Logical Theories of Collective Attitudes and the Belief Base Perspective

Emiliano Lorini, Eloan Rapion

Logical Theories of Collective Attitudes and the Belief Base Perspective

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

Éloan Rapion
ENS Rennes
Rennes, France
eloan.rapion@ens-rennes.fr

ABSTRACT

We present two logics of collective belief with a semantics exploiting the notion of belief base. The semantics distinguishes explicit from implicit belief: an agent's belief of explicit type is a piece of information contained in the agent's belief base, while a belief of implicit type corresponds to a piece of information that is derivable from the agent's belief base. The first part of the paper is devoted to the logic of implicit common belief, while the second presents the logic of explicit common belief. Implicit common belief is defined as a mutual belief of any order. This leads to the usual fixpoint construction of common belief. Explicit common belief is the collective counterpart of explicit individual belief and has a public nature. It moreover implies implicit common belief. We study axiomatic aspects of our logics as well as complexity of satisfiability checking. We show that, while the satisfiability checking problem is EXPTIME-hard for the logic of implicit common belief, it is in PSPACE for the logic of explicit common belief. This makes the latter logic a natural candidate for reasoning about collective attitudes in multi-agent scenarios and applications. We also study a dynamic extension of the logic of explicit common belief in which private and public forms of information dynamics can be modeled.

KEYWORDS

Reasoning about beliefs; modal logic; epistemic logic

ACM Reference Format:
Emiliano Lorini and Éloan Rapion. 2022. Logical Theories of Collective Attitudes and the Belief Base Perspective . In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 9 pages.

1 INTRODUCTION

Epistemic logic is the logic of epistemic attitudes including knowledge and belief. Since the pioneering work of Hintikka [24], it has been widely studied in artificial intelligence (AI) [15, 35] and economics [29]. It offers a rich language for modeling not only agents' beliefs about propositional facts, but also higher-order beliefs. Standard semantics for epistemic logic languages exploit the so-called multi-agent *Kripke models*, namely, multi-relational structures equipped with valuation functions for the interpretation of atomic formulas. Binary relations in a multi-agent Kripke model are called *epistemic accessibility relations* and are used to describe the agents' epistemic states and uncertainties.

Higher-order beliefs are essential constituents of collective attitudes of agents including common knowledge and common belief [43]. According to the standard fixpoint definition, a group of agents J has the common belief that the fact φ is true if and only if the agents in J mutually believe that φ for every order $k \geq 1$, that is, every agent in J believes that φ, every agent in J believes that every agent in J believes that φ, and so on *ad infinitum*.

The earliest analysis of common belief can be found in [16, 28]. The economist Robert Aumann [2] gave the first set-theoretic characterization of common knowledge. Some initial analysis of common belief and common knowledge based on epistemic logic can be found in [4, 6]. The axiomatic and complexity properties of the standard approach to common belief based on the Kripke semantics were fully studied in [20].[1] Weakest systems of common belief and common knowledge not satisfying, e.g., closure under logical consequence for individual beliefs or the least fixpoint principle for common belief were studied in [8, 17, 22, 29, 36]. Common belief has played a fundamental role in the analysis of shared cooperative activity of both human agents [10] and artificial agents [18, 27], intentional communication [12], social conventions [28] and social institutions [30]. It has also be demonstrated to be an essential constituent of the concept of common ground [41], as a basis for discourse understanding and definite reference [11, 39]. Finally, it is central in epistemic game theory [37] in which several solution concepts including iterated deletion of strongly/weakly dominated strategies (IDSDS/IDWDS) [9, 31] and backward induction [3] are justified in the light of common belief that all players are rational.

In this paper, we present a novel approach to common belief which relies on a formal semantics exploiting the notion of belief base. The latter distinguishes explicit from implicit belief: an agent's belief of explicit type is a piece of information contained in the agent's belief base, while a belief of implicit type corresponds to a piece of information that is derivable from the agent's belief base. This semantics and its corresponding logic of explicit and implicit individual belief were introduced in [33] (see also [32, 34]). They generalize the concept of belief base from knowledge representation [21] and, more generally, the database perspective on modeling agents' mental attitudes [40] to the multi-agent case. In this paper, we extend them with collective attitudes of coalitions of agents. We study two logical systems: a logic of explicit and implicit individual belief extended by the notion of implicit common belief, and a logic of explicit and implicit individual belief extended by the notion of explicit common belief. Implicit common belief is the standard

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

[1] The axiomatics of common knowledge and belief given in [20] use a fixpoint axiom and an induction (*alias* least fixpoint) rule. In [19, 26], the induction rule is replaced by an induction axiom. An alternative axiomatization of S5-based common knowledge using a more intuitive version of the induction axiom was recently given in [23].

fixpoint notion of common belief we find in the literature [20], whereas explicit common belief is the collective counterpart of explicit individual belief and has a public nature. We stipulate that there is an explicit common belief in a coalition of agents about a certain fact α, if every agent explicitly believes that α, and every agent explicitly believes that there is an explicit common belief that α. This construction guarantees that explicit common belief implies implicit common belief, but not vice versa. The main reason for distinguishing explicit common belief from implicit common belief is that the logic of the former notion is intrinsically simpler than the logic of the latter one. In particular, we will show that, while the satisfiability checking problem is EXPTIME-hard for the logic of implicit common belief, it is in PSPACE for the logic of explicit common belief. This makes the latter logic a natural candidate for reasoning about collective attitudes of agents in multi-agent scenarios and applications.

The paper is organized as follows. In Section 2 we present the background material on the language and the semantics for explicit and implicit individual belief introduced in [33]. Section 3 is devoted to the logic of implicit common belief. Section 4 focuses on the logic of explicit common belief. We study the axiomatic and complexity properties of both logics. In Section 5, we illustrate the expressiveness of the logic of explicit common belief by using it to formalize an example of coordination between robotic agents. In Section 6, we present a dynamic extension of the logic of explicit common belief in which private and public forms of information dynamics can be modeled.

2 BACKGROUND

This section presents the language and the semantics for agents' individual beliefs of both explicit and implicit type introduced in [33].

2.1 Language

Assume a countably infinite set of atomic propositions $Atm = \{p, q, \ldots\}$ and a finite set of agents $Agt = \{1, \ldots, n\}$. We define the language in two steps. First, define the language $\mathcal{L}_0$ by:

$$\mathcal{L}_0 \stackrel{\text{def}}{=} \quad \alpha \quad ::= \quad p \mid \neg\alpha \mid \alpha_1 \wedge \alpha_2 \mid \triangle_i \alpha,$$

where p ranges over Atm and i ranges over Agt. $\mathcal{L}_0$ is the language for representing explicit beliefs. The formula $\triangle_i \alpha$ is read "agent i has the explicit belief that α". The language $\mathcal{L}$ extends $\mathcal{L}_0$ and is defined by:

$$\mathcal{L} \stackrel{\text{def}}{=} \quad \varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i \varphi,$$

where α ranges over $\mathcal{L}_0$ and i ranges over Agt. The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from α, $\neg$ and $\wedge$ in the standard way. The formula $\Box_i \varphi$ is read "agent i implicitly believes that φ". The abbreviation $\Diamond_i \varphi \stackrel{\text{def}}{=} \neg\Box_i \neg\varphi$ defines the concept of belief compatibility. The formula $\Diamond_i \varphi$ has to be read "φ is compatible with agent i's explicit beliefs".

2.2 Belief Base Semantics

The formal semantics for the language $\mathcal{L}$ exploit belief bases. Unlike the standard Kripke semantics in which possible worlds and epistemic alternatives are primitive, they are here defined from the primitive concept of belief base.

Definition 1 (State). *A state is a tuple $B = \big((B_i)_{i \in Agt}, V\big)$ where $B_i \subseteq \mathcal{L}_0$ is agent i's belief base, and $V \subseteq Atm$ is the actual environment. The set of all states is noted* $\mathbf{S}$.

The following definition specifies truth conditions for formulas in the sublanguage $\mathcal{L}_0$.

Definition 2 (Satisfaction relation). *Let $B = \big((B_i)_{i \in Agt}, V\big) \in \mathbf{S}$. 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 operators in the previous definition: agent i has the explicit belief that α if and only if α is included in her belief base.

States satisfying the property of belief correctness (BC) are of particular interest, namely, states in which agents' explicit beliefs are correct.

Definition 3 (Belief correct state). *Let $B = \big((B_i)_{i \in Agt}, V\big)$ be a state. We say it satisfies belief correctness (BC) if and only if, for every $i \in Agt$ and for every $\alpha \in \mathcal{L}_0$, if $\alpha \in B_i$ then $B \models \alpha$. The set of states satisfying property BC is noted* $\mathbf{S}_{BC}$.

The following definition introduces the notion of doxastic alternative.

Definition 4 (Doxastic alternatives). *Let $i \in Agt$. Then, $\mathcal{R}_i$ is the binary relation on the set* $\mathbf{S}$ *such that, for all $B = \big((B_i)_{i \in Agt}, V\big)$, $B' = \big((B'_i)_{i \in Agt}, V'\big) \in \mathbf{S}$:*

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

$B \mathcal{R}_i B'$ means that B' is a doxastic alternative for agent i at B, that is to say, B' is a state that at B agent i considers 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.

A multi-agent belief model (MAB), or simply model, is defined to be a state supplemented with a set of states, called *context*. The context Cxt is not necessarily equal to the set of all states $\mathbf{S}$, since there could be states in $\mathbf{S}$ incompatible with the general "laws of the domain" and, consequently, with the agents' epistemic states. For example, we might want to exclude from the context Cxt all states in which the propositions "1+1=2" and "1+1=3" are true concomitantly.

Definition 5 (Multi-agent belief model). *A multi-agent belief model (MAB) is a pair (B, Cxt), where $B \in \mathbf{S}$ and $Cxt \subseteq \mathbf{S}$. The class of MABs is noted* $\mathbf{M}$.

Note that in Definition 5 we do not require $B \in Cxt$. The following definition extends Definition 2 to the full language $\mathcal{L}$. Its formulas are interpreted with respect to MABs. (We omit Boolean cases, as they are defined in the usual way.)

Definition 6 (Satisfaction relation (cont.)). *Let $(B, Cxt) \in \mathbf{M}$. Then:*

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

According to the previous definition, agent i implicitly believes that φ if and only if, φ is true at all states in the context that i considers possible.

The notion of belief correctness can be lifted from states to models, as follows.

Definition 7 (Belief correct MAB). *Let $(B, Cxt) \in \mathbf{M}$. We say it satisfies belief correctness (BC) if and only if (i) $B \in Cxt$, and (ii) all states in Cxt satisfy property BC of Definition 3. The class of MABs satisfying property BC is noted $\mathbf{M}_{BC}$.*

It is easy to verify that (B, Cxt) satisfies BC iff $B \in Cxt$ and, for every $i \in Agt$, $\mathcal{R}_i \cap (Cxt \times Cxt)$ is reflexive.

Let $\varphi \in \mathcal{L}$. We say that φ is valid relative to the class $\mathbf{M}$ (resp. $\mathbf{M}_{BC}$), noted $\models_{\mathbf{M}} \varphi$ (resp. $\models_{\mathbf{M}_{BC}} \varphi$), if and only if, for every $(B, Cxt) \in \mathbf{M}$ (resp. $(B, Cxt) \in \mathbf{M}_{BC}$) we have $(B, Cxt) \models \varphi$. We say that φ is satisfiable for the class $\mathbf{M}$ (resp. $\mathbf{M}_{BC}$) if and only if $\neg\varphi$ is not valid for the class $\mathbf{M}$ (resp. $\mathbf{M}_{BC}$).

2.3 Axiomatics and Complexity

In [33, Def. 13], the logic LDA (Logic of Doxastic Attitudes) is defined as follows.

Definition 8 (LDA). *LDA is the extension of classical propositional logic by the following axioms and rule of inference:*

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

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

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

Logic $LDA_{T_{\Box_i}}$ extends LDA by the following additional axiom:

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

$\mathbf{K}_{\Box_i}$ and $\mathbf{Nec}_{\Box_i}$ are, respectively, the K axiom and the rule of necessitation for implicit belief. $\mathbf{Int}_{\triangle_i,\Box_i}$ is the "bridge" axiom between explicit and implicit belief, while $\mathbf{T}_{\Box_i}$ is the T axiom for *correct* implicit belief. In [33, Theorem 3], it is proved that LDA (resp. $\mathrm{LDA}_{T_{\Box_i}}$) is sound and complete for class $\mathbf{M}$ (resp. $\mathbf{M}_{BC}$). It is also proved that satisfiability checking for formulas in $\mathcal{L}$ relative to class $\mathbf{M}$ (resp. $\mathbf{M}_{BC}$) is PSPACE-complete [33, Theorem 6].

3 IMPLICIT COMMON BELIEF

In this section, we study a novel extension of the logic of explicit and implicit individual belief by implicit common belief. The latter is the standard notion of common belief in the literature [15].

3.1 Language and Semantics

We extend the language $\mathcal{L}$ by the implicit shared belief operator $\Box_J\varphi$ and implicit common belief operator $\Box_J^*$ and name $\mathcal{L}^{\Box_J^*}$ the resulting language. We define it as follows:

$$\mathcal{L}^{\Box_J^*} \stackrel{\mathsf{def}}{=} \quad \varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i\varphi \mid \Box_J\varphi \mid \Box_J^*\varphi,$$

with α ranging over $\mathcal{L}_0$, i ranging over Agt and J over the set of non-empty sets of agents (*alias* coalitions) $2^{Agt*} = 2^{Agt} \setminus \{\emptyset\}$.

Formula $\Box_J\varphi$ has to be read "the agents in J share the implicit belief that φ", whereas $\Box_J^*\varphi$ has to be read "the agents J have the implicit common belief that φ". Before giving a semantic interpretation of the implicit shared and common belief operator, we define the mutual belief operator inductively as follows:

$$\Box_J^0\varphi \stackrel{\mathsf{def}}{=} \varphi,$$

$$\Box_J^{k+1}\varphi \stackrel{\mathsf{def}}{=} \Box_J\Box_J^k\varphi \text{ for } k \geq 0.$$

The implicit shared and common belief operator are interpreted relative to a multi-agent belief model (B, Cxt) of Definition 5, as follows:

$$(B, Cxt) \models \Box_J\varphi \iff \forall i \in J : (B, Cxt) \models \Box_i\varphi,$$

$$(B, Cxt) \models \Box_J^*\varphi \iff \forall k \in \mathbb{N}^* : (B, Cxt) \models \Box_J^k\varphi.$$

According to the previous semantic interpretations, implicit shared belief coincides with implicit belief of all agents, while implicit common belief is the same as k-order mutual belief for every $k \in \mathbb{N}^*$.

3.2 Alternative Kripkean Semantics

In [33] an alternative semantics for the language $\mathcal{L}$ is given. It extends the standard multi-relational Kripke semantics of epistemic logic by agents' belief bases. We here use it for interpreting the language $\mathcal{L}^{\Box_J^*}$ and show that it has the same set of validities as the belief base semantics.

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,*
- *$\mathcal{D} : Agt \times W \longrightarrow 2^{\mathcal{L}_0}$ 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 $\mathcal{L}^{\Box_J^*}$ relative to a pair (M, w) with $w \in W$:

$$(M, w) \models p \iff w \in \mathcal{V}(p),$$

$$(M, w) \models \neg\varphi \iff (M, w) \not\models \varphi,$$

$$(M, w) \models \varphi \wedge \psi \iff (M, w) \models \varphi \text{ and } (M, w) \models \psi,$$

$$(M, w) \models \triangle_i\alpha \iff \alpha \in \mathcal{D}(i, w),$$

$$(M, w) \models \Box_i\varphi \iff \forall v \in \mathcal{N}(i, w) : (M, v) \models \varphi,$$

$$(M, w) \models \Box_J\varphi \iff \forall i \in J : (M, w) \models \Box_i\varphi,$$

$$(M, w) \models \Box_J^*\varphi \iff \forall k \in \mathbb{N}^* : (M, w) \models \Box_J^k\varphi.$$

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 say that M satisfies belief correctness (BC) if the following condition holds, for all $i \in Agt$ and for all $w \in W$:

(C2) $w \in \mathcal{N}(i, w)$.

The class of NDMs is noted **NDM**, *while the class of NDMs satisfying property BC is noted* $\mathbf{NDM}_{BC}$.

The pair (M, w) in the previous definition is called 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 $\mathcal{N}(i, w)$ is called agent i's set of *notional* worlds at world w, where the term 'notional' is borrowed from [13, 14] (see, also, [25]). As Condition C1 indicates, an agent's

notional world is a world at which all its explicit beliefs are true. Condition C2 is the counterpart of belief correctness for NDMs.

As usual, we say that formula $\varphi \in \mathcal{L}$ is valid relative to the class **NDM** (resp. $\mathbf{NDM}_{BC}$) if and only if, for every $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V}) \in$ **NDM** (resp. $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V}) \in \mathbf{NDM}_{BC}$) and for every $w \in W$, we have $(M, w) \models \varphi$. We say that φ is satisfiable for the class **NDM** (resp. $\mathbf{NDM}_{BC}$) if and only if $\neg\varphi$ is not valid for the class **NDM** (resp. $\mathbf{NDM}_{BC}$).

Quasi-notional doxastic models (quasi-NDMs) are like notional doxastic models of Definition 9 except that Constraint C1 is replaced by the following weaker constraint:

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

The class of quasi-NDMs is noted **QNDM**, whereas the class of quasi-NDMs satisfying property BC is noted $\mathbf{QNDM}_{BC}$.

A NDM (resp. quasi-NDM) $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ is said to be *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 NDMs (resp. finite quasi-NDMs) is noted finite-**NDM** (resp. finite-**QNDM**). The class of finite NDMs (resp. finite quasi-NDMs) satisfying property BC is noted finite-$\mathbf{NDM}_{BC}$ (resp. finite-$\mathbf{QNDM}_{BC}$).

The following theorem highlights that the five semantics for the language $\mathcal{L}^{\Box_J^*}$ are all equivalent.

Theorem 1. *Let $\varphi \in \mathcal{L}^{\Box_J^*}$. Then, the following five statements are equivalent.*

(1) φ is satisfiable relative to class **QNDM** *(resp.* $\mathbf{QNDM}_{BC}$*),*
*(2) φ is satisfiable relative to class finite-***QNDM** *(resp. finite-*$\mathbf{QNDM}_{BC}$*),*
*(3) φ is satisfiable relative to class finite-***NDM** *(resp. finite-*$\mathbf{NDM}_{BC}$*),*
(4) φ is satisfiable relative to class **NDM** *(resp.* $\mathbf{NDM}_{BC}$*),*
(5) φ is satisfiable relative to class **M** *(resp.* $\mathbf{M}_{BC}$*).*

Sketch of Proof. We use a filtration method to prove that (1) implies (2). In order to prove that (2) implies (3), we use a technique which consists in enlarging an agent i's belief base at a 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. Finally, we prove that (4) and (5) are equivalent. The right-to-left direction is easy since from a MAB we can easily construct the corresponding NDM. The left-to-right direction is less immediate. Indeed, a NDM can be redundant, i.e., it can contain two different worlds with the same valuation of propositional atoms and the same belief bases for the agents. We have to transform a possibly redundant NDM into a non-redundant one. From a non-redundant NDM we can construct the corresponding MAB which satisfies the same formulas. ■

3.3 Axiomatics

The logic LDA–ICB (Logic of Doxastic Attitudes and Implicit Common Belief) and its variant LDA–$\text{ICB}_{T_{\Box_i}}$ extend the logics LDA and $\text{LDA}_{T_{\Box_i}}$ by principles for implicit common belief taken from [20]. They are defined as follows.

Definition 10 (LDA–ICB). *LDA–ICB is the extension of LDA defined in Definition 8 by the following axioms and rule of inference:*

$$\Box_J \varphi \leftrightarrow \bigwedge_{i \in J} \Box_i \varphi \qquad (\mathbf{Def}_{\Box_J})$$

$$\Box_J^* \varphi \rightarrow \Box_J(\varphi \wedge \Box_J^* \varphi) \qquad (\mathbf{FP}_{\Box_J^*})$$

$$\frac{\psi \rightarrow \Box_J(\varphi \wedge \psi)}{\psi \rightarrow \Box_J^* \varphi} \qquad (\mathbf{Ind}_{\Box_J^*})$$

Logic LDA–$ICB_{T_{\Box_i}}$ extends LDA–ICB by Axiom $\mathbf{T}_{\Box_i}$ of Definition 8.

Axiom $\mathbf{FP}_{\Box_J^*}$ is the so-called fixpoint axiom for implicit common belief, while rule $\mathbf{Ind}_{\Box_J^*}$ is the so-called induction rule (*alias* least fixpoint rule).

As usual, for every $\varphi \in \mathcal{L}^{\Box_J^*}$, we write $\vdash_{\text{LDA–ICB}} \varphi$ to mean that φ is deducible in LDA–ICB, 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–ICB 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–ICB.

We say that the set of formulas Γ from $\mathcal{L}^{\Box_J^*}$ is LDA–ICB-consistent if there are no formulas $\varphi_1, \ldots, \varphi_m \in \Gamma$ such that $\vdash_{\text{LDA–ICB}} (\varphi_1 \wedge \ldots \wedge \varphi_m) \rightarrow \bot$. Moreover, φ is LDA–ICB-consistent if $\{\varphi\}$ is LDA–ICB-consistent. Definitions of LDA–$\text{ICB}_{T_{\Box_i}}$-deducibility and LDA–$\text{ICB}_{T_{\Box_i}}$-consistency are analogous. Definitions of strong completeness and weak completeness are the usual ones from modal logic (see, e.g., [7]): a logic Λ is said to be strongly complete with respect to a class of structures S iff every Λ-consistent set of formulas is satisfiable on some element of S. It is said to be weakly complete with respect to S iff every *finite* Λ-consistent set of formulas is satisfiable on some element of S.

We know for sure that the logics LDA–ICB and LDA–$\text{ICB}_{T_{\Box_i}}$ are not strongly complete. Indeed, they do not satisfy the compactness property, i.e., the fact that for every set of $\mathcal{L}^{\Box_J^*}$-formulas Γ, if Γ is LDA–ICB-consistent (resp. LDA–$\text{ICB}_{T_{\Box_i}}$-consistent) then there is $(B, Cxt) \in$ **M** (resp. $(B, Cxt) \in \mathbf{M}_{BC}$) such that $(B, Cxt) \models \psi$, for every $\psi \in \Gamma$. For example, let $\Gamma = \{\neg\Box_J^* p\} \cup \{\Box_{i_1} \ldots \Box_{i_k} p : k \in \mathbb{N}^*, i_1, \ldots, i_k \in J\}$. Γ is LDA–ICB-consistent, but there is no (B, Cxt)
$\in$ **M** such that $(B, Cxt) \models \psi$, for every $\psi \in \Gamma$. However, LDA–ICB and LDA–$\text{ICB}_{T_{\Box_i}}$ are weakly complete.

Theorem 2. *The logic LDA–ICB is sound and weakly complete for the class* **M***, and the logic LDA–$ICB_{T_{\Box_i}}$ is sound and weakly complete for the class* $\mathbf{M}_{BC}$*.*

Sketch of Proof. Soundness is a routine exercise. Thanks to Theorem 1, it is sufficient to prove completeness for the quasi-notional model semantics of Section 3.2. We prove the latter adapting the proof of Theorem 4.3 in [20]. Our proof uses a canonical model argument in which worlds in the canonical model are not maximally consistent sets of all formulas but rather finite maximally consistent subsets of set $\text{Sub}^*(\varphi)$ including all subformulas of an input formula φ, their negations and a finite "epistemic theory" capturing the interrelation between implicit individual and common

belief. The latter is necessary to prove that the canonical model so constructed belongs to the class **QNDM** (resp. $\mathbf{QNDM}_{BC}$). ■

In Section 4, we will replace implicit common belief by explicit common belief. We will show that the resulting logic satisfies the compactness property and is strongly complete. Moreover, it is computationally simpler than the logic of implicit common belief.

3.4 Complexity

It is easy to show that satisfiability checking for formulas in $\mathcal{L}^{\Box^*_J}$ relative to model classes $\mathbf{M}$ and $\mathbf{M}_{BC}$ is EXPTIME-hard. Indeed, LDA–ICB is a conservative extension of the logic of common belief with base logic K^n for individual belief, while $\mathsf{LDA–ICB}_{T_{\Box_i}}$ is a conservative extension of the logic of common belief with base logic KT^n for individual belief. These two logics are known to be EXPTIME-hard for $n > 1$ [20]. As the following theorem indicates, EXPTIME is also a upper bound.

Theorem 3. *Let $n > 1$. Then, satisfiability checking for formulas in $\mathcal{L}^{\Box^*_J}$ relative to the class $\mathbf{M}$ (resp. $\mathbf{M}_{BC}$) is in EXPTIME.*

Sketch of Proof. Given a formula φ, we define a variant of its set of sub-formulas. Then, we consider the set of maximally consistent subsets of such a set, which are intended to be the worlds of a canonical model. We set an algorithm which deletes some of those states, until we can effectively obtain a quasi-NDM. Then, we can simply check if φ is in one of the remaining sets to know if φ is satisfiable. Moreover, the algorithm can be executed in exponential time in the size of φ. ■

For $n = 1$, logics LDA–ICB and $\mathsf{LDA–ICB}_{T_{\Box_i}}$ are clearly PSPACE-complete, since they coincide with the single-agent logics of explicit and implicit belief with base logics K and KT for implicit belief which are known to be PSPACE-complete [33].

4 EXPLICIT COMMON BELIEF

In Section 3, we have presented the logic of explicit and implicit individual belief, and implicit common belief. The aim of this section is study a new concept of collective attitude of explicit type: explicit common belief. Explicit common belief is the collective counterpart of explicit individual belief. We will use symbols of type $\triangledown_J$ to represent it. We stipulate that *there is an explicit common belief that α if everyone explicitly believes that α, and everyone explicitly believes that there is an explicit common belief that α*. The latter highlights the public nature of explicit common belief: explicit common belief implies the agents' awareness of its existence. This is similar to the publicity condition of the concept of *co-presence*, as defined by Clark & Marshall [11]. For example, suppose Ann, Bob, Mary and Paul are sitting at a table on which there is a box. This is a situation of co-presence: each agent correctly believes that there is a box on the table and, moreover, that they are jointly seeing that there is a box on the table so that they commonly believe so.

4.1 Language and Semantics

We name $\mathcal{L}_0^{\triangledown_J}$ the language which results from extending $\mathcal{L}_0$ with operators for explicit shared belief and explicit common belief. It is defined as follows:

$$\mathcal{L}_0^{\triangledown_J} \stackrel{\mathsf{def}}{=} \quad \alpha \quad ::= \quad p \mid \neg\alpha \mid \alpha_1 \wedge \alpha_2 \mid \triangle_i\alpha \mid \triangle_J\alpha \mid \triangledown_J\alpha,$$

with p ranging over Atm and i ranging over Agt. Formula $\triangle_J\alpha$ has to be read "the agents in J share the explicit belief that α", whereas $\triangledown_J\alpha$ has to be read "the agents in J have the explicit common belief that α". The following language $\mathcal{L}^{\triangledown_J}$ extends language $\mathcal{L}_0^{\triangledown_J}$ by implicit individual belief operators:

$$\mathcal{L}^{\triangledown_J} \stackrel{\mathsf{def}}{=} \quad \varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i\varphi,$$

with α ranging over $\mathcal{L}_0^{\triangledown_J}$ and i ranging over Agt.

In order to interpret the new language $\mathcal{L}_0^{\triangledown_J}$, we need to slightly redefine the notion of state of Definition 1 by assuming that an agent i's belief base is now a set of formulas from $\mathcal{L}_0^{\triangledown_J}$.

The explicit shared belief and explicit common belief operators are interpreted relative to a state $B = \big((B_i)_{i\in Agt}, V\big)$ so redefined, as follows:

$$B \models \triangle_J\alpha \iff \forall i \in J : \alpha \in B_i,$$
$$B \models \triangledown_J\alpha \iff \forall i \in J : \alpha \in B_i \text{ and } \triangledown_J\alpha \in B_i.$$

Notice, in particular, the semantic interpretation of the explicit common belief operator: explicit common belief that α is the same as everyone having the explicit belief that α and that there is explicit common belief that α.

Clearly, implicit common belief does not necessarily imply explicit common belief. Indeed, for every $J \in 2^{Agt*}$, there exists $(B, Cxt) \in \mathbf{M}$ such that:

$$(B, Cxt) \models \Box_J^k p \text{ for every } k \in \mathbb{N}^* \text{ and } (B, Cxt) \not\models \triangledown_J p.$$

An example of such a MAB is the following: $Cxt = \{B, B'\}$, $B = \big((B_i)_{i\in Agt}, V\big)$ and $B' = \big((B_i)_{i\in Agt}, V'\big)$, where $B_i = \{q, q \to p\}$ for every $i \in Agt$, $V = \{q\}$ and $V' = \{p, q\}$. On the contrary, explicit common belief necessarily implies implicit common belief, as indicated by the following validity, for every $J \in 2^{Agt*}$:

$$\models \triangledown_J\alpha \to \Box_J^k\alpha \text{ for every } k \in \mathbb{N}^*.$$

Moreover, explicit common belief implies implicit common belief about explicit shared belief, as well as implicit common belief about explicit common belief. For every $J \in 2^{Agt*}$, we have

$$\models \triangledown_J\alpha \to \Box_J^k\triangle_J\alpha \text{ for every } k \in \mathbb{N}^*,$$
$$\models \triangledown_J\alpha \to \Box_J^k\triangledown_J\alpha \text{ for every } k \in \mathbb{N}^*.$$

Like implicit common belief, the explicit common belief operator can alternatively be interpreted relative to notional doxastic models (NDMs) introduced in Section 3.2. To do so, we simply need to replace the semantic interpretation for the modal operator $\Box_J^*$ in Definition 9 by the following semantic interpretation for the modal operators $\triangle_J$ and $\triangledown_J$:

$$(M, w) \models \triangle_J\alpha \iff \forall i \in J : \alpha \in \mathcal{D}(i, w),$$
$$(M, w) \models \triangledown_J\alpha \iff \forall i \in J : \alpha \in \mathcal{D}(i, w) \text{ and } \triangledown_J\alpha \in \mathcal{D}(i, w),$$

where (M, w) is a pointed NDM.

The following theorem is the counterpart of Theorem 1 for the language of explicit common belief.

Theorem 4. *Let $\varphi \in \mathcal{L}^{\triangledown_J}$. Then, the following five statements are equivalent:*

(1) φ is satisfiable relative to class **QNDM** *(resp.* $\mathbf{QNDM}_{BC}$*),*

*(2) φ is satisfiable relative to class finite-***QNDM** *(resp. finite-***QNDM**$_{BC}$*),*
*(3) φ is satisfiable relative to class finite-***NDM** *(resp. finite-***NDM**$_{BC}$*),*
(4) φ is satisfiable relative to class **NDM** *(resp.* **NDM**$_{BC}$*),*
(5) φ is satisfiable relative to class **M** *(resp.* **M**$_{BC}$*).*

Sketch of Proof. The proof is similar to that of Theorem 1. ■

4.2 Axiomatics

The following definition introduces the logic LDA–ECB (Logic of Doxastic Attitudes and Explicit Common Belief) and its variant LDA–ECB$_{T_{\Box_i}}$ for correct beliefs.

Definition 11 (LDA–ECB). *LDA–ECB is the extension of LDA defined in Definition 8 by the following axioms:*

$$\triangle_J\alpha \leftrightarrow \bigwedge_{i\in J} \triangle_i\alpha \qquad (\mathbf{Def}_{\triangle_J})$$

$$\nabla_J\alpha \leftrightarrow (\triangle_J\alpha \wedge \triangle_J\nabla_J\alpha) \qquad (\mathbf{Refl}_{\nabla_J})$$

The logic LDA–ECB$_{T_{\Box_i}}$ extends LDA–ECB by Axiom $\boldsymbol{T}_{\Box_i}$ of Definition 8.

Axiom $\mathbf{Def}_{\triangle_J}$ defines explicit shared belief from the explicit individual belief of all agents. Axiom $\mathbf{Refl}_{\nabla_J}$ is the "publicity" axiom for explicit common belief. Definitions of LDA–ECB-deducibility, LDA–ECB$_{T_{\Box_i}}$-deducibility, LDA–ECB-consistency, LDA–ECB$_{T_{\Box_i}}$-consistency, strong and weak completeness for LDA–ECB and LDA–ECB$_{T_{\Box_i}}$ are analogous to the ones for logics LDA–ICB and LDA–ICB$_{T_{\Box_i}}$ given in Section 3.3. Unlike logics LDA–ICB and LDA–ICB$_{T_{\Box_i}}$, logics LDA–ECB and LDA–ECB$_{T_{\Box_i}}$ satisfy the compactness property and, as the following theorem indicates, are strongly complete.

Theorem 5. *The logic LDA–ECB is sound and strongly complete for the class* **M***, and the logic LDA–ECB$_{T_{\Box_i}}$ is sound and strongly complete for the class* **M**$_{BC}$*.*

Sketch of Proof. Soundness is a routine exercise. Thanks to Theorem 4, it is sufficient to prove completeness for the class **QNDM** (resp. **QNDM**$_{BC}$). Unlike the proof of Theorem 5, we prove the latter by standard canonical model argument: worlds in the canonical model are (infinite) maximally consistent sets of formulas. ■

4.3 Complexity

It is easy to show that satisfiability checking for formulas in $\mathcal{L}^{\nabla_J}$ relative to model classes **M** and **M**$_{BC}$ is PSPACE-hard. Indeed, LDA–ECB is a conservative extension of the multi-modal logic K^n, while LDA–ICB$_{T_{\Box_i}}$ is a conservative extension of the multi-modal logic KT^n. These two logics are known to be PSPACE-hard, even for the case $n = 1$ [20]. As the following theorem indicates, PSPACE is also a upper bound.

Theorem 6. *Satisfiability checking for formulas in $\mathcal{L}^{\nabla_J}$ relative to the class* **M** *(resp.* **M**$_{BC}$*) is in PSPACE.*

Sketch of Proof. We can use the tableau method. To check the satisfiablity of a given formula φ, the goal is to construct a *tableau* (a concept similar to that of a quasi-NDM) which satisfies φ at its root. Then φ is satisfiable if and only if the construction can be completed. The construction algorithm can be executed with a quadratic memory in the size of the formula φ. ■

5 EXAMPLE

We illustrate the language $\mathcal{L}^{\nabla_J}$ with the help of an example of dichotomous coordination game in which agents can achieve their common goal only by making the same choice. In a coordination game, each coordination point is a Nash equilibrium and the general problem is to select one of them in order to ensure that the common goal is effectively achieved.

To this aim, we assume the set of atomic propositions *Atm* includes special atoms of type $pl_i[a]$ with $i \in Agt$ and $a \in Act$, where *Act* is a finite set of action symbols. Atom $pl_i[a]$ has to be read "agent i plays (or chooses) action a". The agents in *Agt* are said to play a dichotomous coordination game with common goal $\varphi_G \in \mathcal{L}^{\nabla_J}$ and action repertoire $X \subseteq Act$, noted $Coord(\varphi_G,X)$, if and only if (i) each agent will choose exactly one action from X, and (ii) the agents will achieve their common goal φ_G if and only if they coordinate by choosing the same action from X. In formal terms:

$$\mathrm{Coord}(\varphi_G,X) \stackrel{\mathsf{def}}{=} \big(\varphi_G \leftrightarrow \big(\bigvee_{a\in X}\bigwedge_{i\in Agt} pl_i[a]\big)\big) \wedge \bigwedge_{i\in Agt}\big(\big(\bigvee_{a\in X} pl_i[a]\big) \wedge \bigwedge_{a,b\in X: a\neq b}(pl_i[a] \rightarrow \neg pl_i[b])\big).$$

In our example, we suppose $Agt = \{1, 2\}$ and $\{mr, ml\} \subseteq Act$. Agents 1 and 2 are two mobile robots standing in front of a narrow passage on its opposite sides, as illustrated in Figure 1. An agent

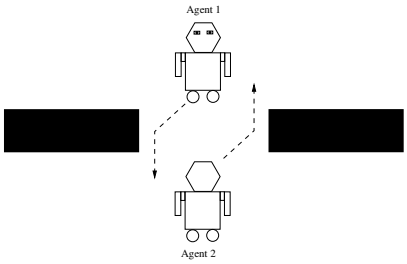

Figure 1: Example of coordination game

can either decide to move forward keeping to the right (action mr) or decide to move forward keeping to the left (action ml). The two agents can only coordinate by making the same choice. Indeed, if they choose different actions (e.g., agent 1 chooses mr while agent 2 chooses ml), they will miscoordinate and, consequently, collide ($coll$). We consider a number of hypotheses that agents 1 and 2 could share in the coordination scenario:

$$\alpha_1 \stackrel{\mathsf{def}}{=} coll \leftrightarrow \big((pl_1[mr] \wedge pl_2[ml]) \vee (pl_1[ml] \wedge pl_2[mr])\big),$$

$$\alpha_2 \stackrel{\mathsf{def}}{=} \bigwedge_{i\in Agt}(pl_i[mr] \vee pl_i[ml]),$$

$$\alpha_3 \stackrel{\mathsf{def}}{=} \bigwedge_{i\in Agt}(pl_i[mr] \rightarrow \neg pl_i[ml]).$$

According to hypothesis α_1, agents 1 and 2 will collide if and only if they choose different actions. Hypotheses α_2 and α_3 just state that an agent will choose exactly one action from $\{mr, ml\}$.

We can prove that explicit common belief of every hypothesis in $\{\alpha_1, \alpha_2, \alpha_3\}$ implies implicit common belief of facing a coordination problem. Indeed, for every $k \in \mathbb{N}^*$, we have:

$$\models_{\mathbf{M}} \Big(\bigwedge_{h\in\{1,2,3\}} \nabla_{Agt}\alpha_h \Big) \rightarrow \Box^k_{Agt} Coord(\neg coll, \{mr, ml\}). \quad (1)$$

In a coordination game like the one depicted in Figure 1, there are different ways to secure coordination. The bottom-up solution relies on repeated interaction: by playing the game several times the agents can learn to coordinate in a such a way that a social convention between them is established. According to Lewis' seminal theory [28], for a social convention to exist in a group of agents, the agents in the group must form a common belief that everyone in the group will conform to the convention by playing the corresponding action in the selected equilibrium. The top-down solution is by means of external norms that induce agents to play a specific equilibrium in the game. For example, in the case of the coordination game of Figure 1 the obligation to choose action mr could be enforced on agents 1 and 2 to secure coordination among them and to make the agents have a common belief of this, under the assumption that they are obedient and will comply with the obligation.

We are going to show how such a top-down mechanism can be formally represented in the language $\mathcal{L}^{\nabla_J}$. The first step in the analysis consists in assuming that the set of atomic propositions Atm includes special atoms of type $obed_i$ and $obl[a]$ for every $i \in Agt$ and $a \in Act$ that have to read, respectively, "agent i is norm compliant (or obedient)" and "it is obligatory to perform action a". As a second step, we define the following hypothesis which specifies the meaning of the notion of norm compliant agent:

$$\alpha_4 \stackrel{\mathsf{def}}{=} \bigwedge_{i,j\in Agt: i\neq j} \Big(obed_i \leftrightarrow \bigwedge_{a\in Act} \big((\triangle_i obl[a] \wedge \triangle_i obed_j) \rightarrow pl_i[a] \big) \Big).$$

According to hypothesis α_4, an agent is norm compliant (or obedient) if and only if, if it explicitly believes that the other agent is also norm compliant and that there is an obligation to perform action a, then it will choose action a. This captures a reciprocal form of obedience: an agent is willing to comply with the obligation only if it believes that the other agent is willing too.

For every $k \in \mathbb{N}^*$, we have the following two validities which highlight the conditions under which obligations determine coordination and common belief about coordination:

$$\models_{\mathbf{M}} \Big(\bigwedge_{i\in Agt} \nabla_{Agt} obed_i \wedge \nabla_{Agt}\alpha_4 \Big) \rightarrow \bigwedge_{a\in Act} \Box^k_{Agt} \big(\triangle_{Agt} obl[a] \rightarrow (pl_1[a] \wedge pl_2[a]) \big), \quad (2)$$

$$\models_{\mathbf{M}} \Big(\bigwedge_{i\in Agt} \nabla_{Agt} obed_i \wedge \bigwedge_{h\in\{1,3,4\}} \nabla_{Agt}\alpha_h \Big) \rightarrow \bigwedge_{a\in Act} \Box^k_{Agt} \big(\triangle_{Agt} obl[a] \rightarrow \neg coll \big). \quad (3)$$

According to validity (2) if the agents in the coordination game of Figure 1 have explicit common belief that each of them is norm compliant and have explicit common belief of what norm compliance means, as specified by hypothesis α_4, then they have implicit common belief that if each of them is aware that it is obligatory to perform a certain action then they will perform the obligatory action. According to validity (3) if the agents in the coordination game of Figure 1 have explicit common belief that each of them is norm compliant and have explicit common belief of each hypothesis in $\{\alpha_1, \alpha_3, \alpha_4\}$, then they have implicit common belief that if each of them is aware that it is obligatory to perform a certain action then they will not collide.

6 DYNAMIC EXTENSION

In this section, we extend the language $\mathcal{L}^{\nabla_J}$ by belief expansion operations. Such an extension allows us to represent private and public forms of information dynamics in a multi-agent domain. We name $\mathcal{L}^{\nabla_J, +_J}$ the resulting language and define it as follows:

$$\mathcal{L}^{\nabla_J, +_J} \stackrel{\mathsf{def}}{=} \quad \varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i\varphi \mid [+_J\alpha]\varphi,$$

where α ranges over $\mathcal{L}_0^{\nabla_J}$, i ranges over Agt and J ranges over 2^{Agt*}. The formula $[+_J\alpha]\varphi$ has to be read "φ holds after every agent in J has expanded her belief base with α". Events of type $+_J\alpha$ are generically named 'informative events'.

The dynamic operator $[+_J\alpha]$ has the following semantic interpretation relative to MABs of Definition 5. Let $B = \big((B_i)_{i\in Agt}, V\big)$ $\in \mathbf{S}$ and let $(B, Cxt) \in \mathbf{M}$. Then:

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

with $V^{+_J\alpha} = V$, $B_i^{+_J\alpha} = B_i \cup \{\alpha\}$ for $i \in J$, and $B_j^{+_J\alpha} = B_j$ for $j \notin J$.

Intuitively speaking, belief expansion by all agents in J with α simply consists in every agent in J adding the information α to her belief base, while all agents outside J keep their beliefs unchanged.

The following proposition provides reduction principles which allow us to transform every formula of the language $\mathcal{L}^{\nabla_J, +_J}$ into an equivalent formula of the language $\mathcal{L}^{\nabla_J}$.

Proposition 1. *The following formulas are valid for class* $\mathbf{M}$*:*

$$\begin{aligned}
&[+_J\alpha]p \leftrightarrow p\\
&[+_J\alpha]\neg\psi \leftrightarrow \neg[+_J\alpha]\psi\\
&[+_J\alpha](\psi_1 \wedge \psi_2) \leftrightarrow ([+_J\alpha]\psi_1 \wedge [+_J\alpha]\psi_2)\\
&[+_J\alpha]\triangle_i\beta \leftrightarrow \triangle_i\beta \text{ if } i \notin J \text{ or } \alpha \neq \beta\\
&[+_J\alpha]\triangle_i\alpha \leftrightarrow \top \text{ if } i \in J\\
&[+_J\alpha]\triangle_{J'}\beta \leftrightarrow \triangle_{J'}\beta \text{ if } \alpha \neq \beta\\
&[+_J\alpha]\triangle_{J'}\alpha \leftrightarrow \triangle_{J'\setminus J}\alpha\\
&[+_J\alpha]\nabla_{J'}\beta \leftrightarrow (\triangle_{J'}\beta \wedge \triangle_{J'\setminus J}\nabla_{J'}\beta) \text{ if } \alpha = \nabla_{J'}\beta\\
&[+_J\alpha]\nabla_{J'}\beta \leftrightarrow (\triangle_{J'\setminus J}\beta \wedge \triangle_{J'}\nabla_{J'}\beta) \text{ if } \alpha = \beta\\
&[+_J\alpha]\nabla_{J'}\beta \leftrightarrow \nabla_{J'}\beta \text{ if } \alpha \neq \nabla_{J'}\beta \text{ and } \alpha \neq \beta\\
&[+_J\alpha]\Box_i\varphi \leftrightarrow \Box_i\varphi \text{ if } i \notin J\\
&[+_J\alpha]\Box_i\varphi \leftrightarrow \Box_i(\alpha \rightarrow \varphi) \text{ if } i \in J
\end{aligned}$$

It is easy to define a a mapping *red* which iteratively applies the valid equivalences of Proposition 1 and transforms any formula φ in $\mathcal{L}^{\nabla_J, +_J}$ into an equivalent formula $red(\varphi)$ in $\mathcal{L}^{\nabla_J}$ of polynomial size.

As a consequence, thanks to Theorem 6, the following complexity result can be proved.

Theorem 7. *Satisfiability checking for formulas in $\mathcal{L}^{\nabla_J,+_J}$ relative to the class* **M** *is PSPACE-complete.*

Let us briefly explore the modeling aspects of the language $\mathcal{L}^{\nabla_J,+_J}$. In this language, we can represent a variety of private and public information dynamics. Before defining them, we introduce the following abbreviation:

$$[+_{J_1}\alpha_1;\ldots;+_{J_k}\alpha_k]\varphi \stackrel{\text{def}}{=} [+_{J_1}\alpha_1]\ldots[+_{J_k}\alpha_k]\varphi,$$

where $[+_{J_1}\alpha_1;\ldots;+_{J_k}\alpha_k]\varphi$ means "φ holds after the informative events $+_{J_1}\alpha_1,\ldots,+_{J_k}\alpha_k$ take place consecutively". The following validity captures a commutativity property for informative events:

$$\models_{\mathbf{M}} [+_{J_1}\alpha_1;\ldots;+_{J_k}\alpha_k]\varphi \leftrightarrow [\sigma(+_{J_1}\alpha_1);\ldots;\sigma(+_{J_k}\alpha_k)]\varphi, \quad (4)$$

where σ is any permutation of the set $\{+_{J_1}\alpha_1,\ldots,+_{J_k}\alpha_k\}$. This property is a consequence of the semantics of informative events based on belief expansion. Since events $+_{J_1}\alpha_1,\ldots,+_{J_k}\alpha_k$ simply expand the belief bases of the agents in $J_1,\ldots,J_k$, the order in which they occur does not matter. This is different from public announcement logic PAL [38] in which the order of announcements can play a role. The following are three examples of informative events of private, semi-private and public form:

$$\begin{aligned}
&\mathrm{priv}(J,\alpha) \stackrel{\text{def}}{=} +_J\alpha,\\
&\mathrm{publ}(J,\alpha) \stackrel{\text{def}}{=} +_J\alpha;+_J\nabla_J\alpha,\\
&\mathrm{semiPriv}(J,J',\alpha) \stackrel{\text{def}}{=} +_J\alpha;+_{J'}(\triangle_J\alpha\vee\triangle_J\neg\alpha).
\end{aligned}$$

The event $\mathrm{priv}(J,\alpha)$ consists in every agent in J privately learning that α, whereas $\mathrm{publ}(J,\alpha)$ means that all agents in J publicly learn that α. Finally, $\mathrm{semiPriv}(J,J',\alpha)$ means that all agents in J privately learn that α, while the agents in J' learn that the agents in J has just learnt whether α is true. An example of the latter is the situation of Ann (A), Bob (B), Mary (M) and Paul (P) sitting at a table on which there is an empty box. In the initial situation, they do not know whether the box is empty since they cannot see what is inside it. Mary and Paul look inside the box in front of everybody. Therefore, Mary and Paul privately learn that the box is empty. Moreover, Ann, Bob, Mary and Paul learn that Mary and Paul have just learnt whether the box is empty. The latter event is represented by $\mathrm{semiPriv}(\{M,P\},\{A,B,M,P\},\mathit{boxEmpty})$. Note that $\mathrm{publ}(J,\alpha)$ and $\mathrm{semiPriv}(J,J',\alpha)$ are definable from $\mathrm{priv}(J,\alpha)$. Specifically, $\mathrm{publ}(J,\alpha)$ is the same as $\mathrm{priv}(J,\alpha)$ followed by $\mathrm{priv}(J,\nabla_J\alpha)$ and $\mathrm{semiPriv}(J,J',\alpha)$ is the same as $\mathrm{priv}(J,\alpha)$ followed by $\mathrm{priv}(J',\triangle_J\alpha\vee\triangle_J\neg\alpha)$. In other words, publicity and semi-privateness are defined from privateness. For instance, publicly learning that α corresponds to the fact that every agent in α privately learns that α and that α has become common belief. The following are interesting properties of private, semi-private and public informative events:

$$\models_{\mathbf{M}} [\mathrm{priv}(J,\alpha)]\bigwedge_{i\in J}\Box_i\alpha, \quad (5)$$

$$\models_{\mathbf{M}} [\mathrm{semiPriv}(J,J',\alpha)]\bigwedge_{i\in J,j\in J'}\big(\Box_i\alpha\wedge\Box_j(\Box_i\alpha\vee\Box_i\neg\alpha)\big), \quad (6)$$

$$\models_{\mathbf{M}} [\mathrm{publ}(J,\alpha)]\Box_J^k\alpha \text{ for every } k\in\mathbb{N}^*, \quad (7)$$

$$\models_{\mathbf{M}} [\mathrm{publ}(J,\alpha)]\Box_J^k\triangle_J\alpha \text{ for every } k\in\mathbb{N}^*, \quad (8)$$

$$\models_{\mathbf{M}} [\mathrm{publ}(J,\alpha)]\Box_J^k\nabla_J\alpha \text{ for every } k\in\mathbb{N}^*. \quad (9)$$

Note in particular validities (7), (8) and (9): after having publicly learnt that α, the agents will have the implicit common belief of α, that they share the belief of α and that they have the explicit common belief of α. For example, going back to the example of Section 5, we have the following validity, for every $k\in\mathbb{N}^*$:

$$\models_{\mathbf{M}}\Big(\bigwedge_{i\in Agt}\nabla_{Agt}obed_i\wedge\nabla_{Agt}\alpha_4\Big)\rightarrow \big[\mathrm{publ}\big(Agt,obl[a]\big)\big]\Box_{Agt}^k(pl_1[a]\wedge pl_2[a]). \quad (10)$$

This guarantees that if the agents in the coordination game of Figure 1 have explicit common belief that each of them is norm compliant and of what norm compliance means then, by publicly learning that action a is obligatory, they form the implicit common belief that each of them will perform the obligatory action.

7 CONCLUSION

Let's take stock. We have studied two logics of collective attitudes: the logic of implicit common belief and the logic of explicit common belief. We have provided axiomatic and complexity results for both logics. While satisfiability checking for the former logic is EXPTIME-hard, it is in PSPACE for the latter. Our complexity result relies on a tableau method. Future work will be devoted to find a polysize reduction of satisfiability checking for the logic of explicit common belief to QBF. This will open up the possibility of exploiting existing efficient QBF solvers for automated reasoning about collective attitudes in multi-agent systems. We also plan to explore the connection between the dynamic extension we presented in Section 6 and dynamic epistemic logic (DEL), namely the extension of epistemic logic by so-called event models [5, 42]. In particular, we would like to precisely characterize the subclass of event models that are captured by our update semantics. Since full DEL is known to be NEXPTIME-hard [1], we know for sure that the latter is necessarily a strict subclass. In this paper, we have studied the logic of implicit common belief and the logic of explicit common belief separately. In future work, we will explore the axiomatic properties and complexity of the unified language including both implicit and explicit common belief. Our conjecture is that satisfiability checking for such rich epistemic language remains in EXPTIME.

ACKNOWLEDGMENTS

This work is supported by the ANR project CoPains ("Cognitive Planning in Persuasive Multimodal Communication"). Support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute is also acknowledged.

REFERENCES

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

[2] R. Aumann. 1976. Agreeing to Disagree. *Annals of Statistics* 4, 6 (1976), 1236–1239.

[3] R. Aumann. 1995. Backward induction and common knowledge of rationality. *Games and Economic Behavior* 8, 1 (1995), 6–19.

[4] M. Bacharach. 1985. Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory* 37, 1 (1985), 167–190.

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

[6] C. Bicchieri. 1989. Self-refuting theories of strategic interaction: A paradox of common knowledge. *Erkenntnis* 30, 1-2 (1989), 69–85.

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

[8] G. Bonanno. 1996. On the Logic of Common Belief. *Games and Economic Behavior* 42, 1 (1996), 305–311.

[9] G. Bonanno. 2008. A syntactic approach to rationality in games with ordinal payoffs. In *Proceedings of LOFT 2008 (Texts in Logic and Games Series)*. Amsterdam University Press, 59–86.

[10] M. Bratman. 1992. Shared cooperative activity. *The Philosophical Review* 101, 2 (1992), 327–41.

[11] H. H. Clark and C. Marshall. 1981. Definite reference and mutual knowledge. In *Elements of Discourse Understanding*, A. Joshi, B. Webber, and I. Sag (Eds.). Cambridge University Press, 10–63.

[12] M. Colombetti. 1999. A modal logic of intentional communication. *Mathematical Social Sciences* 38, 2 (1999), 171–196.

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

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

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

[16] M. F. Friedell. 1969. On the Structure of Shared Awareness. *Behavioral Science* 14, 1 (1969), 28–39.

[17] S. Fukuda. 2020. Formalizing common belief with no underlying assumption on individual beliefs. *Games and Economic Behavior* 121 (2020), 169–189.

[18] B. J. Grosz and S. Kraus. 1996. Collaborative Plans for Complex Group Action. *Artificial Intelligence* 86 (1996), 269–357.

[19] J. Y. Halpern and Y. Moses. 1985. A guide to the modal logics of knowledge and belief: Preliminary draft. In *Proceedings of IJCAI'85*. Morgan Kaufmann, 480–490.

[20] J. Y. Halpern and Y. Moses. 1992. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence* 54, 3 (1992), 319–379.

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

[22] A. Heifetz. 1999. Iterative and fixed point common belief. *Journal of Philosophical Logic* 28, 1 (1999), 61–79.

[23] A. Herzig and E. Perrotin. 2020. On the Axiomatisation of Common Knowledge. In *Proceedings of the 13th Conference on Advances in Modal Logic (AiML 2020)*. College Publications, 309–328.

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

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

[26] D. Lehmann. 1984. Knowledge, common knowledge and related puzzles (Extended Summary). In *Proceedings of the third annual ACM symposium on Principles of distributed computing (PODC'84)*. Morgan Kaufmann, 62–67.

[27] H. J. Levesque, P. R. Cohen, and J. H. T. Nunes. 1990. On Acting Together. In *Proceedings of the 8th National Conference on Artificial Intelligence (AAAI-90)*. AAAI Press / The MIT Press, 94–99.

[28] D. K. Lewis. 1969. *Convention: a philosophical study*. Harvard University Press, Cambridge.

[29] L. Lismont and P. Mongin. 1994. On the Logic of Common Belief and Common Knowledge. *Theory and Decision* 37 (1994), 75–106.

[30] C. List. 2014. Three kinds of collective attitudes. *Erkenntnis* 79, 9 (2014), 1601–1622.

[31] E. Lorini. 2016. A minimal logic for interactive epistemology. *Synthese* 193, 3 (2016), 725–755.

[32] E. Lorini. 2018. In Praise of Belief Bases: Doing Epistemic Logic Without Possible Worlds. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*. AAAI Press, 1915–1922.

[33] E. Lorini. 2020. Rethinking epistemic logic with belief bases. *Artificial Intelligence* 282 (2020).

[34] E. Lorini and F. Romero. 2019. Decision Procedures for Epistemic Logic Exploiting Belief Bases. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2019)*. IFAAMAS, 944–952.

[35] J. J. Meyer and W. van der Hoek. 1995. *Epistemic logic for AI and computer science*. Cambridge University Press.

[36] E. Pacuit. 2017. *Neighborhood Semantics for Modal Logic*. Springer.

[37] A. Perea. 2012. *Epistemic game theory: reasoning and choice*. Cambridge University Press.

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

[39] S. Schiffer. 1972. *Meaning*. Clarendon Press, Oxford.

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

[41] R. Stalnaker. 2002. Common ground. *Linguistics and philosophy* 25, 5/6 (2002), 701–721.

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

[43] P. Vanderschraaf and G. Sillari. 2011. Common knowledge. *Stanford Encyclopedia of Philosophy* (2011). https://plato.stanford.edu/entries/common-knowledge/