



HAL
open science

Base-Based Model Checking for Multi-agent only Believing

Tiago de Lima, Emiliano Lorini, François Schwarzentruher

► **To cite this version:**

Tiago de Lima, Emiliano Lorini, François Schwarzentruher. Base-Based Model Checking for Multi-agent only Believing. 18th European Conference on Logics in Artificial Intelligence (JELIA 2023), Sep 2023, Dresden, Germany. pp.437-445, 10.1007/978-3-031-43619-2_30 . hal-04268823

HAL Id: hal-04268823

<https://univ-artois.hal.science/hal-04268823v1>

Submitted on 2 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Base-based Model Checking for Multi-Agent Only Believing

Tiago de Lima¹, Emiliano Lorini², and François Schwarzentruber³

¹ CRIL, Univ Artois and CNRS, Lens, France, delima@cril.fr

² IRIT, CNRS, Toulouse University, France, lorini@irit.fr

³ ENS Rennes, France, francois.schwarzentruber@ens-rennes.fr

Abstract. We present a novel semantics for the language of multi-agent only believing exploiting belief bases, and show how to use it for automatically checking formulas of this language and of its dynamic extension with private belief expansion operators. We provide a PSPACE algorithm for model checking relying on a reduction to QBF and alternative dedicated algorithm relying on the exploration of the state space. We present an implementation of the QBF-based algorithm and some experimental results on computation time in a concrete example.

1 Introduction

The idea of using belief bases for building a semantics for epistemic logic was initially proposed by Lorini [17, 19]. In [18] it was shown that such a semantics allows to represent the concept of universal epistemic model which is tightly connected with the concept of universal type space studied by game theorists [20]. A qualitative version of the universal type space with no probabilities involved is defined by Fagin et al. [6] (see also [7]). Broadly speaking, a universal epistemic model for a given situation is the most general model which is compatible with that situation. It is the model which only contains information about the situation and makes no further assumption. From an epistemic point view, it can be seen as the model with maximal ignorance with respect to the description of the situation at stake.

Such a universal epistemic model has been shown to be crucial for defining a proper semantics for the concept of multi-agent only knowing (or believing) [14, 12], as a generalization of the concept of single-agent only knowing (or believing) [15].⁴ However, the construction of this semantics is far from being straightforward. Halpern & Lakemeyer [13] use the proof-theoretic notion of canonical model for defining it. The limitation of the canonical model is its being infinite thereby not being exploitable in practice. In a more recent work, Belle & Lakemeyer [2] provided an inductive proof-independent definition of the semantics for multi-agent only knowing which departs from the standard semantics of multi-agent epistemic logic based on multi-relational Kripke structures. Finally,

⁴ As usual, the difference between knowledge and belief lies in the fact that the former is always correct while the latter can be incorrect.

Aucher & Belle [1] have shown how to interpret the language of multi-agent only knowing on standard Kripke structures. Although being independent from the proof theory, these last two accounts are fairly non-standard or quite involved. They rely either on an inductive definition (Belle & Lakemeyer) or on a complex syntactic representation up to certain modal depth (Aucher & Belle) of the multi-agent epistemic structure used for interpreting the multi-agent only knowing language.

In this paper, we concentrate on the logic of multi-agent only believing based on the logic K for beliefs. We show how to use the belief base semantics and its construction of the universal model to automatically check formulas of the corresponding language. The novel contribution of the paper is twofold.

- Although the idea of using belief bases as a semantics for epistemic logic has been proposed in previous work, this is the first attempt to use them in the context of the logic of multi-agent only believing and of its extension with private belief expansion operators.
- Moreover, we are the first to provide a model checking algorithm for the logic of multi-agent only believing, to implement it and to test it experimentally on a concrete example. The belief base semantics helped us to accomplish this task given its compactness and manageability.

The paper is organized as follows. In Section 2, we first recall the belief base semantics introduced in our previous work [17, 19]. We show how to interpret the language of multi-agent only believing and how to define the universal model in it. In Section 3, we introduce an example to illustrate the framework. In Section 4, we move to model checking formulated in the belief base semantics. We provide a PSPACE algorithm for model checking relying on a reduction to QBF and an alternative dedicated algorithm relying on the exploration of the state space. In Section 5, we present an implementation of the QBF-based algorithm and some experimental results on computation time in the example. In Section 6 we propose an extension of the setting with private belief expansion operators, and demonstrate that the model checking problem remains in PSPACE. Section 7 concludes the paper.

2 Language and semantics

The multi-agent epistemic language introduced in [19] has two basic epistemic modalities: one for explicit belief, and another one for implicit belief. An agent’s explicit belief corresponds to a piece of information in the agent’s belief base. An agent’s implicit belief corresponds to a piece of information that is derivable from the agent’s explicit beliefs. In other words, if an agent can derive φ from its explicit beliefs, it implicitly believes *at least* that φ is true. We consider the extension of this epistemic language by complementary modalities for implicitly believing *at most*. The *at least* and *at most* modalities can be combined to represent the concept of *only* believing.

The semantics over which the language is interpreted exploits belief bases. Unlike the standard multi-relational Kripke semantics for epistemic logic in which the agents' epistemic accessibility relations over possible worlds are given as primitive, in this semantics they are computed from the agents' belief bases. Specifically, in this semantics it is assumed that at state S an agent considers a state S' possible (or state S' is epistemically accessible to the agent at state S) if and only if S' satisfies all formulas that are included in the agent's belief base at S . This idea of computing the agents' accessibility relations from the state description is shared with the semantics of epistemic logic based on interpreted systems [8, 16]. However, there is an important difference. While the interpreted system semantics relies on the abstract notion of an agent's local state, in the belief base semantics an agent's local state is identified with its concrete belief base.

2.1 Semantics

Assume a countably infinite set of atomic propositions $Atm = \{p, q, \dots\}$ and a finite set of agents $Agt = \{1, \dots, n\}$. We define the language \mathcal{L}_0 for explicit belief by the following grammar in Backus-Naur Form (BNF):

$$\mathcal{L}_0 \stackrel{\text{def}}{=} \alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \Delta_i\alpha,$$

where p ranges over Atm and i ranges over Agt . \mathcal{L}_0 is the language used to represent explicit beliefs. The formula $\Delta_i\alpha$ reads "agent i has the explicit belief that α ". In our semantics, a state is not a primitive notion but it is decomposed into different elements: one belief base per agent and an interpretation of propositional atoms.

Definition 1 (State). *A state is a tuple $S = ((B_i)_{i \in Agt}, V)$ 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 \mathcal{L}_0 .

Definition 2 (Satisfaction relation). *Let $S = ((B_i)_{i \in Agt}, V) \in \mathbf{S}$. Then,*

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

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 its belief base.

The following definition introduces the agents' epistemic relations. They are computed from the agents' belief bases.

Definition 3 (Epistemic relation). *Let $i \in \text{Agt}$. Then, \mathcal{R}_i is the binary relation on \mathbf{S} such that, for all $S = ((B_i)_{i \in \text{Agt}}, V), S' = ((B'_i)_{i \in \text{Agt}}, V') \in \mathbf{S}$:*

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

$S \mathcal{R}_i S'$ means that S' is an epistemic alternative for agent i at S , that is to say, S' is a state that at S agent i considers possible. The idea of the previous definition is that S' is an epistemic alternative for agent i at S if and only if, S' satisfies all facts that agent i explicitly believes at S .

The following definition introduces the concept of model, namely a state supplemented with a set of states, called *context*. The latter includes all states that are compatible with the agents' common ground, i.e., the body of information that the agents commonly believe to be the case [21].

Definition 4 (Model). *A model is a pair (S, Cxt) with $S \in \mathbf{S}$ and $\text{Cxt} \subseteq \mathbf{S}$. The class of models is noted \mathbf{M} .*

Note that in a model (S, Cxt) , the state S is not necessarily an element of the context Cxt due to the fact that we model belief instead of knowledge. Therefore, the agents' common ground represented by the context Cxt may be incorrect and not include the actual state. If we modeled knowledge instead of belief, we would have to suppose that $S \in \text{Cxt}$.

Let $\Gamma = (\Gamma_i)_{i \in \text{Agt}}$ where, for every $i \in \text{Agt}$, Γ_i represents agent i 's vocabulary. A Γ -universal model is a model containing all states at which an agent i 's explicit beliefs are built from its vocabulary Γ_i . In other words, an agent's vocabulary plays a role analogous to that of the notion of awareness in the formal semantics of awareness [9]. The notion of Γ -universal model is defined as follows.

Definition 5 (Universal model). *The model (S, Cxt) in \mathbf{M} is said to be Γ -universal if $S \in \text{Cxt} = \mathbf{S}_\Gamma$ with:*

$$\mathbf{S}_\Gamma = \left\{ ((B'_i)_{i \in \text{Agt}}, V') \in \mathbf{S} \mid \forall i \in \text{Agt}, B'_i \subseteq \Gamma_i \right\}.$$

The class of Γ -universal models is noted $\mathbf{M}_{\text{univ}}(\Gamma)$.

$\Gamma = (\Gamma_i)_{i \in \text{Agt}}$ is also called agent vocabulary profile. Clearly, when $\Gamma = \mathcal{L}_0^n$, we have $\mathbf{S}_\Gamma = \mathbf{S}$. A model (S, \mathbf{S}) in $\mathbf{M}_{\text{univ}}(\mathcal{L}_0^n)$ is a model with maximal ignorance: it only contains the information provided by the actual state S . For simplicity, we write \mathbf{M}_{univ} instead of $\mathbf{M}_{\text{univ}}(\mathcal{L}_0^n)$.

2.2 Language

In this section, we introduce a language for implicitly believing *at most* and implicitly believing *at least* on the top of the language \mathcal{L}_0 defined above. It is noted \mathcal{L} and defined by:

$$\mathcal{L} \stackrel{\text{def}}{=} \varphi ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i \varphi \mid \Box_i^{\mathbb{G}} \varphi,$$

where α ranges over \mathcal{L}_0 and i ranges over Agt . The other Boolean constructions \top , \perp , \vee , \rightarrow , \oplus , and \leftrightarrow are defined from α , \neg and \wedge in the standard way. The formula $\Box_i\varphi$ is read “agent i at least implicitly believes that φ ”, while $\Box_i^{\mathbb{C}}\varphi$ is read “agent i at most implicitly believes that $\neg\varphi$ ”. Alternative readings of formulas $\Box_i\varphi$ and $\Box_i^{\mathbb{C}}\varphi$ are, respectively, “ φ is true at all states that agent i considers possible” and “ φ is true at all states that agent i does not consider possible”. The latter is in line with the reading of the normal modality and the corresponding “window” modality in the context of boolean modal logics [10]. The duals of the operators \Box_i and $\Box_i^{\mathbb{C}}$ are defined in the usual way, as follows: $\Diamond_i\varphi \stackrel{\text{def}}{=} \neg\Box_i\neg\varphi$ and $\Diamond_i^{\mathbb{C}}\varphi \stackrel{\text{def}}{=} \neg\Box_i^{\mathbb{C}}\neg\varphi$. Formulas in the language \mathcal{L} are interpreted relative to a model (S, Cxt) . (Boolean cases are omitted since they are defined as usual.)

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

$$\begin{aligned} (S, Cxt) \models \alpha &\iff S \models \alpha, \\ (S, Cxt) \models \Box_i\varphi &\iff \forall S' \in Cxt : \text{if } S\mathcal{R}_iS' \text{ then } (S', Cxt) \models \varphi, \\ (S, Cxt) \models \Box_i^{\mathbb{C}}\varphi &\iff \forall S' \in Cxt : \text{if } S\mathcal{R}_i^{\mathbb{C}}S' \text{ then } (S', Cxt) \models \varphi, \end{aligned}$$

with $\mathcal{R}_i^{\mathbb{C}} = (\mathbf{S} \times \mathbf{S}) \setminus \mathcal{R}_i$.

Note that $S\mathcal{R}_i^{\mathbb{C}}S'$ just means that at state S agent i does not consider state S' possible. Moreover, interpretations of the two modalities \Box_i and $\Box_i^{\mathbb{C}}$ are restricted to the actual context Cxt . The only believing modality (\Box_i°) and the universal modality (\mathbf{U}) are defined as follows:

$$\Box_i^{\circ}\varphi \stackrel{\text{def}}{=} \Box_i\varphi \wedge \Box_i^{\mathbb{C}}\neg\varphi.$$

Notions of satisfiability and validity of \mathcal{L} -formulas for the class of models \mathbf{M} are defined in the usual way: φ is satisfiable if there exists $(S, Cxt) \in \mathbf{M}$ such that $(S, Cxt) \models \varphi$, φ is valid if $\neg\varphi$ is not satisfiable.

In [18, Theorem 26], it is shown that when restricting to the fragment of the language \mathcal{L} where formulas containing explicit beliefs are disallowed (i.e., in the definition of \mathcal{L} α is replaced by p), the set of satisfiable formulas relative to the class \mathbf{M}_{univ} is the same as the set of satisfiable formulas relative to the class of qualitative universal type spaces, as defined in [6]. The latter is similar to the class of k -structures, as defined by Belle & Lakemeyer [2].⁵ In Section 4, we will show that Γ -universal models of Definition 5 provide an adequate and compact semantics for model checking formulas of the language \mathcal{L} . But, before delving into model checking, we illustrate our language with the help of an example.

3 Example

We give two variants of the example, the first focused on first-order beliefs and the second focused on second-order beliefs.

⁵ Although it has not been formally proven, we believe that Belle & Lakemeyer’s semantics and Fagin et al.’s semantics are nothing but different formulations of the same class of universal epistemic structures.

Example 1. Agents in Agt are members of a selection committee for an associate professor position. They have to choose which candidates to admit to the second round of selection consisting in an interview. Committee members and candidates work in the same scientific community. Therefore, it is possible that they co-authored some papers in the past. Assume there are m candidates $Cand = \{c_1, \dots, c_m\}$. In order to formalize the example we use atomic propositions of the form $\text{vote}(i, c)$, with $i \in Agt$ and $c \in Cand$, standing for “agent i votes for candidate c ”. The first two rules of the game state that each committee member must vote for exactly one candidate (at least one candidate and no more than one):

$$\alpha_1 \stackrel{\text{def}}{=} \bigwedge_{i \in Agt} \bigvee_{c \in Cand} \text{vote}(i, c),$$

$$\alpha_2 \stackrel{\text{def}}{=} \bigwedge_{i \in Agt} \bigwedge_{c, c' \in Cand, c \neq c'} (\text{vote}(i, c) \rightarrow \neg \text{vote}(i, c')).$$

The third rule states that a member of the committee cannot vote for a candidate with whom she/he co-authored an article in the past:

$$\alpha_3 \stackrel{\text{def}}{=} \bigwedge_{i \in Agt} \bigwedge_{c \in \mathbf{f}(i)} \neg \text{vote}(i, c),$$

where $\mathbf{f} : Agt \rightarrow 2^{Cand}$ is a function mapping each member of the committee to her/his co-authors. A candidate c is admitted to the interview if and only if at least one member of the committee has voted for her/him. This is expressed by the following abbreviation:

$$\text{adm}(c) \stackrel{\text{def}}{=} \bigvee_{i \in Agt} \text{vote}(i, c).$$

Let us consider the variant of the example in which the evaluation committee and the set of candidates have the same cardinality and a committee member co-authored an article with only her/his matching candidate in the linear order. That is, we suppose:

$$|Agt| = |Cand| > 2, \text{ and}$$

$$\forall i \in Agt, \mathbf{f}(i) = \{c_i\}.$$

Furthermore, we suppose that (i) each committee member except the last one votes for her/his next candidate in the linear order, while the last committee member n votes for her/his previous candidate $n-1$; (ii) the vote by a committee member is secret (i.e., a committee member only has epistemic access to her/his vote), (iii) all committee members know the results of the selection, namely, which candidates are admitted to the interview and which are not. The three hypotheses (i), (ii) and (iii) are fully expressed by the state $S_0 = ((B_i)_{i \in Agt}, V)$ such that, for every $1 \leq i < n$,

$$B_i = \{\text{vote}(i, c_{i+1}), \neg \text{adm}(c_1), \text{adm}(c_2), \dots, \text{adm}(c_n), \alpha_1, \alpha_2, \alpha_3\},$$

and moreover,

$$\begin{aligned} B_n &= \{\text{vote}(n, c_{n-1}), \neg \text{adm}(c_1), \text{adm}(c_2), \dots, \text{adm}(c_n), \alpha_1, \alpha_2, \alpha_3\}, \\ V &= \{\text{vote}(1, c_2), \dots, \text{vote}(n-1, c_n), \text{vote}(n, c_{n-1})\}. \end{aligned}$$

The following holds when $|Agt| = |Cand| = 3$:

$$(S_0, \mathbf{S}_\Gamma) \models \varphi_0,$$

with

$$\begin{aligned} \varphi_0 &\stackrel{\text{def}}{=} \Box_1^o \psi_1 \wedge \bigwedge_{i \in \{2,3\}} \Box_i^o \psi_2, \\ \psi_1 &\stackrel{\text{def}}{=} \text{vote}(1, c_2) \wedge \neg \text{vote}(1, c_1) \wedge \neg \text{vote}(1, c_3) \\ &\quad \wedge \text{vote}(2, c_3) \wedge \neg \text{vote}(2, c_1) \wedge \neg \text{vote}(2, c_2) \\ &\quad \wedge \text{vote}(3, c_2) \wedge \neg \text{vote}(3, c_1) \wedge \neg \text{vote}(3, c_3), \\ \psi_2 &\stackrel{\text{def}}{=} \neg \text{vote}(1, c_1) \wedge (\text{vote}(1, c_2) \oplus \text{vote}(1, c_3)) \\ &\quad \wedge \text{vote}(2, c_3) \wedge \neg \text{vote}(2, c_1) \wedge \neg \text{vote}(2, c_2) \\ &\quad \wedge \text{vote}(3, c_2) \wedge \neg \text{vote}(3, c_1) \wedge \neg \text{vote}(3, c_3). \end{aligned}$$

and $\Gamma_i = B_i \cup \neg B_i$ for every $i \in Agt$, (where $\neg B_i$ is the set of negations of the formulas in B_i). (S_0, \mathbf{S}_Γ) so defined is nothing but a Γ -universal model in which the agents' vocabularies include all and only those formulas in their actual belief bases and their negations.

This means that, in the three-agent case, agent 1 only knows for whom an agent voted and for whom she/he did not vote, while agent 2 and agent 3 only know for whom they voted and for whom they did not vote, and that agent 1 voted either for 2 or for 3. Therefore, 2 and 3 do not know for whom 1 voted. Interestingly, when $|Agt| = |Cand| > 3$:

$$(S_0, \mathbf{S}_\Gamma) \not\models \varphi_0.$$

Example 2. It is worth to consider a variant of Example 1 in which agent 1 has higher-order explicit beliefs (i.e., explicit beliefs about other agents' explicit beliefs). Specifically, we consider a state $S'_0 = ((B'_i)_{i \in Agt}, V')$ such that,

$$B'_1 = B_1 \cup \{\Delta_2 \neg \text{adm}(c_1), \Delta_2 \text{adm}(c_2), \dots, \Delta_2 \text{adm}(c_n), \Delta_2 \alpha_1, \Delta_2 \alpha_2, \Delta_2 \alpha_3\}$$

and, for every $1 < i \leq n$:

$$\begin{aligned} B'_i &= B_i, \\ V' &= V, \end{aligned}$$

where B_i and V are defined as above. In other words, committee member 1 explicitly knows that committee member 2 explicitly knows the rules of the game as well as the results of the selection.

Interestingly, when $|Agt| = |Cand| = 3$, the following holds:

$$(S'_0, \mathbf{S}) \models \Box_2^o \psi_2 \wedge \Box_1 \Box_2 \psi_2 \wedge \neg \Box_1 \Box_2^o \psi_2.$$

In words, in the three-agent case, at S'_0 , committee member 2 only knows that ψ_2 , committee member 1 knows that 2 knows ψ_2 , but 1 does not know that 2 only knows that ψ_2 .

4 Model checking

The model checking problem is defined in our framework, as follows:

input: an agent vocabulary profile $\Gamma = (\Gamma_i)_{i \in Agt}$ with Γ_i finite for every $i \in Agt$, a finite state S_0 in \mathbf{S}_Γ , and a formula $\varphi_0 \in \mathcal{L}$;

output: yes if $(S_0, \mathbf{S}_\Gamma) \models \varphi_0$; no otherwise.

Remark 1. We suppose w.l.o.g. that outer most subformulas of φ_0 of the form $\Delta_i \alpha$ are such that $\alpha \in \Gamma_i$. If this is not the case for some subformulas $\Delta_i \alpha$, then the subformula $\Delta_i \alpha$ will be false anyway and can be replaced by \perp .

Direct PSPACE algorithm Figure 1 shows an algorithm $mc(S, \Gamma, \varphi)$ that checks whether $(S, \mathbf{S}_\Gamma) \models \varphi$. Note that \mathbf{S}_Γ is not computed explicitly, but implicitly represented by Γ . In the algorithm, states S are represented as vectors of bits indicating for all i , for each element α of Γ_i whether α belongs to i 's base in S or not. It also encodes the valuation over atomic propositions appearing in Γ and φ_0 . The states S manipulated by the algorithm are of size polynomial in the size of the input. The loop **for** all $S' \in \mathbf{S}_0$ such that $SR_i S'$ works as follows. We consider all the vectors S' . Each such vector S' represents a state in \mathbf{S}_Γ . For each S' we check in polynomial time whether $SR_i S'$.

Reduction to TQBF In order to efficiently implement the algorithm above, we propose a reduction to TQBF (true quantified binary formulas) that mimics the algorithm. To do this, we introduce TQBF propositional variables $x_{\alpha,k}$ for all $\alpha \in \mathcal{L}_0$ and for all integers k . The variables indexed by k are said to be of level k . They correspond to the recursive nesting in the procedure mc described in Figure 1 for the cases $\Box_i \varphi$ and $\Box_i^G \varphi$. For instance, $x_{\alpha,k}$ is true if α is true at some state at depth k . Let X_k be the set of formulas of level k . More precisely, X_k contains exactly formulas $x_{\Delta_i \alpha, k}$ with $\alpha \in \Gamma_i$ for any agent i , and $x_{p,k}$ with p appearing in Γ or φ_0 .

Definition 7. We define the function tr that maps any formula of \mathcal{L} to a QBF-formula by $tr(\varphi_0) := tr_0(\varphi_0)$ with:

- $tr_k(p) = x_{p,k}$,
- $tr_k(\neg \varphi) = \neg tr_k(\varphi)$,
- $tr_k(\varphi \wedge \psi) = tr_k(\varphi) \wedge tr_k(\psi)$,

```

function  $mc(S, \Gamma, \varphi)$ 
  match  $\varphi$  do
    case  $p$ : return  $p \in V$ 
    case  $\Delta_i \alpha$ : return  $\alpha \in B_i$ 
    case  $\neg \psi$ : return not  $mc(S, \Gamma, \psi)$ 
    case  $\psi_1 \wedge \psi_2$ : return  $mc(S, \Gamma, \psi_1)$  and  $mc(S, \Gamma, \psi_2)$ 
    case  $\Box_i \psi$ :
      for all  $S' \in \mathbf{S}_\Gamma$  such that  $\mathcal{SR}_i S'$  do
        if not  $mc(S', \Gamma, \psi)$  return false
      return true
    case  $\Box_i^c \psi$ :
      for all  $S' \in \mathbf{S}_\Gamma$  such that  $\mathcal{SR}_i^c S'$  do
        if not  $mc(S', \Gamma, \psi)$  return false
      return true

```

Fig. 1. Generic algorithm for model checking \mathcal{L} -formulas.

- $tr_k(\Delta_i \alpha) = x_{\Delta_i \alpha, k}$,
- $tr_k(\Box_i \varphi) = \forall X_{k+1}, R_{ik} \rightarrow tr_{k+1}(\varphi)$,
- $tr_k(\Box_i^c \varphi) = \forall X_{k+1}, \neg R_{ik} \rightarrow tr_{k+1}(\varphi)$,

where:

$$R_{ik} := \bigwedge_{\alpha \in \Gamma_i} x_{\Delta_i \alpha, k} \rightarrow tr_{k+1}(\alpha).$$

The translation $tr_k(\Box_i \varphi)$ corresponds to case $\Box_i \varphi$ in the algorithm. State S (resp. S') is represented by the truth values of variables in X_k (resp. X_{k+1}). Formula R_{ik} reformulates $\mathcal{SR}_i S'$.

Proposition 1. *Let $\varphi_0 \in \mathcal{L}$. Then, the following two statements are equivalent:*

- $(S_0, \mathbf{S}_\Gamma) \models \varphi_0$
- $\exists X_0(\text{desc}_{S_0}(X_0) \wedge tr_0(\varphi_0))$ is QBF-true

where:

$$\text{desc}_{S_0}(X_0) := \bigwedge_{i \in \text{Agt}} \left(\bigwedge_{\alpha \in B_i} x_{\Delta_i \alpha, 0} \wedge \bigwedge_{\alpha \in \Gamma_i \setminus B_i} \neg x_{\Delta_i \alpha, 0} \right) \wedge \bigwedge_{p \in V} x_{p, 0} \wedge \bigwedge_{p \notin V} \neg x_{p, 0}$$

with $S_0 = ((B_i)_{i \in \text{Agt}}, V)$.

Proof. Let $\text{val}_S(X_k)$ represent the unique valuation on X_k satisfying $\text{desc}_S(X_k)$. We prove by induction on the structure of φ that $(S, \mathbf{S}_\Gamma) \models \varphi$ iff $\text{val}_S(X_k) \models tr_k(\varphi)$, for all k .

Induction base. Let $\varphi = p$, for some $p \in \text{Atm}$. We have $(S, \mathbf{S}_\Gamma) \models p$ iff $p \in V$ iff $x_{k,p} \in \text{val}_S(X_k)$ iff $\text{val}_S(X_k) \models tr_k(p)$.

Induction step. The cases for operators \neg and \wedge are straightforward. We proceed with the modal operators in the language:

- Let $\varphi = \Delta_i \alpha$. We have: $(S, \mathbf{S}_\Gamma) \models \Delta_i \alpha$ iff $\alpha \in B_i$ iff $x_{k, \Delta_i \alpha} \in \text{val}_S(X_k)$ iff $\text{val}_S(X_k) \models \text{tr}_k(\Delta_i \alpha)$.
- Let $\varphi = \Box_i \psi$. We denote by $\text{val}_S(X_k) + \text{val}_{S'}(X_{k+1})$ the valuation obtained by concatenating the valuation $\text{val}_S(X_k)$ and $\text{val}_{S'}(X_{k+1})$ (we take the truth values of propositions in X_k from the former and the truth values of propositions in X_{k+1} from the latter). We have:

$$\begin{aligned} (S, \mathbf{S}_\Gamma) \models \Box_i \psi &\iff \text{for all } S' \in \mathbf{S}_\Gamma, S\mathcal{R}_i S' \text{ implies } (S', \mathbf{S}_\Gamma) \models \psi \\ &\iff \text{for all } S' \in \mathbf{S}_\Gamma, S\mathcal{R}_i S' \text{ implies } \text{val}_{S'}(X_{k+1}) \models \text{tr}_{k+1}(\psi) \\ &\iff \text{for all } S' \in \mathbf{S}_\Gamma \text{ val}_S(X_k) + \text{val}_{S'}(X_{k+1}) \models R_{ik} \rightarrow \text{tr}_{k+1}(\psi) \\ &\iff \text{val}_S(X_k) \models \forall X_{k+1}, R_{ik} \rightarrow \text{tr}_{k+1}(\psi) \\ &\iff \text{val}_S(X_k) \models \text{tr}_k(\Box_i \alpha) \end{aligned}$$
- Let $\varphi = \Box_i^G \psi$. The proof is analogous to that for operator \Box_i above. In particular, we use relation \mathcal{R}_i^G and formula $\neg R_{ik}$.

Therefore, $(S_0, \mathbf{S}_\Gamma) \models \varphi_0$ iff $\text{val}_{S_0}(X_0) \models \text{tr}_0(\varphi)$. In addition, $\text{val}_{S_0}(X_0) \models \text{tr}_0(\varphi)$ is equivalent to $\exists X_0(\text{desc}_{S_0}(X_0) \wedge \text{tr}_0(\varphi))$ is QBF-true. This concludes the proof. \square

In [18], it is proved that the previous model checking problem formulated in the belief base semantics is PSPACE-hard, already for the fragment of \mathcal{L} with only “at least” implicit belief operators, but with no “at most” implicit belief operators involved. Thus, the fact that the generic model checking problem given in Figure 1 runs in polynomial space as well as Proposition 1 allow us to state the following complexity result.

Theorem 1. *Model checking \mathcal{L} -formulas is PSPACE-complete.*

5 Implementation and experimental results

We implemented a symbolic model checker,⁶ which uses the translation to TQBF. The resulting TQBF is then translated into a binary decision diagram (BDD), in the same way as done in [3]. The program is implemented in Haskell and the BDD library used is HasCacBDD [11]. It was compiled with GHC 9.2.7 in a MacBook Air with a 1.6 GHz Dual-Core Intel Core i5 processor and 16 GB of RAM, running macOS Ventura 13.3.1.

Table 1 shows the performance of the model checker on the examples of Section 3. It shows execution times for different instances. For both examples, the size of the model (*states*) is given by the number of possible valuations times the number of possible multi-agent belief bases: $2^{|Atm|} \times (2^{\text{ratoms}})^{|Agt|}$. The value of *ratoms* is the number of “relevant atoms”. There is one such atom for each formula in Γ , each propositional variable appearing in Γ and in the input formula, each formula α that is a sub-formula of the input formula, plus

⁶ Available at <https://src.koda.cnrs.fr/tiago.de.lima/lda/-/releases/0.0.4.4>

$cands = voters = Agt $	3	4	5	6	7	8	9	10
$ Atm $	9	16	25	36	49	64	81	100
$ratoms$	100	164	244	340	452	580	724	884
$states$	2^{309}	2^{672}	2^{1245}	2^{2076}	2^{3213}	2^{4704}	2^{6597}	2^{8940}
Execution time (sec.)	0.076	0.015	0.026	0.047	0.066	0.101	0.157	0.248

$cands = voters = Agt $	3	4	5	6	7	8	9	10
$ Atm $	9	16	25	36	49	64	81	100
$ratoms$	133	210	305	418	549	698	865	1050
$states$	2^{408}	2^{856}	2^{1550}	2^{2544}	2^{3892}	2^{5648}	2^{7866}	2^{10600}
Execution time (sec.)	0.081	0.063	0.334	3.066	17.588	90.809	KO	KO

Table 1. Symbolic model checker performance on Examples 1 (above) and 2 (below).

one atom for each formula $\Delta_i\alpha$ such that $\alpha \in \Gamma$. The number of states gives an idea of the size of the search space for modal formulas. In principle, to check a formula of the form $\Box^o\varphi$, one must check φ in every state of the model. Because of that, a naive implementation cannot be used. Indeed, in our tests with such a solution, no instance could be solved under the timeout of 10 minutes.

One can notice that the model checker is slower in the case of 3 candidates than in the case of 4 candidates (and in Example 1 the latter is true even up to 7 candidates). The reason is that the input formula is true for 3 candidates, whereas it is false on all the other cases. Checking that a box formula is false is easier, because the checker needs to find only one state where the formula in the scope of the box operator is false. Also note that instances of Example 1 are solved much faster than those of Example 2. This is due to two factors. First, Example 2 has larger belief bases, which imply larger number of states. Second, the input formula of the second example has a larger modal depth, which obliges the checker to generate a larger search tree.

6 Dynamic extension

In this section, we present a simple extension of the language \mathcal{L} by dynamic operators for modeling the agents' belief dynamics of private type. Similar operators were introduced in [19]. The novel result of this section is to show that adding them to the language \mathcal{L} does not increase complexity of the model checking problem. More generally, we provide a simple dynamic extension of the static language of implicitly believing *at least* and implicitly believing *at most* whose model checking problem remains in PSPACE.

The extended language is noted \mathcal{L}^+ and is defined by the following grammar:

$$\mathcal{L}^+ \stackrel{\text{def}}{=} \varphi ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i\varphi \mid \Box_i^{\mathcal{C}}\varphi \mid [+_i\alpha]\varphi,$$

where α ranges over \mathcal{L}_0 and i ranges over Agt .

Events of type $+_i\alpha$ are called informative events. In particular, $+_i\alpha$ is the event of agent i privately expanding its belief base with α .

The formula $[+_i\alpha]\varphi$ is read “ φ holds after the informative event $+_i\alpha$ has occurred”. It has the following semantic interpretation relative to a model.

Definition 8 (Satisfaction relation, cont.). *Let $S = (B_1, \dots, B_n, V) \in \mathbf{S}$ and let $(S, Cxt) \in \mathbf{M}$. Then:*

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

where $B_i^{+_i\alpha} = B_i \cup \{\alpha\}$ and $B_j^{+_i\alpha} = B_j$ for all $j \neq i$.

Intuitively speaking, the private belief expansion of i 's belief base by α simply consists in agent i adding the information that α to its belief base, while all other agents keep their belief bases unchanged. Let us go back to the Example 1 we introduced in Section 3 to illustrate the expressiveness of our dynamic extension.

Example 3. It worth noting that private belief dynamics allow agents to gather new information and to gain new knowledge. Suppose in the three-agent variant of the example agent 2 and agent 3 privately learn that 1 voted for 2. This ensures that there is no longer any information asymmetry between agent 1 and agents 2 and 3. Formally, we have

$$(S_0, \mathbf{S}_\Gamma) \models [+_2 \text{ vote}(1,2)] [+_3 \text{ vote}(1,2)]\chi_0,$$

where

$$\chi_0 \stackrel{\text{def}}{=} \bigwedge_{i \in \{1,2,3\}} \Box_i^o \psi_1,$$

and ψ_1 , S_0 , \mathbf{S}_Γ are defined as in Example 1 in Section 3. This means that, in the three-agent variant of the example, after agent 2 and agent 3 privately learn that agent 1 voted for 2, everybody only knows for whom an agent voted and for whom she/he did not vote.

As the following proposition highlights, we have reduction principles for the dynamic operators.

Proposition 2. *The following equivalences are valid for the class \mathbf{M} :*

$$\begin{aligned} [+_i\alpha]p &\leftrightarrow p \\ [+_i\alpha]\neg\varphi &\leftrightarrow \neg[+_i\alpha]\varphi \\ [+_i\alpha](\varphi_1 \wedge \varphi_2) &\leftrightarrow ([+_i\alpha]\varphi_1 \wedge [+_i\alpha]\varphi_2) \\ [+_i\alpha]\Delta_j\beta &\leftrightarrow \Delta_j\beta \text{ if } i \neq j \text{ or } \alpha \neq \beta \\ [+_i\alpha]\Delta_i\alpha &\leftrightarrow \top \\ [+_i\alpha]\Box_j\varphi &\leftrightarrow \Box_j\varphi \text{ if } i \neq j \\ [+_i\alpha]\Box_i\varphi &\leftrightarrow \Box_i(\alpha \rightarrow \varphi) \\ [+_i\alpha]\Box_j^{\mathbb{G}}\varphi &\leftrightarrow \Box_j^{\mathbb{G}}\varphi \text{ if } i \neq j \\ [+_i\alpha]\Box_i^{\mathbb{G}}\varphi &\leftrightarrow (\Box_i(\neg\alpha \rightarrow \varphi) \wedge \Box_i^{\mathbb{G}}\varphi) \end{aligned}$$

Proof. We only prove cases $\Box_i\varphi$ and $\Box_i^G\varphi$, since other cases are straightforward.

$$\begin{aligned} (S, Cxt) \models [+_i\alpha]\Box_i\varphi &\iff (S^{+i\alpha}, Cxt) \models \Box_i\varphi, \\ &\iff \forall S' \in Cxt : \text{if } S^{+i\alpha}\mathcal{R}_i S' \text{ then } (S', Cxt) \models \varphi, \\ &\iff \forall S' \in Cxt : \text{if } S\mathcal{R}_i S' \text{ and } S' \models \alpha \text{ then } (S', Cxt) \models \varphi, \\ &\iff (S, Cxt) \models \Box_i(\alpha \rightarrow \varphi). \end{aligned}$$

$$\begin{aligned} (S, Cxt) \models [+_i\alpha]\Box_i^G\varphi &\iff (S^{+i\alpha}, Cxt) \models \Box_i^G\varphi, \\ &\iff \forall S' \in Cxt : \text{if } S^{+i\alpha}\mathcal{R}_i^G S' \text{ then } (S', Cxt) \models \varphi, \\ &\iff \forall S' \in Cxt : \text{if } S\mathcal{R}_i^G S' \text{ or } (S\mathcal{R}_i S' \text{ and } S' \models \neg\alpha) \text{ then} \\ &\quad (S', Cxt) \models \varphi, \\ &\iff (S, Cxt) \models \Box_i(\neg\alpha \rightarrow \varphi) \wedge \Box_i^G\varphi. \quad \square \end{aligned}$$

Model checking for formulas in the language \mathcal{L}^+ is analogous to model checking for formulas in \mathcal{L} we defined in Section 4. The valid equivalences in Proposition 2 could be used to find a procedure for reducing model checking for formulas in \mathcal{L}^+ to model checking for formulas in \mathcal{L} . The problem is that such reduction is exponential due to the fact that every time we find a formula of type $[+_i\alpha]\Box_i^G\varphi$ we have to duplicate it into two parts $\Box_i(\neg\alpha \rightarrow \varphi)$ and $\Box_i^G\varphi$.

Fortunately we can easily adapt the generic algorithm presented in Section 4 in order to obtain a PSPACE procedure for model checking formulas of the language \mathcal{L}^+ . It is sufficient to add the following case for the dynamic operators to the main routine of the algorithm in Figure 1:

case $[+_i\alpha]\psi$: **return** $mc(S^{+i\alpha}, \Gamma, \psi)$

The resulting algorithm clearly runs in polynomial space. Thus, we can generalize the complexity result given in Theorem 1 to the language \mathcal{L}^+ .

Theorem 2. *Model checking \mathcal{L}^+ -formulas is PSPACE-complete.*

7 Conclusion

This paper describes optimal procedures for model checking multi-agent only believing formulas. As far as we know, we are the first to tackle the problem of automating model checking for the logic of multi-agent only believing or knowing. We implemented these procedures and presented some experimental results on computation time. Moreover, we extended the formalism with private belief expansion operators and showed that model checking remains PSPACE-complete. In the future, we plan to implement the dynamic extension presented in Section 6 and to extend the setting to introspective agents whose logic of belief (resp. knowledge) is K45 (resp. S5). Last but not least, we intend to apply our semantics for multi-agent only believing and model checking approach to epistemic planning. We believe that the compactness of our semantics can offer an advantage in terms of ease of implementation compared to the multi-relational Kripke semantics traditionally used in the context of epistemic planning [4, 5].

References

1. Aucher, G., Belle, V.: Multi-agent only knowing on planet kripke. In: Yang, Q., Wooldridge, M.J. (eds.) Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015. pp. 2713–2719. AAAI Press (2015), <http://ijcai.org/Abstract/15/384>
2. Belle, V., Lakemeyer, G.: Multi-agent only-knowing revisited. In: Lin, F., Sattler, U., Truszczyński, M. (eds.) Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010. AAAI Press (2010), <http://aaai.org/ocs/index.php/KR/KR2010/paper/view/1361>
3. van Benthem, J., van Eijck, J., Gattinger, M., Su, K.: Symbolic model checking for dynamic epistemic logic - S5 and beyond. *Journal of Logic and Computation* **28**(2), 367–402 (2018)
4. Bolander, T., Andersen, M.B.: Epistemic planning for single- and multi-agent systems. *Journal of Applied Non-Classical Logics* **21**(1), 656–680 (2011)
5. Bolander, T., Jensen, M.H., Schwarzentruher, F.: Complexity results in epistemic planning. In: Yang, Q., Wooldridge, M.J. (eds.) Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015. pp. 2791–2797. AAAI Press (2015), <http://ijcai.org/Abstract/15/395>
6. Fagin, R., Halpern, J.Y., Vardi, M.Y.: A model-theoretic analysis of knowledge. *Journal of the ACM* **38**(2), 382–428 (1991). <https://doi.org/10.1145/103516.128680>
7. Fagin, R., Geanakoplos, J., Halpern, J.Y., Vardi, M.Y.: The hierarchical approach to modeling knowledge and common knowledge. *International Journal of Game Theory* **28**(3), 331–365 (1999)
8. Fagin, R., Halpern, J., Moses, Y., Vardi, M.: Reasoning about Knowledge. MIT Press, Cambridge (1995)
9. Fagin, R., Halpern, J.Y.: Belief, awareness, and limited reasoning. *Artificial Intelligence* **34**(1), 39–76 (1987)
10. Gargov, G., Passy, S.: A note on boolean modal logic. In: Petkov, P.P. (ed.) *Mathematical Logic*, pp. 299–309. Springer, Boston, MA (1990). https://doi.org/10.1007/978-1-4613-0609-2_21, https://doi.org/10.1007/978-1-4613-0609-2_21
11. Gattinger, M.: HasCacBDD (2023), <https://github.com/m4lvin/HasCacBDD>, version 0.1.0.4, Feb. 3, 2023
12. Halpern, J.Y.: Reasoning about only knowing with many agents. In: Fikes, R., Lehnert, W.G. (eds.) Proceedings of the 11th National Conference on Artificial Intelligence. Washington, DC, USA, July 11-15, 1993. pp. 655–661. AAAI Press / The MIT Press (1993), <http://www.aaai.org/Library/AAAI/1993/aaai93-098.php>
13. Halpern, J.Y., Lakemeyer, G.: Multi-agent only knowing. *J. Log. Comput.* **11**(1), 41–70 (2001). <https://doi.org/10.1093/logcom/11.1.41>, <https://doi.org/10.1093/logcom/11.1.41>
14. Lakemeyer, G.: All they know: A study in multi-agent autoepistemic reasoning. In: Bajcsy, R. (ed.) Proceedings of the 13th International Joint Conference on Artificial Intelligence. Chambéry, France, August 28 - September 3, 1993. pp. 376–381. Morgan Kaufmann (1993), <http://ijcai.org/Proceedings/93-1/Papers/053.pdf>
15. Levesque, H.J.: All I know: A study in autoepistemic logic. *Artif. Intell.* **42**(2-3), 263–309 (1990). [https://doi.org/10.1016/0004-3702\(90\)90056-6](https://doi.org/10.1016/0004-3702(90)90056-6), [https://doi.org/10.1016/0004-3702\(90\)90056-6](https://doi.org/10.1016/0004-3702(90)90056-6)

16. Lomuscio, A., Qu, H., Raimondi, F.: MCMAS: an open-source model checker for the verification of multi-agent systems. *International Journal on Software Tools for Technology Transfer* **19**, 9–30 (2017)
17. Lorini, E.: In praise of belief bases: Doing epistemic logic without possible worlds. In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*. pp. 1915–1922. AAAI Press (2018)
18. Lorini, E.: Exploiting belief bases for building rich epistemic structures. In: Moss, L.S. (ed.) *Proceedings of the Seventeenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2019)*. EPTCS, vol. 297, pp. 332–353 (2019)
19. Lorini, E.: Rethinking epistemic logic with belief bases. *Artificial Intelligence* **282** (2020). <https://doi.org/https://doi.org/10.1016/j.artint.2020.103233>
20. Mertens, J.F., Zamir, S.: Formulation of bayesian analysis for games with incomplete information. *International Journal of Game Theory* **14**, 1–29 (1985). <https://doi.org/https://doi.org/10.1007/BF01770224>
21. Stalnaker, R.: Common ground. *Linguistics and Philosophy* **25(5-6)**, 701–721 (2002)

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this contribution is published in "Logics in Artificial Intelligence", and is available at https://doi.org/10.1007/978-3-031-43619-2_30.