a loose sense. A category-theoretical analysis of the relationship between multi-agent belief models and Kripke models is beyond the scope of this paper.

[12] Van Benthem studies, among the others, the connection between standard epistemic models and plausibility models [11] as well as the connection between plausibility models and evidence models [78].

8.2. Product update

Following Kooi & Renne [50], we define the following concept of arrow update model. It shall be conceived as an event which is responsible for updating the agents' beliefs (epistemic change) and/or the truth values of the atomic formulas (factual change).

Definition 17 (Arrow update model). *An arrow update model is a pair $U = (O, \tau, post)$ consisting of a finite nonempty set of outcomes O, a partial function*

$$\tau : Agt \times O \times O \rightharpoonup \mathcal{LANG}_{\mathsf{LDA}} \times \mathcal{LANG}_{\mathsf{LDA}},$$

called the arrow function, and a postcondition function

$$post : O \times Atm^{+} \longrightarrow \mathcal{LANG}_{\mathsf{LDA}}.$$

A pointed arrow update model is a pair (U, o) where $U = (O, \tau, post)$ is an arrow update model and $o \in O$.

Our definition of arrow update model slightly differs from Kooi & Renne (K&R)'s original definition, as it includes the postcondition function which is not included in K&R's definition. Note that the postcondition function is usually integrated in the definition of action model in order to model factual change in DEL, in opposition to epistemic change (see, e.g., [77, 81]). Another minor difference with Kooi & Renne's definition concerns the arrow function τ that they specify as a total function with domain $Agt \times O$ and codomain $\mathcal{LANG}_{\mathsf{LDA}} \times O \times \mathcal{LANG}_{\mathsf{LDA}}$. However, this difference has no implication, since the partiality of our arrow function can be 'simulated' in K&R by the fact that, possibly, $\tau(i, o) = (\bot, o', \bot)$ for some $o, o' \in O$ and $i \in Agt$.

For notational convenience, for every $i \in Agt$ and for every $o', o'' \in O$, we write $\tau_1(i, o', o'')$ to denote the formula φ such that $\tau(i, o', o'') = (\varphi, \psi)$ for some ψ, and $\tau_2(i, o', o'')$ to denote the formula ψ such that $\tau(i, o', o'') = (\varphi, \psi)$ for some φ.

A pointed arrow update model (U, o) is applied to a pointed Kripke model (K, s) to generate a new pointed Kripke model, called general product update, which is defined as follows.

Definition 18 (Product update). *Let (U, o) be a pointed arrow update model with $U = (O, \tau, post)$ and let (K, s) be a pointed Kripke model with $K = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \pi)$. The product update of (U, o) and (K, s) is the pointed Kripke model $\big(K \otimes U, (s, o)\big)$ with $K \otimes U = (S', \Rightarrow'_1, \ldots, \Rightarrow'_n, \pi')$ such that:*

- *$S' = S \times O$,*
- *for every $i \in Agt$ and for every $(s', o') \in S'$,*

$$\Rightarrow'_i\big((s', o')\big) = \big\{(s'', o'') \in S' : s' \Rightarrow_i s'', \tau(i, o', o'') \text{ is defined}, (K, s') \models \tau_1(i, o', o'') \text{ and } (K, s'') \models \tau_2(i, o', o'')\big\},$$

- *for every $x \in Atm^{+}$, $\pi'(x) = \big\{(s', o') \in S' : (K, s') \models post(o', x)\big\}$.*

General product update increases the size of the initial model by size-change factor $|O|$. Indeed, for every outcome o' in O and for every state s' in the Kripke model to be updated, an o'-indexed copy (s', o') of s' is generated. Moreover, for every agent $i \in Agt$ and for every pair of outcomes (o', o''), the arrow update model specifies the *source condition* $\tau_1(i, o', o'')$ that has to be satisfied by a state s' and the *target condition* $\tau_2(i, o', o'')$ that has to be satisfied by another state s'', to guarantee that in the updated Kripke model there will be a i-arrow from (s', o') to (s'', o''). Intuitively, this means that an agent's uncertainty between states in the original Kripke model is carried over to outcome-indexed copies of those states if and only if the original states satisfy the source and target conditions for the corresponding outcomes. The postcondition function $post$ is responsible for changing the truth values of atomic formulas. In particular, it specifies the formula $post(o', x)$ that must be satisfied by a state s' to make the atomic formula x true at its outcome-indexed copy (s', o').

From the postcondition function $post$, we can identify the set of assignments occurring at a given outcome $o' \in O$, denoted by

$$Assign(o') = \{x \hookrightarrow \psi : x \in Atm^+, \psi \in \mathcal{LANG}_{\mathsf{LDA}} \text{ and } post(o', x) = \psi\}.$$

Following van Ditmarsch et al. [84], we call $x \hookrightarrow \psi$ an assignment since it corresponds to the operation of assigning the value of ψ to the value of x.

8.3. Private belief expansion as a compact version of private product update

In this section, we show how the operation of private belief expansion defined in Section 7.2 can be mapped to a specific kind of product update of private type in which the beliefs of a single agent are updated, while the beliefs of all other agents are kept unchanged. We call it *private arrow update model*. It consists of two outcomes o_1 and o_2. Outcome o_1 is the outcome at which agent i shrinks her set of doxastic alternatives to φ-states and at which assignment $x \hookrightarrow \psi$ takes place. Outcome o_2 is the outcome at which nothing happens.

Definition 19 (Private arrow update model). *The private arrow update model for agent i by information φ and assignment $x \hookrightarrow \psi$ is the arrow update model $U^{(\varphi, x \hookrightarrow \psi)_i} = (O, \tau, post)$ such that:*

- $O = \{o_1, o_2\}$*;*
- *for every $k, h \in \{1, 2\}$ and for every $j \in Agt$, $o(j, o_k, o_h)$ is defined if and only if $(k = 1$ and $h = 2)$ or $k = h = 2$;*
- $o(i, o_1, o_2) = (\top, \varphi)$ *and* $o(i, o_2, o_2) = (\top, \top)$*;*
- $o(j, o_1, o_2) = (\top, \top)$ *and* $o(j, o_2, o_2) = (\top, \top)$ *for every $j \neq i$;*
- *for every $y \in Atm^+$ and for every $k \in \{1, 2\}$:*

$$post(o_k, y) = \begin{cases} y & \text{if } k = 2 \text{ or } x \neq y, \\ \psi & \text{if } k = 1 \text{ and } x = y. \end{cases}$$

Let us see what the model obtained by product update with the private arrow update model of Definition 19 looks like. Let (K, s) be a pointed Kripke model with $K = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \pi)$ and $s \in S$. $\big(K \otimes U^{(\varphi, x \hookrightarrow \psi)_i}, (s, o_1)\big)$ is the product update of (K, s) and the pointed private arrow update model $(U^{(\varphi, x \hookrightarrow \psi)_i}, o_1)$, where $K \otimes U^{(\varphi, x \hookrightarrow \psi)_i}$ is the Kripke model $(S', \Rightarrow'_1, \ldots, \Rightarrow'_n, \pi')$ which is defined as follows:

$$S' = \{(s', o_1) : s' \in S\} \cup \{(s', o_2) : s' \in S\},$$

for all $s' \in S$ and for all $j \in Agt$:

$$\begin{aligned}
\Rightarrow'_j ((s', o_1)) &= \big\{(s'', o_2) : s' \Rightarrow_j s'' \text{ and } (M, s'') \models \varphi\big\} \text{ if } i = j,\\
\Rightarrow'_j ((s', o_1)) &= \big\{(s'', o_2) : s' \Rightarrow_j s''\big\} \text{ if } i \neq j,\\
\Rightarrow'_j ((s', o_2)) &= \big\{(s'', o_2) : s' \Rightarrow_j s''\big\},
\end{aligned}$$

and for all $y \in Atm^+$:

$$\pi'(y) = \begin{cases}
\big\{(s', o_k) \in S' : k = 1 \text{ and } (K, s') \models \psi\big\} \cup & \\
\big\{(s', o_k) \in S' : k = 2 \text{ and } (K, s') \models y\big\} & \text{if } x = y,\\
\big\{(s', o_k) \in S' : k \in \{1, 2\} \text{ and } (K, s') \models y\big\} & \text{if } x \neq y.
\end{cases}$$

We call $\big(K \otimes U^{(\varphi, x \hookrightarrow \psi)_i}(s, o_1)\big)$ *private update* of (K, s) for agent i by information φ and assignment $x \hookrightarrow \psi$. To make notation more compact, in what follows, we write $(K, s)^{(\varphi, x \hookrightarrow \psi)_i}$ instead of $\big(K \otimes U^{(\varphi, x \hookrightarrow \psi)_i}(s, o_1)\big)$.

In order to formally define the correspondence between private belief base expansion and private update, we exploit the notion of bisimulation. Let us recall its definition.

A bisimulation between two Kripke models $K = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \pi)$ and $K' = (S', \Rightarrow'_1, \ldots, \Rightarrow'_n, \pi')$ is a nonempty binary relation $\leftrightarroweq \subseteq S \times S'$ such that whenever $s \leftrightarroweq s'$ for $s \in S$ and $s' \in S$ we have that:

Atomic harmony: for every $x \in Atm^+$, $s \in \pi(x)$ if and only if $s' \in \pi'(x)$,

Zig: for every $s'' \in S$, if $s \Rightarrow_i s''$, then there exists $s''' \in S'$ such that $s' \Rightarrow'_i s'''$ and $s'' \leftrightarroweq s'''$,

Zag: for every $s'' \in S'$, if $s' \Rightarrow'_i s''$, then there exists $s''' \in S$ such that $s \Rightarrow_i s'''$ and $s''' \leftrightarroweq s''$.

Two pointed Kripke models (K, s) and (K', s') with $K = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \pi)$ and $K' = (S', \Rightarrow'_1, \ldots, \Rightarrow'_n, \pi')$ are said to be bisimilar if there exists a bisimulation $\leftrightarroweq \subseteq S \times S'$ such that $s \leftrightarroweq s'$.

As a consequence of [16, Theorem 2.20], we have that if (K, s) and (K', s') are bisimilar then they satisfy the same formulas in $\mathcal{LANG}_{\mathsf{LDA}}$.

The following theorem is the core result of this section.

Theorem 11. *The pointed Kripke model* $transf\big((B^{+_i\alpha}, Cxt)\big)$ *and the pointed Kripke model* $transf\big((B, Cxt)\big)^{(\alpha,\triangle_i\alpha\hookrightarrow\top)_i}$ *are bisimilar.*

PROOF. Let $transf\big((B, Cxt)\big) = (K, s)$ with $K = (S, \Rightarrow_1, \ldots, \Rightarrow_n, \pi)$ and $s \in S$. Furthemore, let $transf\big((B^{+_i\alpha}, Cxt)\big) = (K', s')$ with $K' = (S', \Rightarrow'_1, \ldots, \Rightarrow'_n, \pi')$ and $s' \in S'$. Finally, let $(K, s)^{(\alpha,\triangle_i\alpha\hookrightarrow\top)_i} = (K'', s'')$ with $K'' = (S'', \Rightarrow''_1, \ldots, \Rightarrow''_n, \pi'')$.

We recall that:

$$S' = \big\{s_{B'} : B' \in Cxt \cup \{B^{+_i\alpha}\}\big\}, \text{ and}$$
$$S'' = \big\{(s_{B'}, o_1) : B' \in Cxt \cup \{B\}\big\} \cup \big\{(s_{B'}, o_2) : B' \in Cxt \cup \{B\}\big\}.$$

Moreover, $s = s_B$, $s' = s_{B^{+_i\alpha}}$ and $s'' = (s_B, o_1)$.

We define the binary relation $\leftrightarrows \subseteq S' \times S''$, as follows:

$$\leftrightarrows(s_{B^{+_i\alpha}}) = \{(s_B, o_1)\},$$
$$\leftrightarrows(s_{B'}) = \{(s_{B'}, o_2)\} \text{ for every } B' \neq B^{+_i\alpha}.$$

It is routine exercise to show that $\leftrightarrows$ defines a bisimulation between K' and K'' and, consequently, (K', s') and $\big(K'', (s, o_1)\big)$ are bisimilar. ■

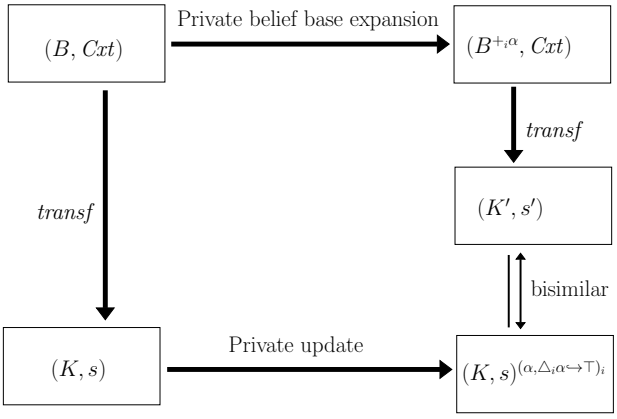


Figure 3: Connection between private belief base expansion and private update

Theorem 11 highlights that the Kripke model obtained by applying private belief base expansion with α to agent i first, and then the transformation functor $transf$, is bisimilar with the Kripke model obtained by applying the transformation functor $transf$ first, and then private update for i by information α and assignment $\triangle_i\alpha \hookrightarrow \top$. This result is summarized in Figure 3. From Theorem 11, it follows that, for every $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$,

$$transf\big((B^{+_i\alpha}, Cxt)\big) \models \varphi \text{ if and only if } transf\big((B, Cxt)\big)^{(\alpha,\triangle_i\alpha\hookrightarrow\top)_i} \models \varphi.$$

Observe that product update by the private arrow update model of Definition 19 duplicates the size of the original Kripke model, since, for every state s', two copies (s', o_1) and (s', o_2) are generated. Consequently, the size of the original Kripke model increases exponentially in the length of the sequence of private update operations.[13] On the contrary, private belief base expansion as defined in Section 7.2, simply requires to add a piece of information to the local belief base of a single agent, while keeping the belief bases of all other agents unchanged. Consequently, the size of the original multi-agent belief model increases at most linearly in the length of the sequence of private belief base expansion operations.

In the light of the previous observation and of Theorem 11, we can safely conclude that the dynamic version of the logic LDA offers a parsimonious and practical approach to modeling private belief change in a multi-agent setting, compared to the DEL approach. It is worth noting that dynamic extensions of LDA studied above also have an advantage compared to dynamic extensions of F&H's logic of general awareness, such as the ones studied in [82, 80]. The semantics of F&H's logic extends the Kripkean semantics of epistemic logic by the syntactic notion of awareness. In this logic, explicit belief change is derivative of awareness and implicit belief change, i.e., an explicit belief may change as a consequence of either awareness change or implicit belief change, since explicit belief is defined in terms of implicit belief and awareness. Therefore, modeling private (explicit or implicit) belief change in F&H's logic requires to exponentially increase the size of the initial model in the length of the sequence of private update operations. Let us emphasize again that this drawback is not encountered in the LDA framework.

Before concluding, a mention must be made to the fact that, although modeling private belief change in LDA is simpler than modeling it in DEL, there are forms of DEL update such as semi-private announcements about generic formulas in $\mathcal{LANG}_{\mathsf{LDA}}$ that do not seem to be representable in the LDA framework.

9. Perspectives and conclusion

We have presented a family of logics of explicit and implicit belief with a semantics based on belief bases. We have explored several aspects of these logics both at the conceptual level and at the mathematical and computational level. This includes the distinction between belief and knowledge, the connection with the logic of general awareness, different dimensions of introspection for explicit and implicit beliefs, dynamic aspects related to belief change, axiomatizability of these logics, complexities of their satisfiability problems and comparison with the DEL approach to multi-agent belief change. We have also illustrated the application potential of the LDA framework with the aid of a concrete example taken from the domain of conversational agents.

[13] Indeed, let $x_0, \ldots, x_k \in Atm^+$, $i_0, \ldots, i_k \in Agt$ and $\psi_0, \ldots, \psi_k \in \mathcal{LANG}_{\mathsf{LDA}}$. Moreover, let $K_0 = (S_0, \Rightarrow_1, \ldots, \Rightarrow_n, \pi)$ with $s_0 \in S_0$ and, for every $k \geq 1$, let:

$$(K_k, s_k) = (K_{k-1}, s_{k-1})^{(\alpha_{k-1}, x_{k-1} \hookrightarrow \psi_{k-1})_{i_{k-1}}}$$

with $K_k = (S_k, \Rightarrow'_1, \ldots, \Rightarrow'_n, \pi')$. Then, $|S_k| = |S_0| \times 2^k$.

Directions of future research are manifold. Some of them were already mentioned in the previous sections. Among them, we plan to study the proof-theoretic aspects of the introspective variants of LDA, we discussed in Section 6. Moreover, we intend to investigate the proof-theoretic and complexity aspects for the dynamic extension of LDA by the forgetting operators $[-_i\alpha]$. Finally, we plan to extend LDA by operators for private belief base contraction and revision, along the lines we briefly discussed in Section 7.4. Before concluding, we mention further perspectives of future research.

Restoring introspection. As we observed in Section 7.2, private belief base expansion does not necessarily preserve positive introspection on implicit beliefs. This is due to the "locality" of the private belief base expansion operation. Private belief base expansion updates an agent's actual belief base without updating the states that the agent considers possible from the actual state. Consequently, if the agent expands her belief base with α, she will start to believe that α, without necessarily believing that she believes that α. This is a natural property of belief expansion for *non-introspective agents* whose beliefs are not necessarily closed under positive introspection.

The previous observation naturally raises the following problem: how can we restore positive introspection on implicit beliefs after the occurrence of a belief base expansion operation? Solving this problem is important if we want to model belief base expansion for an *introspective agent*, who does not fail to form the higher-order belief that she believes that α, when forming the belief that α.

A solution to this problem relies on the notion of *introspective discernment* (ID). The fact that an agent has introspective discernment means that all her doxastic alternatives are for her subjectively-equivalent to the actual state, where two states are 'subjectively-equivalent' for an agent if the agent has the same belief base at the two states. In other words, an introspectively discerning agent cannot consider possible a situation at which her belief base differs from her actual belief base. Therefore, (i) if an introspectively discerning agent explicitly believes that α, then she implicitly believes that she explicitly believes that α, and (ii) if she does not believe that α explicitly, then she implicitly believes that she does not believe that α explicitly.

Introspective discernment guarantees recovery of an agent's introspective access to her implicit beliefs, after the occurrence of a belief base expansion operation. Indeed, after having expanded her belief base with α, an introspectively discerning agent "recomputes" her doxastic accessibility relation by shrinking her set of doxastic alternatives to the states in which α holds and, then, by closing it under positive and negative introspection.[14]

Let us sketch a formal analysis of this notion of introspective discernment. The first thing to be defined is the doxastic accessibility relation for an introspectively discerning agent i, denoted by $\mathcal{R}_i^{ID}$. Specifically, $\mathcal{R}_i^{ID}$ is a binary relation on the set of multi-agent belief bases $\mathbf{B}$ such that, for all $B = (B_1, \ldots, B_n, V), B' = (B'_1, \ldots, B'_n, V') \in$

[14]This corresponds to the operation of taking the minimal expansion that includes α and is closed under positive and negative introspection.

B:

$$
\begin{array}{l}
B\mathcal{R}_i^{ID}B' \text{ if and only if} \\
(i)\ \forall \alpha \in B_i : B' \models \alpha, \text{ and} \\
(ii)\ B_i = B'_i.
\end{array}
$$

$B\mathcal{R}_i^{ID}B'$ means that B' is a doxastic alternative for the introspectively discerning agent i at B. According to the previous definition, if agent i is introspectively discerning, then i's set of doxastic alternatives at B (i.e., $\mathcal{R}_i^{ID}(B)$) includes all and only those states that satisfy i's explicit beliefs, and that are for i subjectively equivalent to B. Note that $\mathcal{R}_i^{ID} \subseteq \mathcal{R}_i$ since item (i) is exactly the definition of the relation $\mathcal{R}_i$ (Definition 5 in Section 3.1).

By means of the new relation $\mathcal{R}_i^{ID}$, we can define a variant of the language $\mathcal{LANG}_{\mathsf{LDA}}$ in which every implicit belief operator $\Box_i$ is replaced by an introspectively discerning variant of it of the form $\Box_i^{ID}$, where $\Box_i^{ID}\varphi$ has to be read "if agent i was introspectively discerning, she would implicitly believe that φ" (or "agent i implicitly believes that φ, under the assumption that she is introspectively discerning"). Likewise the operator $\Box_i$, the operator $\Box_i^{ID}$ is interpreted relative to a MAB (B, Cxt), as follows:

$$(B, Cxt) \models \Box_i^{ID}\varphi \iff \forall B' \in Cxt : \text{if } B\mathcal{R}_i^{ID}B' \text{ then } (B', Cxt) \models \varphi.$$

Notions of satisfiability and validity for this variant of the language $\mathcal{LANG}_{\mathsf{LDA}}$ in which operators $\Box_i$ are replaced by operators $\Box_i^{ID}$ are defined in the usual way.

Intuitively speaking, the operator $\Box_i^{ID}$ should be understood as describing the body of information that agent i can form either through deduction from her belief base and the common ground or through introspection on what she explicitly believes and on what she does not believe explicitly.

As the following proposition indicates, if an agent is introspectively discerning, then her implicit beliefs are closed under positive and negative introspection.

Proposition 11. *Let $i \in Agt$. Then, the relation $\mathcal{R}_i^{ID}$ is transitive and Euclidean.*

PROOF. We first prove transitivity. Suppose $B\mathcal{R}_i^{ID}B'$ and $B'\mathcal{R}_i^{ID}B''$. The latter implies that $B' \models \alpha$ for all $\alpha \in B_i$, $B'' \models \alpha$ for all $\alpha \in B'_i$ and $B_i = B'_i = B''_i$. Hence, $B'' \models \alpha$ for all $\alpha \in B_i$, since $B_i = B'_i$. It follows that $B\mathcal{R}_i^{ID}B''$.

Let us prove that $\mathcal{R}_i^{ID}$ is Euclidean. Suppose $B\mathcal{R}_i^{ID}B'$ and $B\mathcal{R}_i^{ID}B''$. The latter implies that $B' \models \alpha$ and $B'' \models \alpha$ for all $\alpha \in B_i$, and $B_i = B'_i = B''_i$. Hence, $B'' \models \alpha$ for all $\alpha \in B'_i$, since $B_i = B'_i$. It follows that $B'\mathcal{R}_i^{ID}B''$. ■

The following four validities capture some fundamental properties of this operator:

$$\models_{\mathbf{MAB}} \triangle_i\alpha \rightarrow \Box_i^{ID}\triangle_i\alpha, \qquad (\mathbf{PID}_{\triangle_i})$$

$$\models_{\mathbf{MAB}} \neg\triangle_i\alpha \rightarrow \Box_i^{ID}\neg\triangle_i\alpha, \qquad (\mathbf{NID}_{\triangle_i})$$

$$\models_{\mathbf{MAB}} \Box_i^{ID}\varphi \rightarrow \Box_i^{ID}\Box_i^{ID}\varphi, \qquad (\mathbf{PID}_{\Box_i^{ID}})$$

$$\models_{\mathbf{MAB}} \neg\Box_i^{ID}\varphi \rightarrow \Box_i^{ID}\neg\Box_i^{ID}\varphi. \qquad (\mathbf{NID}_{\Box_i^{ID}})$$

Principles $\mathbf{PID}_{\triangle_i}$ and $\mathbf{NID}_{\triangle_i}$ are direct consequences of the definition of the accessibility relation $\mathcal{R}_i^{ID}$ and of the interpretation of the operator $\Box_i^{ID}$, whereas principles $\mathbf{PID}_{\Box_i^{ID}}$ and $\mathbf{NID}_{\Box_i^{ID}}$ are direct consequences of Proposition 11 and of the interpretation of the operator $\Box_i^{ID}$. They capture, respectively, positive introspective discernment on explicit beliefs (principle $\mathbf{PID}_{\triangle_i}$), negative introspective discernment on explicit beliefs (principle $\mathbf{NID}_{\triangle_i}$), positive introspective discernment on implicit beliefs (principle $\mathbf{PID}_{\Box_i^{ID}}$) and negative introspective discernment on implicit beliefs (principle $\mathbf{NID}_{\Box_i^{ID}}$). $\mathbf{PID}_{\Box_i^{ID}}$ and $\mathbf{NID}_{\Box_i^{ID}}$ are nothing but the modal logic Axioms 4 and 5 for the $\Box_i^{ID}$-operator.

Likewise the operator $\Box_i$, if the operator $\Box_i^{ID}$ is interpreted relative to the model class $\mathbf{MAB}_{\{BC\}}$, it satisfies the "knowledge implies truth" principle:

$$\models_{\mathbf{MAB}_{\{BC\}}} \Box_i^{ID}\varphi \rightarrow \varphi. \qquad (\mathbf{T}_{\Box_i^{ID}})$$

Indeed, a MAB (B, Cxt) satisfies belief correctness if and only if $B \in Cxt$ and, for every $i \in Agt$ and for every $B' \in Cxt$, $B'\mathcal{R}_i^{ID}B'$.

Therefore, $\Box_i^{ID}$ becomes a S5-modality in the context of the model class $\mathbf{MAB}_{\{BC\}}$.

The interesting aspect of the operator $\Box_i^{ID}$ is its behavior in the presence of the private belief base expansion operator $[+_i\alpha]$, whose interpretation was given in Section 7.2 (Definition 16).

As the following validities highlight, an introspectively discerning agent will keep positive and negative introspection on her implicit beliefs, when performing a belief base expansion:

$$\models_{\mathbf{MAB}} [+_i\alpha](\Box_i^{ID}\varphi \rightarrow \Box_i^{ID}\Box_i^{ID}\varphi), \qquad (17)$$

$$\models_{\mathbf{MAB}} [+_i\alpha](\neg\Box_i^{ID}\varphi \rightarrow \Box_i^{ID}\neg\Box_i^{ID}\varphi). \qquad (18)$$

This contrasts with the behavior of the operator $\Box_i$ for which — as we observed in Section 7.2 — preservation of positive introspection on beliefs under private belief base expansion is not guaranteed.

Another interesting difference between the operators $\Box_i$ and $\Box_i^{ID}$ concerns stability of implicit beliefs under inference. As we observed in Section 7.2, if an agent implicitly believes that φ, then she will still implicitly believe that φ after inferring that α. Moreover, if she does not believe that φ implicitly, then she will not believe that φ implicitly after inferring that α. This stability property is not satisfied by the operator $\Box_i^{ID}$. Indeed, the following two formulas are satisfiable for the class $\mathbf{MAB}$:

$$\Box_i^{ID}\varphi \wedge \neg[\mathit{infer}(i,\alpha)]\Box_i^{ID}\varphi,$$
$$\neg\Box_i^{ID}\varphi \wedge \neg[\mathit{infer}(i,\alpha)]\neg\Box_i^{ID}\varphi.$$

To see this, let $B = (B_1, V)$, $B' = (B'_1, V')$, $B'' = (B''_1, V'')$, $Cxt = \{B, B', B''\}$ with $B_1 = B'_1 = B''_1 = \emptyset$ and $V = V' = V'' = \{p\}$. It is easy to check that

$$(B, Cxt) \models \Box_1^{ID}\neg\triangle_1 p \wedge \Box_1^{ID}p \wedge \neg[+_1 p]\Box_1^{ID}\neg\triangle_1 p.$$

Thus,

$$(B, Cxt) \models \Box_1^{ID}\neg\triangle_1 p \wedge \neg[\mathit{infer}(1,p)]\Box_1^{ID}\neg\triangle_1 p.$$

Similarly, we can observe that

$$(B, Cxt) \models \neg\Box_1^{ID}\triangle_1 p \wedge \neg[infer(1,p)]\neg\Box_1^{ID}\triangle_1 p.$$

Future work will be devoted to study more in detail this variant of the logic LDA in which $\Box_i$-operators are replaced by $\Box_i^{ID}$-operators. Specifically, for each model class $\mathbf{MAB}_X$ with $X \subseteq \{BC, GC\}$, we plan to study complexity of its satisfiability problem and provide a proof theory.

Operators for common ground and distributed belief. In future work, we also plan to extend LDA by a common ground operator $\blacksquare$ with the following interpretation relative to a MAB (B, Cxt):

$$(B, Cxt) \models \blacksquare\varphi \iff \forall B' \in Cxt : (B', Cxt) \models \varphi.$$

The operator $\blacksquare$ is the syntactic counterpart of the context Cxt. It corresponds to the universal modal operator studied in modal logic [44]. The interest of having it in the language lies in the possibility of completing the conceptual framework depicted in Figure 1 by formalizing the influence of the agents' common ground on their deductive processes and, consequently, the connection between the agents' explicit beliefs and their implicit beliefs via their common ground. The following is an example of validity relative to the class $\mathbf{MAB}$ capturing such a connection:

$$\models_{\mathbf{MAB}} \big(\triangle_i\alpha \wedge \blacksquare(\alpha \to \beta)\big) \to \Box_i\beta. \tag{19}$$

Common ground is a kind of collective belief. As we pointed out in Section 3.1, it is conceivable as the body of information shared by the agents which is in the background of their inference processes.

There is another kind of collective belief that deserves to be studied in the context of the LDA framework. It is the notion of distributed belief that, similarly to individual belief, can be either of explicit type or of implicit type. Distributed explicit belief is the results of pooling together the agents' belief bases. Specifically, the agents in the coalition $J \in 2^{Agt*} = \big(2^{Agt} \setminus \{\emptyset\}\big)$ have a distributed explicit belief that α, denoted by $\triangle_J\alpha$, if and only if some agent in J explicitly believes that α. In formal terms, $\triangle_J\alpha$ is interpreted relative to a multi-agent belief base, as follows:

$$B \models \triangle_J\alpha \iff \alpha \in B_J,$$

with $B_J = \bigcup_{i\in J} B_i$. In order to define distributed implicit belief we have to generalize the doxastic accessibility relation of Definition 5 to coalitions, as follows:

$$B\mathcal{R}_J B' \text{ if and only if } \forall\alpha \in B_J : B' \models \alpha.$$

By means of the relation $\mathcal{R}_J$, we can provide a semantic interpretation of the distributed implicit belief operator relative to a MAB, which parallels the interpretation of the individual implicit belief operator given in Definition 6:

$$(B, Cxt) \models \Box_J\varphi \iff \forall B' \in Cxt : \text{if } B\mathcal{R}_J B' \text{ then } (B', Cxt) \models \varphi.$$

This means that the agents in the coalition J have a distributed implicit belief that φ, denoted by $\Box_J\varphi$, if and only if φ is true at all states that satisfy the explicit beliefs of the agents in J. The following validities capture some interesting relationships between individual belief and distributed belief:

$$\models_{\mathbf{MAB}} \triangle_J\alpha \leftrightarrow \bigvee_{i\in J} \triangle_i\alpha, \tag{20}$$

$$\models_{\mathbf{MAB}} \triangle_J\alpha \rightarrow \Box_J\alpha, \tag{21}$$

$$\models_{\mathbf{MAB}} \Box_{\{i\}}\varphi \leftrightarrow \Box_i\varphi, \tag{22}$$

$$\models_{\mathbf{MAB}} \Box_J\varphi \rightarrow \Box_{J'}\varphi \text{ if } J \subseteq J'. \tag{23}$$

The first validity reduces distributed explicit belief that α to the existence of an individual explicit belief that α in the coalition. The second validity is a generalization of Axiom $\mathbf{Int}_{\triangle_i,\Box_i}$ given in Section 4 to coalitions of agents. The third and fourth validities are standard principles for distributed implicit belief according to which distributed belief of a singleton coalition is the same as individual belief of the agent in the singleton, and the larger the sub-coalition, the greater the distributed belief of that subcoalition. In future work, we plan to study the proof-theoretic as well as the complexity aspects of this LDA extension by distributed explicit and implicit belief.

Connection with machine learning. As a long-term objective, we plan to explore the connection between LDA and machine learning. Specifically, we plan to combine the LDA framework with machine learning methods, such as inductive logic programming (ILP) [65] and learning of Horn clauses [7], in order to acquire information to be added to the agents' belief bases through experience. More generally, the idea is to construct an agent's belief base through inductive methods based on machine learning techniques, and then to exploit the information contained in the agents' belief base for deductive reasoning. 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. The challenging aspect of the integration between the LDA framework and machine learning techniques such as ILP and learning of Horn clauses lies in the fact that the latter techniques are mostly developed in the context of propositional logic and first-order logic. We will need to adapt them to the multimodal setting of the logic LDA.

Last but not least, we plan to develop decision procedures based on tableaux for all variants of LDA presented in Sections 4 and 6. We expect these decision procedures to be exploitable in the context of AI applications including social robots and conversational agents, as the ones briefly discussed in Section 7.3. In [57] we made the first steps into this direction by proposing tableau-based decision procedures for the logics LDA, $\mathsf{LDA}_{\{\mathbf{D}_{\Box_i}\}}$ and $\mathsf{LDA}_{\{\mathbf{T}_{\Box_i}\}}$. Nonetheless, there is still a long way ahead for what concerns their implementation as well as the experimental analysis of their performance in comparison with the performance of existing decision procedures for standard epistemic logic.

Acknowledgments

Support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute and from the ANR project CoPains ("Cognitive Planning in Persuasive Multimodal Communication", grant number ANR-18-CE33-0012) is gratefully acknowledged. The author is also grateful to Johan van Benthem and Andreas Herzig for their insightful comments on the content of this work, as well as to the three anonymous reviewers of the paper for their critical comments and thoughtful suggestions.

References

[1] T. Ågotnes and N. Alechina. Full and relative awareness: a decidable logic for reasoning about knowledge of unawareness. In *Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge (TARK'07)*, pages 6–14. ACM, 2007.

[2] T. Ågotnes and N. Alechina. The dynamics of syntactic knowledge. *Journal of Logic and Computation*, 17(1):83–116, 2007.

[3] T. Ågotnes and N. Alechina. A logic for reasoning about knowledge of unawareness. *Journal of Logic, Language and Information*, 23(2):197–217, 2014.

[4] S. V. Albrecht and P. Stone. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence*, 258:66–95, 2018.

[5] N. Alechina, B. Logan, and M. Whitsey. A complete and decidable logic for resource-bounded agents. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004)*, pages 606–613. IEEE Computer Society, 2004.

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

[7] D. Angluin, M. Frazier, and L. Pitt. Learning conjunctions of horn clauses. *Machine Learning*, 9(2-3):147–164, 1992.

[8] S. N. Artëmov. The logic of justification. *Review of Symbolic Logic*, 37(2):174–203, 2008.

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

[10] P. Balbiani, D. Fernández-Duque, and E. Lorini. A logical theory of belief dynamics for resource-bounded agents. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS 2016)*, pages 644–652. ACM, 2016.

[11] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In *Proceedings of LOFT 7*, volume 3 of *Texts in Logic and Games*, pages 13–60. Amsterdam University Press, 2008.

[12] A. Baltag, L. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. In Itzhak Gilboa, editor, *Proceedings of the Seventh Conference on Theoretical Aspects of Rationality and Knowledge (TARK'98)*, pages 43–56, San Francisco, CA, 1998. Morgan Kaufmann.

[13] M. Banerjee and D. Dubois. A simple logic for reasoning about incomplete knowledge. *International Journal of Approximate Reasoning*, 55:639–653, 2014.

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

[15] F. Berto. Impossible worlds and the logic of imagination. *Erkenntnis*, 82:1277–1297, 2017.

[16] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.

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

[18] T. Bolander and M. B. Andersen. Epistemic planning for single- and multi-agent systems. *Journal of Applied Non-Classical Logics*, 21(1):656–680, 2011.

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

[20] D. Bonnay and P. Égré. Inexact knowledge with introspection. *Journal of Philosophical Logic*, 38(2):179–227, 2009.

[21] R. Booth, T. Meyer, I. Varzinczak, and R. Wassermann. On the link between partial meet, kernel, and infra contraction and its application to horn logic. *Journal of Artificial Intelligence Research*, 42:31–53, 2011.

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

[23] C. Breazeal, J. Gray, and M. Berlin. An embodied cognition approach to mindreading skills for socially intelligent robots. *The International Journal of Robotics Research*, 20(4):656–680, 2009.

[24] T. Charrier, A. Herzig, E. Lorini, F. Maffre, and F. Schwarzentruber. Building epistemic logic from observations and public announcements. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning (KR 2016)*, pages 268–277. AAAI Press, 2016.

[25] D. C. Dennett. *The Intentional Stance*. MIT Press, Cambridge, Massachusetts, 1987.

[26] D. C. Dennett. Précis of the intentional stance. *Behaviorial and Brain Sciences*, 11:495–546, 1988.

[27] H. N. Duc. Reasoning about rational, but not logically omniscient, agents. *Journal of Logic and Computation*, 7(5):633–648, 1997.

[28] R. A. Eberle. A logic of believing, knowing and inferring. *Synthese*, 26:356–382, 1974.

[29] R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1987.

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

[31] D. Fernández-Duque, A. Nepomuceno-Fernández, E. Sarrión-Morillo, F. Soler-Toscano, and F. R. Velázquez-Quesada. Forgetting complex propositions. *Logic Journal of the IGPL*, 23(6):942–965, 2015.

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

[33] A. I. Goldman. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, 2006.

[34] J. Grant, S. Kraus, and D. Perlis. A logic for characterizing multiple bounded agents. *Autonomous Agents and Multi-Agent Systems*, 3(4):351–387, 2000.

[35] J. Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(2):319–379, 1992.

[36] J. Y. Halpern and L. C. Rêgo. Reasoning about knowledge of unawareness. *Games and Economic Behavior*, 67(2):503–525, 2009.

[37] S. O. Hansson. *Belief Base Dynamics*. PhD thesis, Uppsala University, Sweden, 1991.

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

[39] S. O. Hansson. Kernel contraction. *Journal of Symbolic Logic*, 59(3):845–859, 1994.

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

[41] S. O. Hansson and R. Wassermann. Local change. *Studia Logica*, 70(1):49–76, 2002.

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

[43] A. Heifetz, M. Meyer, and B. C. Schipper. Interactive unawareness. *Journal of Economic Theory*, 130:78–94, 2006.

[44] E. Hemaspaandra. The price of universality. *Notre Dame Journal of Formal Logic*, 37(2):174–203, 1996.

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

[46] M. Jago. Epistemic logic for rule-based agents. *Journal of Logic, Language and Information*, 18(1):131–158, 2009.

[47] A. J. I. Jones. On the logic of self-deception. *South American Journal of Logic*, 1:387–400, 2015.

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

[49] K. Konolige. *A deduction model of belief*. Morgan Kaufmann Publishers, Los Altos, 1986.

[50] B. Kooi and B. Renne. Generalized arrow update logic. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2011)*, pages 205–211. ACM, 2011.

[51] B. Kooi and B. Renne. Arrow update logic. *Review of Symbolic Logic*, 4(4): 536–559, 2011.

[52] A. Kratzer. The notional category of modality. In H.-J. Eikmeyer and H. Rieser, editors, *Words, Worlds, and Contexts*. de Gruyter, Berlin / New York, 1981.

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

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

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

[56] E. Lorini. In praise of belief bases: Doing epistemic logic without possible worlds. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pages 1915–1922. AAAI Press, 2018.

[57] E. Lorini and F. Romero. Decision procedures for epistemic logic exploiting belief bases. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019)*, pages 944–952. IFAAMAS, 2019.

[58] C. Lutz. Complexity and succinctness of public announcement logic. In *Proceedings of the Fifth international Joint Conference on Autonomous agents and Multiagent Systems*, pages 137–143. ACM, 2006.

[59] D. Makinson. How to give it up: A survey of some formal aspects of the logic of theory change. *Synthese*, 62:347–363, 1985.

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

[61] J.-J. C. Meyer and W. van der Hoek. *Epistemic Logic for AI and Theoretical Computer Science*. Cambridge University Press, Oxford, 1995.

[62] G. Milliez, M. Warnier, A. Clodic, and R. Alami. 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*, pages 1103–1109. IEEE Press, 2014.

[63] S. Modica and A. Rustichini. Awareness and partitional information structures. *Theory and Decision*, 37:107–124, 1994.

[64] R. C. Moore and G. G. Hendrix. Computational models of belief and the semantics of belief sentences. In S. Peters and E. Saarinen, editors, *Processes, Beliefs, and Questions*, volume 16 of *Synthese Language Library*, pages 107–127. Cambridge University Press, 2011.

[65] S. Muggleton and L. de Raedt. Inductive logic programming: theory and methods. *Journal of Logic Programming*, 19-20:629–679, 1994.

[66] B. Nebel. Syntax-based approaches to belief revision. In P. Gärdenfors, editor, *Belief Revision*, pages 52–88. Cambridge University Press, Cambridge, 1992.

[67] C. Peters. 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*, pages 163–170, 2005.

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

[69] D. V. Pynadath, N. Wang, and S. C. Marsella. 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)*, volume 8108 of *LNCS*, pages 44–57. Springer, 2013.

[70] A. Rott. "Just because": Taking belief bases seriously. In *Logic Colloquium '98: Proceedings of the 1998 ASL European Summer Meeting*, volume 13 of *Lecture Notes in Logic*, pages 387–408. Association for Symbolic Logic, 1998.

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

[72] Y. Shoham. Logical theories of intention and the database perspective. *Journal of Philosophical Logic*, 38(6):633–648, 2009.

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

[74] J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007.

[75] J. van Benthem. Tracking information. In K. Bimbó, editor, *J. Michael Dunn on Information Based Logics*, Outstanding Contributions to Logic, pages 363–389. Springer, 2016.

[76] J. van Benthem and F. R. Velázquez-Quesada. The dynamics of awareness. *Synthese*, 177:5–27, 2010.

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

[78] J. van Benthem, D. Fernández-Duque, and E. Pacuit. Evidence logic: A new look at neighborhood structures. In *Proceedings of the Ninth Conference on Advances in Modal Logic (AiML 9)*, pages 97–118. College Publications, 2012.

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

[80] H. van Ditmarsch and T. French. Semantics for knowledge and change of awareness. *Journal of Logic, Language and Information*, 23(2):169–195, 2014.

[81] H. van Ditmarsch and B. Kooi. Semantic results for ontic and epistemic change. *CoRR*, abs/cs/0610093, 2006. URL `http://arxiv.org/abs/cs/0610093`.

[82] H. van Ditmarsch, T. French, and F. R. Velázquez-Quesada. Action models for knowledge and awareness. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems, (AAMAS 2012)*, pages 1091–1098. IFAAMAS, 2012.

[83] H. P. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Kluwer Academic Publishers, 2007. ISBN 1402058381.

[84] Hans P. van Ditmarsch, Wiebe van der Hoek, and Barteld P. Kooi. Dynamic epistemic logic with assignment. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2005)*, pages 141–148, 2005.

[85] van Ditmarsch H., A. Herzig, J. Lang, and P. Marquis. Introspective forgetting. *Synthese*, 169(2):405–423, 2009.

[86] F. R. Velázquez-Quesada. Explicit and implicit knowledge in neighbourhood models. In *Proceedings of the Fourth International Workshop on Logic, Rationality and Interaction (LORI 2013)*, volume 8196 of *LNCS*, pages 239–252. Springer, 2013.

[87] H. Wansing. A general possible worlds framework for reasoning about knowledge and belief. *Studia Logica*, 49:523–539, 1990.

[88] R. Wassermann. Resource bounded belief revision. *Erkenntnis*, 50(2-3):429–446, 1999.

[89] T. Williamson. Inexact knowledge. *Mind*, 101:217–242, 1992.

[90] A. F. T. Winfield. Experiments in artificial theory of mind: From safety to storytelling. *Frontiers in Robotics and AI*, 5(75), 2018.

[91] 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. The grand challenges of science robotics. *Science Robotics*, 3(14), 2018.

AppendixA. Proof of Theorem 1 and Theorem 3

AppendixA.1. Proof of Theorem 1

We divide the proof of Theorem 1 in three parts, each part corresponding to the proof of one of the following equivalence results: (i) satisfiability relative to quasi-notional models implies satisfiability relative to finite quasi-notional models, (ii) satisfiability relative to finite quasi-notional models implies satisfiability relative to finite notional models, and (iii) satisfiability relative to multi-agent belief bases is equivalent to satisfiability relative to notional models.

Satisfiability relative to quasi-NDMs implies satisfiability relative to finite quasi-NDMs. We use a filtration argument to show that if a formula φ of the language $\mathcal{LANG}_{\mathsf{LDA}}$ is true in a (possibly infinite) quasi-NDM then it is true in a finite quasi-NDM.

Let $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ be a (possibly infinite) quasi-NDM and let $\Sigma \subseteq \mathcal{LANG}_{\mathsf{LDA}}$ be an arbitrary finite set of formulas which is closed under subformulas. (Cf. Definition 2.35 in [16] for a definition of subformulas closed set of formulas.) Let the equivalence relation $\equiv_\Sigma$ on W be defined as follows. For all $w, v \in W$:

$$w \equiv_\Sigma v \text{ iff } \forall \varphi \in \Sigma : (M, w) \models \varphi \text{ iff } (M, v) \models \varphi.$$

Let $[w]_\Sigma$ be the equivalence class of the world w with respect to the equivalence relation $\equiv_\Sigma$.

We define W_Σ to be the filtrated set of worlds with respect to Σ:

$$W_\Sigma = \{[w]_\Sigma : w \in W\}.$$

Clearly, W_Σ is a finite set.

Let us define the filtrated valuation function $\mathcal{V}_\Sigma$. For every $p \in Atm$, we define:

$$\begin{array}{ll} \mathcal{V}_\Sigma(p) = \{[w]_\Sigma : (M, w) \models p\} & \text{if } p \in Atm(\Sigma), \\ \mathcal{V}_\Sigma(p) = \emptyset & \text{otherwise.} \end{array}$$

The next step in the construction consists in defining the filtrated doxastic function. For every $i \in Agt$ and for every $[w]_\Sigma \in W_\Sigma$, we define:

$$\mathcal{D}_\Sigma(i, [w]_\Sigma) = \big(\bigcap_{w \in [w]_\Sigma} \mathcal{D}(i, w) \big) \cap \Sigma.$$

Finally, for every $i \in Agt$ and for every $[w]_\Sigma \in W_\Sigma$, we define agent i's set of notional worlds at $[w]_\Sigma$ as follows:

$$\mathcal{N}_\Sigma(i, [w]_\Sigma) = \{[v]_\Sigma \in W_\Sigma : \exists w \in [w]_\Sigma, \exists v \in [v]_\Sigma \text{ such that } v \in \mathcal{N}(i, w)\}.$$

We call the model $M_\Sigma = (W_\Sigma, \mathcal{D}_\Sigma, \mathcal{N}_\Sigma, \mathcal{V}_\Sigma)$ the filtration of M under Σ.

We can state the following filtration lemma.

Lemma 4. *Let $\varphi \in \Sigma$ and let $w \in W$. Then, $(M, w) \models \varphi$ if and only if $(M_\Sigma, [w]_\Sigma) \models \varphi$.*

PROOF. The proof is by induction on the structure of φ. For the ease of exposition, we prove our result for the language $\mathcal{LANG}_{\mathsf{LDA}}$ in which the "diamond" operator $\Diamond_i$ is taken as primitive and the "box" operator $\Box_i$ is defined from it. Since the two operators are inter-definable, this does not affect the validity of our result.

The case $\varphi = p$ is immediate from the definition of $\mathcal{V}_\Sigma$. The boolean cases $\varphi = \neg\psi$ and $\varphi = \psi_1 \wedge \psi_2$ follow straightforwardly from the fact that Σ is closed under subformulas. This allows us to apply the induction hypothesis.

Let us prove the case $\varphi = \triangle_i\alpha$.

($\Rightarrow$) Suppose $(M, w) \models \triangle_i\alpha$ with $\triangle_i\alpha \in \Sigma$. Thus, $\alpha \in \mathcal{D}(i, w)$. Hence, by definition of $\mathcal{D}_\Sigma(i, [w]_\Sigma)$ and the fact that Σ is closed under subformulas, we have $\alpha \in \mathcal{D}_\Sigma(i, [w]_\Sigma)$. It follows that $(M_\Sigma, [w]_\Sigma) \models \triangle_i\alpha$.

($\Leftarrow$) For the other direction, suppose $(M_\Sigma, [w]_\Sigma) \models \triangle_i\alpha$ with $\triangle_i\alpha \in \Sigma$. Thus, $\alpha \in \mathcal{D}_\Sigma(i, [w]_\Sigma)$. Hence, by definition of $\mathcal{D}_\Sigma(i, [w]_\Sigma)$, $\alpha \in \mathcal{D}(i, w)$.

Let us conclude the proof for the case $\varphi = \Diamond_i\psi$. It is easy to check that $\mathcal{N}_\Sigma$ gives rise to the smallest filtration and that the following two properties hold for all $w, v \in W$ and for all $i \in Agt$:

(i) if $v \in \mathcal{N}(i, w)$ then $[v]_\Sigma \in \mathcal{N}_\Sigma(i, [w]_\Sigma)$, and

(ii) if $[v]_\Sigma \in \mathcal{N}_\Sigma(i, [w]_\Sigma)$ then for all $\Diamond_i\varphi \in \Sigma$, if $(M, v) \models \varphi$ then $(M, w) \models \Diamond_i\varphi$.

($\Rightarrow$) Suppose $(M, w) \models \Diamond_i\psi$ with $\Diamond_i\psi \in \Sigma$. Thus, there exists $v \in \mathcal{N}(i, w)$ such that $(M, v) \models \psi$. By the previous item (i), $[v]_\Sigma \in \mathcal{N}_\Sigma(i, [w]_\Sigma)$. Since Σ is closed under subformulas, we have $\psi \in \Sigma$. Thus, by the induction hypothesis, $(M_\Sigma, [v]_\Sigma) \models \psi$. It follows that $(M_\Sigma, [w]_\Sigma) \models \Diamond_i\psi$.

($\Leftarrow$) For the other direction, suppose $(M_\Sigma, [w]_\Sigma) \models \Diamond_i\psi$ with $\Diamond_i\psi \in \Sigma$. Thus, there exists $[v]_\Sigma \in \mathcal{N}_\Sigma(i, [w]_\Sigma)$, such that $(M_\Sigma, [v]_\Sigma) \models \psi$. Since Σ is closed under subformulas, by the induction hypothesis, we have $(M, v) \models \psi$. By the item (ii) above, it follows that $(M, w) \models \Diamond_i\psi$. ■

The following proposition highlights that M_Σ is the right model construction, as it is an element of the class of finite quasi-NDMs.

Proposition 12. *The tuple $M_\Sigma = (W_\Sigma, \mathcal{D}_\Sigma, \mathcal{N}_\Sigma, \mathcal{V}_\Sigma)$ is a finite quasi-NDM. Moreover, for every $x \in \{GC, BC\}$, if M satisfies the condition x then M_Σ satisfies it as well.*

PROOF. Clearly, M_Σ is finite. We are going to prove that it satisfies the Condition C1* in Definition 12.

By Lemma 4, if $\alpha \in \big(\bigcap_{w\in[w]_\Sigma} \mathcal{D}(i, w)\big) \cap \Sigma$ then $||\alpha||_{M_\Sigma} = \{[v]_\Sigma : v \in ||\alpha||_M\}$. Moreover, as M is a quasi-NDM, we have

$$\mathcal{N}(i, w) \subseteq \bigcap_{\alpha\in\mathcal{D}(i,w)} ||\alpha||_M \subseteq \bigcap_{\alpha\in\big(\bigcap_{w\in[w]_\Sigma} \mathcal{D}(i,w)\big)\cap\Sigma} ||\alpha||_M.$$

Hence, by definitions of $\mathcal{N}_\Sigma(i, [w]_\Sigma)$ and $\mathcal{D}_\Sigma$,

$$\mathcal{N}_\Sigma(i, [w]_\Sigma) \subseteq \bigcap_{\alpha\in\mathcal{D}_\Sigma(i,[w]_\Sigma)} ||\alpha||_{M_\Sigma}.$$

Moreover, it is easy to verify that if M satisfies global consistency of Definition 10 then M_Σ satisfies it as well, and if M satisfies belief correctness of Definition 11 then M_Σ satisfies it as well. ■

The next lemma states the equivalence between the semantics in terms of quasi-NDMs and the semantics in terms of finite quasi-NDMs.

Lemma 5. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$ and let $X \subseteq \{GC, BC\}$. Then, if φ is satisfiable for the class $\mathbf{QNDM}_X$, then φ is satisfiable for the class finite-$\mathbf{QNDM}_X$.*

PROOF. Let M be a possibly infinite quasi-NDM satisfying every condition in X and let w be a world in M such that $(M, w) \models \varphi$. Moreover, let $sub(\varphi)$ be the set of subformulas of φ. Then, by Lemma 4 and Proposition 12, $(M_{sub(\varphi)}, [w]_{sub(\varphi)}) \models \varphi$ and $M_{sub(\varphi)}$ is a finite quasi-NDM. Moreover, Proposition 12 guarantees that $M_{sub(\varphi)}$ satisfies every condition in X. ■

Satisfiability relative to finite quasi-NDMs implies satisfiability relative to finite NDMs. Our second result concerns the equivalence between the semantics in terms of finite notional models and the semantics in terms of finite quasi-notional models. It is clearly expressed by the following lemma.

Lemma 6. *Let $\varphi \in \mathcal{LANG}_{\mathsf{LDA}}$ and let $X \subseteq \{GC, BC\}$. Then, if φ is satisfiable for the class finite-$\mathbf{QNDM}_X$, then φ is satisfiable for the class finite-$\mathbf{NDM}_X$.*

PROOF. The strategy of the proof consists in enlarging an agent i's belief base at world w (i.e., $\mathcal{D}(i, w)$) so that agent i's set of doxastic alternatives at w (i.e., $\mathcal{N}(i, w)$) shrinks and perfectly coincides with the set of worlds in which all formulas in agent i's belief base at w are true, as required by Condition C1 in Definition 9.

Let $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ be a finite quasi-NDM that satisfies φ, i.e., there exists $w \in W$ such that $(M, w) \models \varphi$. Let

$$\mathcal{T}(M) = \cup_{w \in W, i \in Agt} Atm(\mathcal{D}(i, w))$$

be the *terminology* of model M including all atomic propositions that are in the explicit beliefs of some agent at some world in M. Since M is finite, $\mathcal{T}(M)$ is finite too.

Let us introduce an injective function:

$$f : Agt \times W \longrightarrow Atm \setminus (\mathcal{T}(M) \cup Atm(\varphi))$$

which assigns an identifier to every agent in Agt and world in W. The fact that Atm is infinite while W, $\mathcal{T}(M)$ and $Atm(\varphi)$ are finite guarantees that such an injection exists.

The next step consists in defining the new model $M' = (W', \mathcal{D}', \mathcal{N}', \mathcal{V}')$ with $W' = W$, $\mathcal{N}' = \mathcal{N}$ and where $\mathcal{D}'$ and $\mathcal{V}'$ are defined as follows.

For every $i \in Agt$ and for every $w \in W$:

$$\mathcal{D}'(i, w) = \mathcal{D}(i, w) \cup \{f(i, w)\}.$$

Moreover, for every $p \in Atm$:

$$\begin{aligned} &\mathcal{V}'(p) = \mathcal{V}(p) && \text{if } p \in \mathcal{T}(M) \cup Atm(\varphi), \\ &\mathcal{V}'(p) = \mathcal{N}(i,w) && \text{if } p = f(i,w), \\ &\mathcal{V}'(p) = \emptyset && \text{otherwise.} \end{aligned}$$

It is easy to verify that M' satisfies Condition C1 in Definition 9. In particular, we have $\mathcal{N}'(i,w) = \bigcap_{\alpha \in \mathcal{D}'(i,w)} ||\alpha||_{M'}$ for all $i \in Agt$ and for all $w \in W'$. Thus, more generally, M' is a finite NDM.

Furthermore, it is easy to check that, for every condition $x \in \{GC, BC\}$, if M satisfies x then M' satisfies it as well. Indeed, for every $i \in Agt$, $\mathcal{N}'(i,w) = \mathcal{N}(i,w)$. Thus, if $\mathcal{N}(i,w) \neq \emptyset$ then $\mathcal{N}'(i,w) \neq \emptyset$ and if $w \in \mathcal{N}(i,w)$ then $w \in \mathcal{N}'(i,w)$.

By induction on the structure of φ, we prove that, for all $w \in W$, "$(M,w) \models \varphi$ iff $(M',w) \models \varphi$".

The case $\varphi = p$ is immediate from the definition of $\mathcal{V}'$. By the induction hypothesis, we can prove the boolean cases $\varphi = \neg\psi$ and $\varphi = \psi_1 \wedge \psi_2$ in a straightforward manner.

Let us prove the case $\varphi = \triangle_i \alpha$.

($\Rightarrow$) Suppose $(M,w) \models \triangle_i \alpha$. Then, we have $\alpha \in \mathcal{D}(i,w)$. Hence, by the definition of $\mathcal{D}'$, $\alpha \in \mathcal{D}'(i,w)$. Thus, $(M',w) \models \triangle_i \alpha$.

($\Leftarrow$) Suppose $(M',w) \models \triangle_i \alpha$. Then, we have $\alpha \in \mathcal{D}'(i,w)$. The definition of $\mathcal{D}'$ ensures that $\alpha \neq f(i,w)$, since $f(i,w) \notin Atm(\triangle_i \alpha)$. Thus, $\alpha \in \mathcal{D}(i,w)$ and, consequently, $(M,w) \models \triangle_i \alpha$.

Let us prove the case $\varphi = \Box_i \psi$. $(M,w) \models \Box_i \psi$ means that $(M,v) \models \psi$ for all $v \in \mathcal{N}(i,w)$. By induction hypothesis and the fact that $\mathcal{N}(i,w) = \mathcal{N}'(i,w)$, the latter is equivalent to $(M',v) \models \psi$ for all $v \in \mathcal{N}'(i,w)$. The latter means that $(M',w) \models \Box_i \psi$.

Since M satisfies φ and "$(M,w) \models \varphi$ iff $(M',w) \models \varphi$" for all $w \in W$, M' satisfies φ as well. ■

Satisfiability relative to multi-agent belief bases is equivalent to satisfiability relative to NDMs. Our third result concerns the equivalence between the multi-agent belief base semantics and the notional model semantics.

Lemma 7. *Let $\varphi \in \mathcal{LAN}\mathcal{G}_{\mathsf{LDA}}$ and let $X \subseteq \{GC, BC\}$. Then, φ is satisfiable for the class $\mathbf{MAB}_X$ if and only if φ is satisfiable for the class $\mathbf{NDM}_X$.*

PROOF. As for the left-to-right direction, we prove the following weaker result: if φ is satisfiable for the class $\mathbf{MAB}_X$, then φ is satisfiable for the class $\mathbf{QNDM}_X$.

Let (B, Cxt) be a MAB with $B = (B_1, \ldots, B_n, V)$ and such that $(B, Cxt) \models \varphi$. We define the structure $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ as follows:

- $W = \{w_{B'} : B' \in Cxt \cup \{B\}\}$,
- for every $i \in Agt$ and for every $w_{B'} \in W$: $\mathcal{D}(i, w_{B'}) = B'_i$,
- for every $i \in Agt$ and for every $w_{B'} \in W$: $\mathcal{N}(i, w_{B'}) = \{w_{B''} \in W : B'' \in Cxt \text{ and } B' \mathcal{R}_i B''\}$,

- for every $p \in Atm$: $\mathcal{V}(p) = \{w_{B'} \in W : B' \models p\}$.

One can show that M so defined is a quasi-NDM and that, for every condition $x \in \{GC, BC\}$, if (B, Cxt) satisfies x then M satisfies it as well. Moreover, by induction on the structure of φ, one can prove that, for all $w_{B'} \in W$, $(M, w_{B'}) \models \varphi$ iff $(B', Cxt) \models \varphi$. Thus, $(M, w_B) \models \varphi$, since $(B, Cxt) \models \varphi$.

We have proved that if φ is satisfiable for the class $\mathbf{MAB}_X$, then φ is satisfiable for the class $\mathbf{QNDM}_X$. Thus, by Lemmas 5 and 6, we have that if φ is satisfiable for the class $\mathbf{MAB}_X$, then φ is satisfiable for the class $\mathbf{NDM}_X$.

We now prove the right-to-left direction. Let $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ be a NDM and let w be a world in W such that $(M, w) \models \varphi$.

Let us say that a NDM $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ is redundant if and only if there are $w, v \in W$ such that $w \neq v$ and $w \sim v$, where $w \sim v$ denotes the fact that $\{p \in Atm : w \in \mathcal{V}(p)\} = \{p \in Atm : v \in \mathcal{V}(p)\}$ and, for all $i \in Agt$, $\mathcal{D}(i, w) = \mathcal{D}(i, v)$. Intuitively, a redundant NDM is a NDM which contains two identical worlds, where the criterion for saying that two worlds are identical is that valuations of propositional atoms and agents' belief bases are the same in the two worlds. The class of redundant NDMs includes all NDMs of this kind. The class of non-redundant NDMs is the complementary class including all NDMs which are not redundant. A non-redundant pointed NDM is a pointed NDM (M, w) such that M is non-redundant.

We can show that if a formula is satisfiable for the class of NDMs then it is satisfiable for the class of non-redundant NDMs. Indeed, to every NDM $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ we can associate a NDM $M^{nr} = (W^{nr}, \mathcal{D}^{nr}, \mathcal{N}^{nr}, \mathcal{V}^{nr})$ such that:

$$W^{nr} = \{[v] : v \in W\} \text{ with } [v] = \{u \in W : v \sim u\},$$

for every $[v] \in W'$, for every $i \in Agt$ and for every $p \in Atm$:

$$\begin{aligned}
&\mathcal{D}^{nr}(i, [v]) = \mathcal{D}(v),\\
&\mathcal{N}^{nr}(i, [v]) = \{[u] \in W^{nr} : \exists v \in [v], \exists u \in [u] \text{ such that } u \in \mathcal{N}(i, v)\},\\
&\mathcal{V}^{nr}(p) = \{[v] \in W^{nr} : v \in \mathcal{V}(p)\}.
\end{aligned}$$

Clearly, M^{nr} is non-redundant. Moreover, by induction on the structure of the formula, we can prove that, for every $v \in W$ and for every $\psi \in \mathcal{LANG}_{\mathsf{LDA}}$, $(M, v) \models \psi$ if and only if $(M^{nr}, [v]) \models \psi$. The boolean cases and the case $\psi = \triangle_i \alpha$ are obvious. The only interesting case is $\psi = \Box_i \chi$. Let us prove it.

Suppose $(M, v) \models \Box_i \chi$. The latter means that $(M, u) \models \chi$ for all $u \in \mathcal{N}(i, v)$. By definition of $\mathcal{N}^{nr}(i, [v])$, $u \in \mathcal{N}(i, v)$ implies $[u] \in \mathcal{N}^{nr}(i, [v])$. Thus, by induction hypothesis, $(M^{nr}, [u]) \models \chi$ for all $[u] \in \mathcal{N}^{nr}(i, [v])$. The latter means that $(M^{nr}, [v]) \models \Box_i \chi$.

Now, suppose $(M^{nr}, [v]) \models \Box_i \chi$. Thus, $(M^{nr}, [u]) \models \chi$ for all $[u] \in \mathcal{N}^{nr}(i, [v])$. Hence, by induction hypothesis, $(M, u) \models \chi$ for all $[u] \in \mathcal{N}^{nr}(i, [v])$.

$[u] \in \mathcal{N}^{nr}(i, [v])$ means that $u' \in \mathcal{N}(i, v')$ for some $v' \in [v]$ and for some $u' \in [u]$. By Condition C1 in Definition 9, if $v \sim v'$ and $u' \in \mathcal{N}(i, v')$ then $u' \in \mathcal{N}(i, v)$. Moreover, if $u \sim u'$ and $u' \in \mathcal{N}(i, v)$ then $u \in \mathcal{N}(i, v)$, since $u \sim u'$ implies "$(M, u) \models \alpha$ iff $(M, u') \models \alpha$" for all $\alpha \in \mathcal{LANG}_0$. Therefore, $[u] \in \mathcal{N}^{nr}(i, [v])$ implies $u \in \mathcal{N}(i, v)$.

Therefore, $(M, u) \models \chi$ for all $u \in \mathcal{N}(i, v)$. The latter means that $(M, v) \models \Box_i \chi$. Thus, we can conclude that $(M^{nr}, [w]) \models \varphi$, since $(M, w) \models \varphi$.

Intuitively speaking, the reason for applying a truth-preserving transformation of the initial NDM M into the non-redundant NDM M^{nr} is to obtain a MAB which is isomorphic to M^{nr} and which is bound to satisfy the same formulas as M^{nr} in the light of this isomorphism.[15] Let us define such a MAB formally.

To every $[v] \in W^{nr}$, we associate a tuple $B^{[v]} = (B_1^{[v]}, \ldots, B_n^{[v]}, V^{[v]})$ such that (i) $V^{[v]} = \{p \in Atm : [v] \in \mathcal{V}^{nr}(p)\}$ and (ii) $B_i^{[v]} = \mathcal{D}^{nr}(i, [v])$ for each $i \in Agt$. Moreover, we define the context $Cxt = \{B^{[v]} : [v] \in W^{nr}\}$. It is easy to show that, for every $B^{[v]} \in Cxt$, $(B^{[v]}, Cxt)$ is a MAB and that, for every condition $x \in \{GC, BC\}$, if M satisfies x then $(B^{[v]}, Cxt)$ satisfies it as well.

The mapping $f : [v] \mapsto B^{[v]}$ clearly defines a bijection from W^{nr} to Cxt such that $(M^{nr}, [v]) \models \varphi$ iff $\big(f([v]), Cxt\big) \models \varphi$. Thus, $\big(f([w]), Cxt\big) \models \varphi$, since $(M^{nr}, [w]) \models \varphi$. ■

Conclusion. The previous Lemmas 5, 6 and 7 are sufficient for proving Theorem 1. Indeed, as Figure A.4 highlights, the five semantics defined in Section 3 are all equivalent relative to the language $\mathcal{LANG}_{\mathsf{LDA}}$, since from every node in the graph we can reach all other nodes.

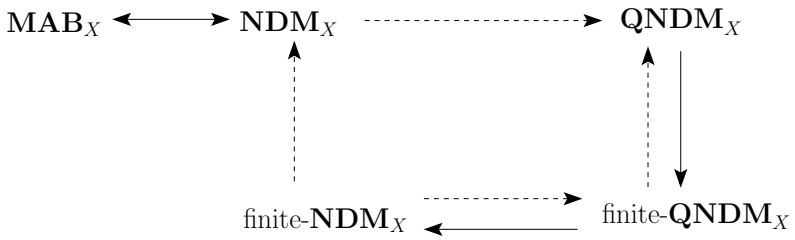


Figure A.4: Relations between semantics for the language $\mathcal{LANG}_{\mathsf{LDA}}$. An arrow means that satisfiability relative to the first class of structures implies satisfiability relative to the second class of structures. Full arrows correspond to the results stated in Lemmas 5, 6 and 7. Dotted arrows denote relations that follow straightforwardly given the inclusion between classes of structures.

AppendixA.2. Proof of Theorem 3

Suppose φ is satisfiable for the class $\mathbf{NDM}_X$. Thus, by Corollary 1, it is LDA_X-consistent. Hence, by Theorem 2, it is satisfiable for the class $\mathbf{QNDM}_X$. From the proof of Lemma 5, we can observe that if φ is satisfiable for the class $\mathbf{QNDM}_X$ then there exists a finite $M \in \mathbf{QNDM}_X$ satisfying φ such that (i) M includes at most 2^n worlds, (ii) the atomic propositions outside $Atm(sub(\varphi))$ are false at every world of M, and (iii) the belief base of an agent at a world of M contains only formulas

[15] Note that this transformation is not required for the proof of the left-to-right direction of the lemma since, by definition, a MAB cannot contain two identical copies of the same state. In this sense, MABs are "intrinsically" non-redundant.

from $sub(\varphi)$, where n is the size of $sub(\varphi)$. The construction in the proof of Lemma 6 ensures that from M, we can build a finite $M' \in \mathbf{NDM}_X$ satisfying φ for which condition (i) holds and such that (iv) the atomic propositions outside $Atm(sub(\varphi)) \cup Y$ are false at every world of M', and (v) the belief base of an agent at a world of M' contains only formulas from $sub(\varphi) \cup Y$, where Y is an arbitrary set of atoms from $Atm \setminus (Atm(\varphi))$ of size at most $2^n \times |Agt|$. Thus, in order to verify whether φ is satisfiable for the class $\mathbf{NDM}_X$, we fix a Y and check satisfiability of φ for all NDMs in $\mathbf{NDM}_X$ satisfying conditions (i), (iv) and (v). There are finitely many NDMs of this kind.