Multi-Agent Belief Base Revision

Emiliano Lorini[1] and **François Schwarzentruber**[2]

[1]IRIT-CNRS, Toulouse University, France
[2]Univ Rennes, IRISA, CNRS, France
emiliano.lorini@irit.fr, francois.schwarzentruber@ens-rennes.fr

Abstract

We present a generalization of belief base revision to the multi-agent case. In our approach agents have belief bases containing both propositional beliefs and higher-order beliefs about their own beliefs and other agents' beliefs. Moreover, their belief bases are split in two parts: the mutable part, whose elements may change under belief revision, and the core part, whose elements do not change. We study a belief revision operator inspired by the notion of screened revision. We provide complexity results of model checking for our approach as well as an optimal model checking algorithm. Moreover, we study complexity of epistemic planning formulated in the context of our framework.

1 Introduction

Since the seminal work of Alchourrón, Gärdenfors and Makinson (AGM) [Alchourrón *et al.*, 1985], belief revision has been one of the central concepts in knowledge representation (KR). An important distinction in KR is between belief set (or theory) change and belief base change. While belief sets are deductively closed, belief bases may not be. This implies that belief sets are necessarily infinite since they contain all tautologies, whereas belief bases can be finite. Finiteness of belief bases offers an advantage for AI applications whereby autonomous agents (i.e., robots, embodied conversational agents) are conveniently described by the finite body of information in their databases. Existing formal theories of belief base change [Hansson, 1993; 1999; Benferhat *et al.*, 2002] are mainly focused on the single-agent case and describe beliefs by propositional logic formulas.

In more recent times, there has been a growing interest in understanding and modelling belief change in the context of multi-agent systems. Epistemic logic (EL) [Fagin *et al.*, 1995] and dynamic epistemic logic (DEL) [van Ditmarsch *et al.*, 2007] have become the main tools for formalizing epistemic states (e.g., knowledge, belief) and their dynamics in a multi-agent setting. Some extensions of DEL by the notion of belief revision were proposed [van Benthem, 2007; Aucher, 2005; van Ditmarsch, 2005; Baltag and Smets, 2008]. DEL-based approaches to belief revision are based on Kripke semantics which assign to each agent a plausibility ordering (also called epistemic entrenchment ordering) on possible worlds, or a qualitative plausibility measure in the style of rank-based theory [Spohn, 1988] and possibility theory [Dubois and Prade, 1988]. The latter allow to capture minimal change in an agent's belief revision process. The problem of such semantics is that they are not compact thereby making model checking not exploitable in practice. Indeed, the epistemic model — with its possible worlds and agents' plausibility orderings — with respect to which a given property has to be checked is often huge and its correspondence with the informal description of the initial situation is not always immediate.

To overcome these limitations of DEL, we present a generalization of belief base revision to the multi-agent case. Unlike existing single-agent belief base revision approaches in which agents' beliefs are propositional, in our approach agents can have both propositional beliefs and higher-order beliefs about their own beliefs and other agents' beliefs. The idea of using belief bases as a semantics for multi-agent epistemic logic was put forth in [Lorini, 2018; 2020]. In this paper we enrich and expand this idea by the notion of belief revision. Our approach to multi-agent belief revision is more parsimonious than standard DEL-based approaches. Indeed, it only uses belief bases and no plausibility ordering or plausibility measure is needed. This provides an advantage for formal verification since the formulation of model checking in our approach is more compact than in existing DEL-based approaches using Kripke semantics. Compactness also has an impact on the representation of epistemic planning, namely, the generalization of classical planning in which the goal to be achieved is not necessarily a state of the world but some belief states of one or more agents. The initial proposal for epistemic planning was to use a standard logic of knowledge or belief together with a representation of actions in terms of event models of DEL [Bolander and Andersen, 2011; Löwe *et al.*, 2011]. Given the high expressivity of DEL, it turned out that checking existence of a solution plan becomes quickly undecidable [Bolander and Andersen, 2011; Aucher and Bolander, 2013; Lê Cong *et al.*, 2018]. Epistemic planning with simple event models leading to a decidable fragment was studied by [Kominis and Geffner, 2015]. Moreover, decidable epistemic planning in a fragment of standard epistemic logic in which states are only described by conjunc-

tions of epistemic literals (i.e., formulas that do not contain any conjunction or disjunction) and in a variant of epistemic logic based on the notion of observability was studied, respectively, by [Muise *et al.*, 2015] and [Cooper *et al.*, 2016]. The standard formulation of epistemic planning in DEL is based on model checking [Bolander *et al.*, 2015]. This requires a full description of the initial epistemic model that, as emphasized above, is not well-suited for implementation and practical applications. On the contrary, our approach only requires to describe the agents' initial belief bases leading to a compact representation of epistemic planning in which epistemic actions are private belief revision operations.

Outline. Section 2 presents a modal language for modeling both explicit and implicit beliefs of multiple agents. While an agent's explicit belief is a piece of information in the agent's belief base, we call implicit belief a statement that can be inferred from the agent's explicit beliefs. The language has special operators for representing the consequences of an agent's private belief base revision operation. Section 3 presents the semantics exploiting belief bases with respect to which our language is interpreted. Taking inspiration from the notion of screened revision [Makinson, 1997] and, more generally from non-prioritized belief base revision [Hansson *et al.*, 2001], we assume that an agent's belief base is split in two parts: the mutable part, whose elements may change under belief revision, and the core part, whose elements do not change. Sections 4 and 5 are devoted, respectively, to the formulation of model checking and epistemic planning in our setting and to the study and their complexities. In Section 6, we illustrate epistemic planning with the help of a concrete example.

2 Language

This section presents a language for representing agents' individual beliefs of both explicit and implicit type as well as the consequences of a private belief base revision operation on an agent's beliefs. An agent's explicit beliefs are of two types: core beliefs and mutable beliefs. Core beliefs are stable pieces of information in an agent's belief base that do not change under belief revision. On the contrary, mutable beliefs are volatile, they may change over time if they turn out to be inconsistent with what the agent perceives.

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:

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

where p ranges over Atm and i ranges over Agt. $\mathcal{L}_0$ is the language for representing explicit beliefs. The formula $\triangle_i^c\alpha$ is read "agent i has the core belief that α", whereas $\triangle_i^m\alpha$ is read "agent i has the mutable (or volatile) belief that α". We define the explicit belief operator $\triangle_i\alpha$, read "agent i explicitly believes that α", as follows: $\triangle_i\alpha \stackrel{\text{def}}{=} \triangle_i^c\alpha \vee \triangle_i^m\alpha$.

The language $\mathcal{L}$ extends $\mathcal{L}_0$ and is defined by:

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

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 $\alpha, \neg$ and $\wedge$ in the standard way.

The formula $\Box_i^c\varphi$ is read "φ is derivable from agent i's core beliefs", while $\Box_i\varphi$ is read "φ is derivable from agent i's explicit beliefs". For simplicity, $\Box_i\varphi$ is also read "agent i implicitly believes that φ".

We define the dual operators $\Diamond_i^c$ and $\Diamond_i$ as follows: $\Diamond_i^c\varphi \stackrel{\text{def}}{=} \neg\Box_i^c\neg\varphi$, and $\Diamond_i\varphi \stackrel{\text{def}}{=} \neg\Box_i\neg\varphi$. Formula $\Diamond_i^c\varphi$ has to be read "φ is compatible (or consistent) with agent i's core beliefs", whereas $\Diamond_i\varphi$ has to be read "φ is compatible (or consistent) with agent i's explicit beliefs". The construction $[*_i\alpha]\varphi$ is read "φ holds after agent i has privately revised her belief base with α".

3 Semantics

In this section, we present a formal semantics for the language $\mathcal{L}$ exploiting belief bases. It refines the belief base semantics from [Lorini, 2020] by splitting an agent's belief base in two parts, the core and the mutable belief base. 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 $B = \big((C_i)_{i \in Agt}, (M_i)_{i \in Agt}, V\big)$ where $C_i \subseteq \mathcal{L}_0$ is agent i's core belief base, $M_i \subseteq \mathcal{L}_0$ is agent i's mutable belief base and $V \subseteq Atm$ is the actual environment. The set of all states is denoted by $\mathbf{S}$.*

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

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

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

Observe in particular the set-theoretic interpretation of the explicit belief operators in the previous definition: agent i has the core (resp. mutable) belief that α if and only if α is included in her core (resp. mutable) belief base.

Given $\beta \in \mathcal{L}_0$, we define $\mathbf{S}(\beta) = \{B \in \mathbf{S} : B \models \beta\}$. $\mathbf{S}(\beta)$ is the set of states induced by the integrity constraint β. This constraint limits the set of states with respect to which formulas are evaluated. Conceptually, it captures the information in the agents common ground, that is, the body of information that the agents commonly believe to be the case [Stalnaker, 2002].

We are interested in a specific subclass of states in which agents' belief bases are consistent.

Definition 3 (Belief consistent state). *Let $B = \big((C_i)_{i \in Agt}, (M_i)_{i \in Agt}, V\big) \in \mathbf{S}$ and $\beta \in \mathcal{L}_0$. The state B is said to be β-belief consistent (BC_β) if and only if $||B_i||_{\mathbf{S}(\beta)} \neq \emptyset$ for every $i \in Agt$, where $||X||_{\mathbf{S}(\beta)} = \{B' \in \mathbf{S}(\beta) : \forall \alpha \in X, B' \models \alpha\}$ for each $X \subseteq \mathcal{L}_0$ and $B_i = C_i \cup M_i$. The set of all β-belief consistent states is denoted by $\mathbf{S}_{BC_\beta}$.*

The set $||X||_{\mathbf{S}(\beta)}$ in the previous definition is called X's truth set under the integrity constraint β.

Note that Definition 3 can be formulated in terms of propositional consistency. To see this, let $\mathcal{L}_{\mathsf{PROP}}$ be the propositional language built from the set of atomic formulas $Atm^+ = Atm \cup \{p_{i,\alpha}^{\mathsf{c}}, p_{i,\alpha}^{\mathsf{m}} : i \in Agt \text{ and } \alpha \in \mathcal{L}_0\}$ and let tr be the following translation from $\mathcal{L}_0$ to $\mathcal{L}_{\mathsf{PROP}}$:

$$tr(p) = p, \qquad\qquad tr(\neg\alpha) = \neg tr(\alpha),$$
$$tr(\alpha_1 \wedge \alpha_2) = tr(\alpha_1) \wedge tr(\alpha_2), \quad tr(\triangle_i^{\mathsf{c}}\alpha) = p_{i,\alpha}^{\mathsf{c}},$$
$$tr(\triangle_i^{\mathsf{m}}\alpha) = p_{i,\alpha}^{\mathsf{m}}.$$

For each $X \subseteq \mathcal{L}_0$, we define $tr(X) = \{tr(\alpha) : \alpha \in X\}$. Moreover, we say that X is propositionally consistent if and only if $\perp \notin Cn(tr(X))$, where Cn is the classical deductive closure operator over the propositional language $\mathcal{L}_{\mathsf{PROP}}$. The following holds.

Proposition 1. *Let $X \subseteq \mathcal{L}_0$ and $\beta \in \mathcal{L}_0$. Then, $||X||_{\mathbf{S}(\beta)} \neq \emptyset$ if and only if $tr(X \cup \{\beta\})$ is propositionally consistent.*

Proof. Let W be the set of all valuations for the propositional language $\mathcal{L}_{\mathsf{PROP}}$. There exists a bijection $f : W \longrightarrow \mathbf{S}$ such that $f(w) \models \alpha$ iff $w \models tr(\alpha)$, for all $w \in W$ and $\alpha \in \mathcal{L}_0$. Propositional consistency of $tr(X \cup \{\beta\})$ means that we can find a valuation $w \in W$ where all formulas in $tr(X \cup \{\beta\})$ hold. Thus, there exists $B' \in \mathbf{S}(\beta)$ such that $f(w) = B'$ and $B' \models \beta'$ for all $\beta' \in X$. The left-to-right direction can be proved in an analogous way. $\square$

By the previous proposition, the notion of belief consistent state can be defined in an equivalent way by replacing $||B_i||_{\mathbf{S}(\beta)} \neq \emptyset$ with $\perp \notin Cn(tr(B_i \cup \{\beta\}))$ in Definition 3.

We denote by $MCS(\alpha, X, Y, \beta)$ the set of maximally Y-consistent subsets of the set X of $\mathcal{L}_0$-formulas including α under the integrity constraint β. The definition will be used at a later stage to precisely define the interpretation of the belief revision operator.

Definition 4. *Let $X, Y \subseteq \mathcal{L}_0$ and $\alpha \in \mathcal{L}_0$. Then, $X' \in MCS(\alpha, X, Y, \beta)$ if and only if:*

- $X' \subseteq X \cup \{\alpha\}$,
- $\alpha \in X'$,
- $||X' \cup Y||_{\mathbf{S}(\beta)} \neq \emptyset$, *and*
- *there is no X'' with $X' \subset X'' \subseteq X$ and $||X'' \cup Y||_{\mathbf{S}(\beta)} \neq \emptyset$.*

By Proposition 1, the set $MCS(\alpha, X, Y, \beta)$ can be defined in an equivalent way by replacing $||X' \cup Y||_{\mathbf{S}(\beta)} \neq \emptyset$ by $\perp \notin Cn(tr(X' \cup Y \cup \{\beta\}))$ in the third item and $||X'' \cup Y||_{\mathbf{S}(\beta)} \neq \emptyset$ by $\perp \notin Cn(tr(X'' \cup Y \cup \{\beta\}))$ in the fourth item of Definition 4.

The following definition introduces the concept of epistemic alternative.

Definition 5 (Epistemic alternatives). *Let $i \in Agt$. Then, $\mathcal{R}_i$ and $\mathcal{R}_i^{\mathsf{c}}$ are the binary relations on the set of states $\mathbf{S}$ such that, for all $B = ((C_i)_{i \in Agt}, (M_i)_{i \in Agt}, V), B' = ((C_i')_{i \in Agt}, (M_i')_{i \in Agt}, V') \in \mathbf{S}$:*

$$B\mathcal{R}_i^{\mathsf{c}}B' \text{ if and only if } \forall\alpha \in C_i : B' \models \alpha,$$
$$B\mathcal{R}_iB' \text{ if and only if } \forall\alpha \in B_i : B' \models \alpha,$$

where we recall that $B_i = C_i \cup M_i$ is agent i's belief base.

$B\mathcal{R}_i^{\mathsf{c}}B'$ means that B' is an epistemic alternative relative to agent i's core beliefs at B, while $B\mathcal{R}_iB'$ means that B' is an epistemic alternative relative to agent i's whole belief base at B. The idea of the previous definitions is that an agent's epistemic alternative is computed from the information in her belief base. Clearly, we have $\mathcal{R}_i \subseteq \mathcal{R}_i^{\mathsf{c}}$ since the core belief base is contained in the whole belief base.

The following definition specifies the interpretation of formulas in $\mathcal{L}$. They are interpreted with respect to pairs (B, β) with $B \in \mathbf{S}$ the actual state and $\beta \in \mathcal{L}_0$ the integrity constraint. (We omit Boolean cases, defined in the usual way.)

Definition 6 (Satisfaction relation). *Let $B = ((C_i)_{i \in Agt}, (M_i)_{i \in Agt}, V) \in \mathbf{S}$ and $\beta \in \mathcal{L}_0$. Then:*

$$(B, \beta) \models \alpha \iff B \models \alpha,$$
$$(B, \beta) \models \square_i^{\mathsf{c}}\varphi \iff \text{for all } B' \in \mathbf{S}(\beta) : \text{if } B\mathcal{R}_i^{\mathsf{c}}B' \text{ then}$$
$$(B', \beta) \models \varphi,$$
$$(B, \beta) \models \square_i\varphi \iff \text{for all } B' \in \mathbf{S}(\beta) : \text{if } B\mathcal{R}_iB' \text{ then}$$
$$(B', \beta) \models \varphi,$$
$$(B, \beta) \models [*_i\alpha]\varphi \iff (rev(B, i, \alpha, \beta), \beta) \models \varphi,$$

where $B \models \alpha$ is defined as in Definition 1 and

$$rev(B, i, \alpha, \beta) = ((C_i')_{i \in Agt}, (M_i')_{i \in Agt}, V')$$

such that $V' = V, \forall j \in Agt : C_j' = C_j$, and

$$\forall j \in Agt : M_j' = \begin{cases} M_j & \text{if } i \neq j \text{ or } ||C_i \cup \{\alpha\}||_{\mathbf{S}(\beta)} = \emptyset, \\ \bigcap_{X \in MCS(\alpha, M_i, C_i, \beta)} X & \text{otherwise.} \end{cases}$$

As the previous definition highlights, agent i implicitly believes that φ if and only if φ holds in all agent i's epistemic alternatives that are compatible with the integrity constraint β. Since the latter are computed from agent i's belief base, the implicit belief operator $\square_i$ captures the information that is deducible from the information in agent i's belief base under the integrity constraint β. Similarly, the operator $\square_i^{\mathsf{c}}$ captures the information that is derivable from agent i's core beliefs under the integrity constraint β.

The interpretation of the belief revision operator $[*_i\alpha]$ is based on state update. Specifically, the result of the private belief revision operation $*_i\alpha$ occurring in the state B under the integrity constraint β is the updated state $rev(B, i, \alpha, \beta)$ in which (i) core belief bases or mutable belief bases of the agents different from i are not modified, and (ii) the input α is added to agent i's mutable belief base only if it is consistent with agent i's core beliefs and the integrity constraint. If the latter is the case, then agent i's updated mutable belief base is equal to the intersection of the subsets of agent i's mutable belief base which are maximally consistent with respect to agent i's core belief base under the integrity constraint and which include the input α. This belief revision operation can be conceived as a generalization of the notion of screened revision [Makinson, 1997] to the multi-agent case. Note that instead of taking the intersection of *all* maximally consistent subsets (MCSs) of the initial belief base, we could have taken a *selection* of them, e.g., only MCSs whose cardinality is maximal. This parallels the distinction between full

meet and partial meet revision in the belief revision literature [Alchourrón *et al.*, 1985]. This would have had no impact on the complexity results presented below in Sections 4 and 5.

As the following proposition indicates, the belief revision operation $*_i\alpha$ is well-defined as it preserves belief consistency.

Proposition 2. *Let $i \in Agt$, $\alpha, \beta \in \mathcal{L}_0$ and $B \in \mathbf{S}_{BC_\beta}$. Then, $rev(B,i,\alpha,\beta) \in \mathbf{S}_{BC_\beta}$.*

We conclude this section by defining the notions of β-validity (viz. validity under the integrity constraint β).

Definition 7 (β-Validity)**.** *Let $\varphi \in \mathcal{L}$ and $\beta \in \mathcal{L}_0$. We say that φ is β-valid, noted $\models_\beta \varphi$, if and only if, for every $B \in \mathbf{S}_{BC_\beta}$ we have $(B, \beta) \models \varphi$. We say that φ is β-satisfiable if and only if $\neg\varphi$ is not β-valid.*

Some interesting validities are listed in the following two propositions.

Proposition 3. *Let $i, j \in Agt$ and $\beta \in \mathcal{L}_0$. Then,*

$$\text{if } \models_\beta \varphi \text{ then } \models_\beta \blacksquare\varphi \text{ with } \blacksquare \in \{\Box_i^\mathsf{c}, \Box_i, [*_i\alpha]\}, \quad (1)$$

$$\models_\beta \Box_i^\mathsf{c}\varphi \to \Box_i\varphi, \quad (2)$$

$$\models_\beta \Diamond_i\top, \quad (3)$$

$$\models_\beta [*_j\alpha]\Diamond_i\top. \quad (4)$$

As item (1) indicates, the implicit belief and belief revision operators are closed under necessitation. According to validity (2), if a formula is derivable from an agent's core beliefs, then it is also derivable from the agent's whole belief base. According to validity (3), an agent's belief base must be consistent. This is guaranteed by the fact that the notion of validity of Definition 7 is defined relative to β-belief consistent states. Validity (4) is a consequence of Proposition 2: an agent's belief base remains consistent after the occurrence of a private belief revision operation.

Proposition 4. *Let $i, j \in Agt$, $x \in \{\mathsf{c}, \mathsf{m}\}$ and $\beta \in \mathcal{L}_0$. Then,*

$$\models_\beta \Diamond_i^\mathsf{c}\alpha \leftrightarrow [*_i\alpha]\triangle_i^\mathsf{m}\alpha, \quad (5)$$

$$\models_{\alpha'} \triangle_i^\mathsf{c}\alpha' \leftrightarrow [*_i\alpha]\triangle_i^\mathsf{c}\alpha', \quad (6)$$

$$\models_\beta \triangle_j^x\alpha' \leftrightarrow [*_i\alpha]\triangle_j^x\alpha' \text{ if } i \neq j, \quad (7)$$

$$\models_\beta (\triangle_i^\mathsf{m}\alpha' \wedge \Diamond_i\alpha) \to [*_i\alpha]\triangle_i^\mathsf{m}\alpha'. \quad (8)$$

Validity (5) expresses a form of success postulate: α is consistent with an agent's core belief base if and only if, after revising her beliefs with α, α is included in the agent's mutable belief base. Validities (6) and (7) just mean that core beliefs do not change as a consequence of belief revision. Moreover, since a belief revision operation is private, it only modifies the mutable beliefs of the agent who performs it. Finally, according to validity (8), if the input of belief revision is already consistent with an agent's belief base, then no mutable belief is removed from the agent's belief base. This means that the agent merely expands her belief base without contracting it.

4 Model Checking

The following is a compact formulation of the model checking problem for the language $\mathcal{L}$. In the formulation of model

---

Algorithm 1 Generic algorithm for model checking.

procedure $mc(B, \varphi)$
    match φ **do**
        case p: **return** $B \models p$
        case $\triangle_i^\mathsf{c}\alpha$: **return** $\alpha \in C_i$
        case $\triangle_i^\mathsf{m}\alpha$: **return** $\alpha \in M_i$
        case $\neg\psi$: **return** not $mc(B, \psi)$
        case $\psi_1 \wedge \psi_2$: **return** $mc(B, \psi_1)$ and $mc(B, \psi_2)$
        case $\Box_G^\mathsf{c}\psi$:
            for all $B' \in \mathbf{S}(\beta)$ such that $B\mathcal{R}_i^\mathsf{c}B'$ **do**
                if not $mc(B', \psi)$ **return** false
        case $\Box_G\psi$:
            for all $B' \in \mathbf{S}(\beta)$ such that $B\mathcal{R}_iB'$ **do**
                if not $mc(B', \psi)$ **return** false
            return true
        case $[*_i\alpha]\psi$: **return** $mc(rev(B,i,\alpha,\beta), \psi)$

---

checking we have to suppose $B \in \mathbf{S}_{BC_\beta}$. This guarantees that agents' beliefs are consistent in the actual state.

> Model checking with integrity constraint
> *Given*: $\varphi \in \mathcal{L}$, $\beta \in \mathcal{L}_0$ and a finite $B \in \mathbf{S}_{BC_\beta}$.
> *Question*: Do we have $(B, \beta) \models \varphi$?

where the state $B = \big((C_i)_{i \in Agt}, (M_i)_{i \in Agt}, V\big)$ is said to be finite if V, every C_i and every M_i are finite.

Following [Lorini, 2019], we can prove the following lower bound for our model checking problem.

Theorem 1. *Model checking with integrity constraint is PSPACE-hard.*

In the rest of this section, we explain that the model checking problem is in PSPACE.

Algorithm 1 sums up the algorithm for the model checking of a formula φ in a given finite state $B \in \mathbf{S}_{BC_\beta}$ (we check that $B \in \mathbf{S}_{BC_\beta}$ with propositional consistency checking; if $B \notin \mathbf{S}_{BC_\beta}$ then the input is not correct). The first three cases are base cases. The next two cases are for Boolean connectives. For the $\Box_i^\mathsf{c}\psi$ and $\Box_i\psi$ cases, we enumerate all possible states B' containing correct subformulas of formulas in the initial B and in the initial formula φ. Remark that the current B' only takes a polynomial amount of space. Checking $B\mathcal{R}_i^\mathsf{c}B'$ or $B\mathcal{R}_iB'$ can be performed in polynomial time. Finally, the last case is the revision case. It remains to show that $\bigcap_{X \in MCS(\alpha,M_i,C_i,\beta)} X$ can be computed in polynomial space. To this end, we will consider all formulas and, for each of them, we check that it is in all subsets in $MCS(\alpha,M_i,C_i,\beta)$. To this aim, we explain how $MCS(\alpha,M_i,C_i,\beta)$ can be enumerated in polynomial space. To do that, we enumerate all $X' \subseteq X \cup \{\alpha\}$ containing α. Then, we check that $X' \cup Y \cup \{\beta\}$ is propositionally consistent. We also check that any superset X'' of X' containing just one extra formula is such that $X'' \cup Y \cup \{\beta\}$ is inconsistent. As propositional consistency is in NP, the overall computation is in polynomial space.

As the depth of nested calls in $mc(B, \varphi)$ is bounded by the size of φ, we obtain the PSPACE membership.

Theorem 2. *Model checking with integrity constraint is in PSPACE.*

5 Epistemic Planning

In this section, we first consider one of the simplest epistemic planning problem, where actions are revision operators equipped with preconditions: $(\varphi, [*_i\alpha])$ where $\varphi \in \mathcal{L}$ is a precondition, i is an agent and $\alpha \in \mathcal{L}_0$. The reader may easily imagine natural extensions in which actions could have ontic effects. For conveying the main idea of our approach, we keep the presentation as simple as possible, we stick to the following refined version of epistemic planning.

A plan is a linear sequence of elements $(\psi, [*_i\alpha])$ with $\psi \in \mathcal{L}$, $i \in AGT$ and $\alpha \in \mathcal{L}_0$. We define the formula $\langle\pi\rangle\varphi$ by induction: $\langle\epsilon\rangle\varphi = \varphi$, and $\langle(\psi, [*_i\alpha])\pi\rangle\varphi = \psi \wedge [*_i\alpha]\langle\pi\rangle\varphi$.

Epistemic planning
Given: $\varphi \in \mathcal{L}$, $\beta \in \mathcal{L}_0$, a finite $B \in \mathbf{S}_{BC_\beta}$, and a finite subset $\mathbb{A}$ of elements $(\psi, [*_i\alpha])$ with $\psi \in \mathcal{L}$, $i \in AGT$ and $\alpha \in \mathcal{L}_0$.
Question: Does there exist a finite sequence π of elements in $\mathbb{A}$ such that $(B, \beta) \models \langle\pi\rangle\varphi$?

Given how our framework is settled, the epistemic planning is not more difficult than classical planning (STRIPS for instance).

Theorem 3. *Epistemic planning is in PSPACE.*

Proof. Consider the following graph. Nodes are states. Edges are transitions from a state B' to $rev(B', i, \alpha, \beta)$ for some action $(\psi, [*_i\alpha])$ in $\mathbb{A}$ and ψ holds in B'. That graph is of exponential size since we restrict formulas in bases to be subformulas of φ, β, subformulas of formulas appearing in B and formulas occurring in the repertoire $\mathbb{A}$. Solving an epistemic planning instance reduces to reachability in that graph from the initial state B to a final state satisfying φ. We solve that reachability problem as follows. We store the current state starting from B. At each step, we non-deterministically choose the next action $(\psi, [*_i\alpha])$ to execute. We check that ψ holds in the current state and compute the next state. The algorithm accepts when the goal φ holds in the current base. As the model checking is in PSPACE, our algorithm is non-deterministic in polynomial space. We conclude by Savitch's theorem [Savitch, 1970]. $\square$

Our actions modify the current state and then modify possible worlds implicitly (because possible worlds are computed from the current state). On the contrary, in the epistemic planning based on Dynamic Epistemic Logic (DEL) [Bolander *et al.*, 2020], actions explicitly modify possible worlds. DEL allows for very complex actions and the DEL planning problem is undecidable. Our framework has the advantage of being very close to classical planning.

We conclude this section by considering the following variant of epistemic planning, called uniform epistemic planning, in which we enforce an agent i to know that the plan is successful. In particular, uniform epistemic planning consists in checking whether agent i *knows how* to achieve a certain goal φ through the execution of a plan.

Uniform epistemic planning
Given: $\varphi \in \mathcal{L}$, $\beta \in \mathcal{L}_0$, a finite $B \in \mathbf{S}_{BC_\beta}$, an agent $i \in AGT$, and a finite subset $\mathbb{A}$ of elements $(\psi, [*_i\alpha])$ with $\psi \in \mathcal{L}$, $i \in AGT$ and $\alpha \in \mathcal{L}_0$
Question: Does there exist a finite sequence π of elements in $\mathbb{A}$ such that $(B, \beta) \models \Box_i\langle\pi\rangle\varphi$?

Theorem 4. *Uniform epistemic planning is in EXPSPACE.*

Proof. We now consider the graph whose nodes are information sets (sets of states) for agent i. There is an edge from I to I' if there is an action $(\psi, [*_i\alpha])$ in $\mathbb{A}$ such ψ holds in all states in I and I' is the set of states $rev(B', i, \alpha, \beta)$ where B' in I. Now, we ask for a path in that graph from the set of possible states for agent i in B to a set of states all satisfying φ. As the graph is doubly-exponential, uniform epistemic planning is in NEXPSPACE = EXPSPACE. $\square$

6 Example

We illustrate the planning problem defined in the previous section with the help of a concrete example. Let $Agt = \{i_A, i_B, i_C, i_M\}$. The four agents Ann ($i_A$), Bob ($i_B$), Charles ($i_C$) and Mary ($i_M$) are mobile robots meeting at a roundabout. Each robot has to decide at which moment it will pass. In order to avoid a collision, the robots have to pass sequentially one after the other. We assume an external planning agent is connected to each robot in the system. Its aim is to assign to each robot a priority level from the set of priority levels $Lev = \{1, \dots, 4\}$ in such a way that a collision is avoided. Each priority level specifies a time in the sequence at which a robot has to pass. We assume that the set Atm includes three types of atomic formulas $pr_{i,k}$, $pa_{i,k}$ and $ought_{i,k}$ for every $i \in Agt$ and $k \in Lev$. They have to be read, respectively, "priority level k is assigned to robot i", "robot i will pass at time k" and "robot i ought to pass at time k". Following reductionist theories of intention [Sellars, 1968] according to which an intention is reducible to a first-person normative judgment about what one shall/ought to do, we introduce the following abbreviation:

$$intend_{i,k} \stackrel{\text{def}}{=} \Box_i ought_{i,k},$$

where $intend_{i,k}$ is read "agent i intends to pass at time k".

We assume the following three pieces of information are part of the robots' common ground, that is, they constitute the integrity constraint of the planning problem under consideration:

$$\alpha_1 \stackrel{\text{def}}{=} \bigwedge_{i \in Agt} \bigwedge_{k \in Lev} (pa_{i,k} \rightarrow \triangle_i pr_{i,k}),$$

$$\alpha_2 \stackrel{\text{def}}{=} \bigwedge_{i \in Agt} \bigwedge_{k,k' \in Lev: k \neq k'} (pr_{i,k} \rightarrow \neg pr_{i,k'}),$$

$$\alpha_3 \stackrel{\text{def}}{=} \bigwedge_{i \in Agt} \bigwedge_{k,k' \in Lev: k \neq k'} (\triangle_i pr_{i,k} \rightarrow \neg \triangle_i pr_{i,k'}).$$

According to α_1, a robot will pass at a certain time k only if it explicitly believes that it has priority level k. According to α_2, two different priority levels cannot be assigned to the same robot. According to α_3, a robot cannot believe that two different priority levels were assigned to it. Constraints α_2

and α_3 presuppose that the external planning agent can assign at most one priority level to one robot.

We suppose each robot i has the following piece of information α_i in its core belief base:

$$\alpha_i \overset{\text{def}}{=} \bigwedge_{k \in Lev} \left(pr_{i,k} \wedge \bigwedge_{j \neq i} \neg pa_{j,k} \right) \rightarrow ought_{i,k}.$$

Formula α_i specifies what a robot should do depending on its priority level and the other robots' priority levels. Specifically, if k is the only priority level which was assigned to robot i and no other robot has the same priority level, then robot i ought to pass at time k. To conclude, we suppose that each robot egocentrically believes that it has priority level 1, thereby being authorized to pass before the others. Specifically, for every robot i we assume $pr_{i,1}$ is the only piece of information in its mutable belief base. Moreover, the priority levels have been assigned in the following way but have not been communicated yet to the robots: Mary should pass first, followed by Bob, then Ann, and Charles as last.

To sum up, we define the actual state $B = \left((C_i)_{i \in Agt}, (M_i)_{i \in Agt}, V \right)$ as follows: (i) for every $i \in Agt$, $C_i = \{\alpha_i\}$ and $M_i = \{pr_{i,1}\}$, and (ii) $V = \{pr_{i_\mathsf{M},1}, pr_{i_\mathsf{B},2}, pr_{i_\mathsf{A},3}, pr_{i_\mathsf{C},4}, ought_{i_\mathsf{M},1}, ought_{i_\mathsf{B},2}, ought_{i_\mathsf{A},3}, ought_{i_\mathsf{C},4}\}$. The definition of V presupposes that in the initial situation the robots have not decided yet at which time to pass since the priority levels have not been communicated to them. The goal φ_G is to ensure that (i) each robot forms an intention to pass at certain time, and (ii) there are no two different robots with the same intention (i.e., the intention to pass at the same time). The latter condition is necessary to prevent the robots from colliding. Formally,

$$\varphi_G \overset{\text{def}}{=} \bigwedge_{i \in Agt} \bigvee_{k \in Lev} intend_{i,k} \wedge$$
$$\bigwedge_{i,j \in Agt : i \neq j} \bigwedge_{k \in Lev} (intend_{i,k} \rightarrow \neg intend_{j,k}).$$

In order to achieve this goal, the external planning agent can (i) inform a robot that a certain priority level is assigned to it, and (ii) inform a robot that another robot has been informed that a certain priority level is assigned to it. We assume that informative actions of the planning agent are truthful, i.e., a robot can be informed that a certain fact is true only if the fact is indeed true. Thus, the repertoire of actions of the planning agent is defined as follows:

$$\mathbb{A} = \left\{ (pr_{i,k}, [*_i pr_{i,k}]) : i \in Agt \text{ and } k \in Lev \right\} \cup$$
$$\left\{ (\triangle_i pr_{i,k}, [*_j \triangle_i pr_{i,k}]) : i,j \in Agt, i \neq j \text{ and } k \in Lev \right\}.$$

It is routine exercise to verify that the following plan

$$\pi = (pr_{i_\mathsf{M},1}, [*_{i_\mathsf{M}} pr_{i_\mathsf{M},1}])(pr_{i_\mathsf{B},2}, [*_{i_\mathsf{B}} pr_{i_\mathsf{B},2}])$$
$$(pr_{i_\mathsf{A},3}, [*_{i_\mathsf{A}} pr_{i_\mathsf{A},3}])(pr_{i_\mathsf{C},4}, [*_{i_\mathsf{C}} pr_{i_\mathsf{C},4}])$$
$$(\triangle_{i_\mathsf{B}} pr_{i_\mathsf{B},2}, [*_{i_\mathsf{M}} \triangle_{i_\mathsf{B}} pr_{i_\mathsf{B},2}])(\triangle_{i_\mathsf{A}} pr_{i_\mathsf{A},3}, [*_{i_\mathsf{M}} \triangle_{i_\mathsf{A}} pr_{i_\mathsf{A},3}])$$
$$(\triangle_{i_\mathsf{C}} pr_{i_\mathsf{C},4}, [*_{i_\mathsf{M}} \triangle_{i_\mathsf{C}} pr_{i_\mathsf{C},4}])(\triangle_{i_\mathsf{M}} pr_{i_\mathsf{M},1}, [*_{i_\mathsf{B}} \triangle_{i_\mathsf{M}} pr_{i_\mathsf{M},1}])$$
$$(\triangle_{i_\mathsf{A}} pr_{i_\mathsf{A},3}, [*_B \triangle_{i_\mathsf{A}} pr_{i_\mathsf{A},3}])(\triangle_{i_\mathsf{C}} pr_{i_\mathsf{C},4}, [*_B \triangle_{i_\mathsf{C}} pr_{i_\mathsf{C},4}])$$
$$(\triangle_{i_\mathsf{M}} pr_{i_\mathsf{M},1}, [*_A \triangle_{i_\mathsf{M}} pr_{i_\mathsf{M},1}])(\triangle_{i_\mathsf{B}} pr_{i_\mathsf{B},2}, [*_A \triangle_{i_\mathsf{B}} pr_{i_\mathsf{B},2}])$$
$$(\triangle_{i_\mathsf{C}} pr_{i_\mathsf{C},4}, [*_A \triangle_{i_\mathsf{C}} pr_{i_\mathsf{C},4}])(\triangle_{i_\mathsf{M}} pr_{i_\mathsf{M},1}, [*_{i_\mathsf{C}} \triangle_{i_\mathsf{M}} pr_{i_\mathsf{M},1}])$$
$$(\triangle_{i_\mathsf{B}} pr_{i_\mathsf{B},2}, [*_{i_\mathsf{C}} \triangle_B pr_{i_\mathsf{B},2}])(\triangle_{i_\mathsf{A}} pr_{i_\mathsf{A},3}, [*_{i_\mathsf{C}} \triangle_{i_\mathsf{A}} pr_{i_\mathsf{A},3}])$$

is a solution of our epistemic planning problem. Indeed, we have $(B, \alpha_1 \wedge \alpha_2 \wedge \alpha_3) \models \langle \pi \rangle \varphi_G$. The plan just consists in sequentially informing the robots about their priority levels and then sequentially informing the robots that the others have been informed about their priority levels.

7 Conclusion

We proposed a multi-agent framework that encompasses explicit/implicit beliefs and private belief revision. Our syntactic approach has strong advantages compared to the traditional semantic approaches based on Kripke structures. First, whereas Kripke structures are typically exponential in the number of atomic propositions, our syntactic representation with belief bases scales up in general. Secondly, the syntactic approach enables to elegantly define revision operators, compared to existing approaches which use cumbersome plausibility orderings or plausibility measures on possible worlds [van Benthem, 2007; Aucher, 2005; van Ditmarsch, 2005; Baltag and Smets, 2008]. Thirdly, the distinction between explicit/implicit beliefs comes for free in our approach. Fourthly, our model checking problem is in PSPACE, and that gives a hope of having reasonably efficient implementations of both the model checking and the epistemic planning problem, whereas epistemic planning in DEL is undecidable.

Perspectives of future work are manifold. Our framework can easily be adapted to capture introspective agents who have perfect knowledge of their own core and mutable belief base. To this aim, Definition 5 of epistemic alternative should be tuned as follows:

$$B\mathcal{R}_i^\mathsf{c}B' \text{ iff } C_i = C_i', M_i = M_i' \text{ and } \forall \alpha \in C_i : B' \models \alpha,$$
$$B\mathcal{R}_iB' \text{ iff } C_i = C_i', M_i = M_i' \text{ and } \forall \alpha \in B_i : B' \models \alpha.$$

The binary relations $\mathcal{R}_i^\mathsf{c}$ and $\mathcal{R}_i$ so defined are transitive and Euclidean. Interestingly, unlike DEL in which introspection over beliefs could be lost due to private belief change, in our approach it is necessarily preserved. Indeed, the agents' epistemic accessibility relations are recomputed from belief bases after the occurrence of a private belief revision operation. For this introspective variant, the model checking and epistemic planning remain in PSPACE. Future work will be devoted to an in-depth analysis of the logical properties of belief base revision in the case of introspective agents. Another direction of future work is to move beyond private belief base revision and to explore semi-private forms (e.g., an agent i sees that α while another agent j sees that i sees that α, but i is not aware of this). Last but not least, we plan to compare the expressivity of existing DEL-based approaches to belief revision and the expressivity of our belief base approach.

Acknowledgments

References

[Alchourrón *et al.*, 1985] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50:510–530, 1985.

[Aucher and Bolander, 2013] Guillaume Aucher and Thomas Bolander. Undecidability in epistemic planning. In *IJCAI 2013*, pages 27–33. AAAI Press, 2013.

[Aucher, 2005] Guillaume Aucher. A combined system for update logic and belief revision. In *Proceedings of PRIMA 2004*, volume 3371 of *LNAI*, pages 1–18. Springer-Verlag, 2005.

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

[Benferhat *et al.*, 2002] Salem Benferhat, Didier Dubois, Henri Prade, and Mary-Anne Williams. A practical approach to revising prioritized knowledge bases. *Studia Logica*, 70(1):105–130, 2002.

[Bolander and Andersen, 2011] Thomas Bolander and Mikkel B. Andersen. Epistemic planning for single- and multi-agent systems. *Journal of Applied Non-Classical Logics*, 21(1):9–34, 2011.

[Bolander *et al.*, 2015] Thomas Bolander, Martin Holm Jensen, and François Schwarzentruber. Complexity results in epistemic planning. In *IJCAI 2015*, pages 2791–2797. AAAI Press, 2015.

[Bolander *et al.*, 2020] Thomas Bolander, Tristan Charrier, Sophie Pinchinat, and François Schwarzentruber. DEL-based epistemic planning: Decidability and complexity. *Artificial Intelligence*, 287, 2020.

[Cooper *et al.*, 2016] Martin C. Cooper, Andreas Herzig, Faustine Maffre, Frédéric Maris, and Pierre Régnier. A simple account of multi-agent epistemic planning. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI 2016)*, pages 193–201, 2016.

[Dubois and Prade, 1988] Didier Dubois and Henri Prade. *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, 1988.

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

[Hansson *et al.*, 2001] Sven O. Hansson, Eduardo Fermé, John Cantwell, and Marcelo Falappa. Credibility limited revision. *Journal of Symbolic Logic*, 66:1581–1596, 2001.

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

[Hansson, 1999] Sven O. Hansson. A survey of non-prioritized belief revision. *Erkenntnis*, 50:413–427, 1999.

[Kominis and Geffner, 2015] Filippos Kominis and Hector Geffner. Beliefs in multiagent planning: from one agent to many. In *ICAPS 2015*, pages 147–155. AAAI Press, 2015.

[Lê Cong *et al.*, 2018] Sébastien Lê Cong, Sophie Pinchinat, and François Schwarzentruber. Small undecidable problems in epistemic planning. In *IJCAI 2018*, pages 4780–4786, 2018.

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

[Lorini, 2019] Emiliano Lorini. Exploiting belief bases for building rich epistemic structures. In L. S. Moss, editor, *Proceedings of the Seventeenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2019)*, volume 297 of *EPTCS*, pages 332–353, 2019.

[Lorini, 2020] Emiliano Lorini. Rethinking epistemic logic with belief bases. *Artificial Intelligence*, 282, 2020.

[Löwe *et al.*, 2011] Benedikt Löwe, Eric Pacuit, and Andreas Witzel. DEL planning and some tractable cases. In *Proceedings of the 3rd International International Workshop on Logic, Rationality and Interaction (LORI 2011)*, pages 179–192. Springer Berlin Heidelberg, 2011.

[Makinson, 1997] David Makinson. Screened revision. *Theoria*, 63:14–23, 1997.

[Muise *et al.*, 2015] Christian Muise, Vaishak Belle, Paolo Felli, Sheila A. McIlraith, Tim Miller, Adrian R. Pearce, and Liz Sonenberg. Planning over multi-agent epistemic states: A classical planning approach. In *AAAI 2015*, pages 3327–3334. AAAI Press, 2015.

[Savitch, 1970] Walter J. Savitch. Relationships between nondeterministic and deterministic tape complexities. *Journal of Computer and System Sciences*, 4(2):177–192, 1970.

[Sellars, 1968] Wilfrid Sellars. *Science and Metaphysics: Variations on Kantian Themes*. Routledge & Kegan Paul Ltd, London, 1968.

[Spohn, 1988] Wolfgan Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. In W. L. Harper and B. Skyrms, editors, *Causation in decision, belief change and statistics*, pages 105–134. Kluwer, 1988.

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

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

[van Ditmarsch *et al.*, 2007] Hans P. van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Kluwer Academic Publishers, 2007.

[van Ditmarsch, 2005] Hans P. van Ditmarsch. Prolegomena to dynamic logic for belief revision. *Synthese*, 147(2):229–275, 2005.