Base-based model checking for multi-agent only believing

Tiago de Lima[1] Emiliano Lorini[2] François Schwarzentruber[3]

[1]CRIL, Univ Artois and CNRS, France

[2]IRIT, CNRS, Toulouse University, France

[3]ENS Rennes, France

MAFTEC 11
Tarbes, France, 19 January 2024

Agent i **at least believes** that φ iff φ is true in all situations that i considers possible.

Agent i **at most believes** that φ iff φ is false in all situations that i does not consider possible.

Agent i **only believes** that φ iff i at least and at most believes that φ.

Semantics is involved and rely on universal models (which are huge).[1]

That makes it infeasible in practice.

[1] See [Aucher and Belle, 2015], [Belle and Lakemeyer, 2010] and [Halpern and Lakemeyer, 2001]

Lorini[2] proposed an epistemic logic using **belief bases** (instead of Kripke models).

It has two kinds of epistemic operators:

- explicit belief: present in the belief base
- implicit belief: inferred from explicit beliefs

Advantages:

- specifications are easier
- universal models can be described succinctly

In this paper: we use this idea to implement a model checker for multi-agent only believing.

[2] [Lorini 2020]

Assume:

- Atomic propositions: $Atm = \{p, q, \ldots\}$ (countably infinite)
- Agents: $Agt = \{1, \ldots, n\}$ (finite)

Language ($\mathcal{L}$):

$$\alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \triangle_i \alpha \qquad (\mathcal{L}_0)$$

$$\varphi ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i \varphi \mid \Box_i^{\mathsf{C}} \varphi$$

Abbreviation: $\Box_i^{\mathrm{O}} \varphi \stackrel{\text{def}}{=} \Box_i \varphi \wedge \Box_i^{\mathsf{C}} \neg\varphi$

Assume:

- Atomic propositions: $Atm = \{p, q, \ldots\}$ (countably infinite)
- Agents: $Agt = \{1, \ldots, n\}$ (finite)

Language ($\mathcal{L}$):

$$\alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \triangle_i \alpha \qquad (\mathcal{L}_0)$$

$$\varphi ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i \varphi \mid \Box_i^{\complement} \varphi$$

Abbreviation: $\Box_i^{\mathrm{O}} \varphi \stackrel{\text{def}}{=} \Box_i \varphi \wedge \Box_i^{\complement} \neg\varphi$

Readings:

- $\triangle_i \alpha$: agent i has the explicit belief that α
- $\Box_i \varphi$: agent i at least implicitly believes that φ
- $\Box_i^{\complement} \varphi$: agent i at most implicitly believes that $\neg\varphi$
- $\Box_i^{\mathrm{O}} \varphi$: agent i implicitly only believes that φ

State: $S = ((B_i)_{i \in Agt}, V)$, where:

- $B_i \subseteq \mathcal{L}_0$ (i's belief base)
- $V \subseteq Atm$ (actual environment)

The set of all states is noted **S**.

State: $S = ((B_i)_{i \in Agt}, V)$, where:

- $B_i \subseteq \mathcal{L}_0$ (i's belief base)
- $V \subseteq Atm$ (actual environment)

The set of all states is noted **S**.

Satisfaction relation for $\mathcal{L}_0$:

$$
\begin{array}{lll}
S \models p & \text{iff} & p \in V \\
S \models \neg\alpha & \text{iff} & S \not\models \alpha \\
S \models \alpha_1 \wedge \alpha_2 & \text{iff} & S \models \alpha_1 \text{ and } S \models \alpha_2 \\
S \models \triangle_i \alpha & \text{iff} & \alpha \in B_i
\end{array}
$$

Epistemic alternative:

$$S\mathcal{R}_i S' \text{ iff } S' \models \alpha, \text{ for all } \alpha \in B_i$$

S' is an **epistemic alternative** for agent i at S iff S' satisfies all explicit beliefs of i at S.

Model: (S, Cxt), where:

- $S \in \mathbf{S}$ (actual state)
- $Cxt \subseteq \mathbf{S}$ (context)

Remark

S is **not necessarily** in Cxt, because we model belief (not knowledge).

To model knowledge, we would have to suppose that $S \in Cxt$.

Satisfaction relation for $\mathcal{L}$:

$$
\begin{array}{lll}
(S, Cxt) \models \alpha & \text{iff} & S \models \alpha \\
(S, Cxt) \models \Box_i \varphi & \text{iff} & \text{for all } S' \in Cxt, \text{ if } S\mathcal{R}_i S' \text{ then } (S', Cxt) \models \varphi \\
(S, Cxt) \models \Box_i^{\complement} \varphi & \text{iff} & \text{for all } S' \in Cxt, \text{ if } S\mathcal{R}_i^{\complement} S' \text{ then } (S', Cxt) \models \varphi
\end{array}
$$

where $\mathcal{R}_i^{\complement} = (\mathbf{S} \times \mathbf{S}) \setminus \mathcal{R}_i$.

Lorini[3] showed that this is “equivalent” to epistemic logic.

Valid principles:

$$\triangle_i\varphi \rightarrow \Box_i\varphi \tag{I}$$

$$(\Box_i\varphi \wedge \Box_i(\varphi \rightarrow \psi)) \rightarrow \Box_i\psi \tag{K}$$

$$\text{From } \varphi \text{ infer } \Box_i\varphi \tag{N}$$

$$\text{From } \varphi \text{ and } \varphi \rightarrow \psi \text{ infer } \psi \tag{MP}$$

Additional principles can be obtained:

$$\Box_i\varphi \rightarrow \varphi \quad \text{if } S \in \mathit{Cxt} \text{ and } \mathcal{R}_i \text{ is reflexive} \quad \text{(Belief correctness)}$$

$$\neg\Box_i\varphi \rightarrow \Box_i\neg\Box_i\varphi \quad \text{if } S \in \mathit{Cxt} \text{ and } S\mathcal{R}_iS' \text{ iff } B_i = B_i' \quad \text{(Introspection)}$$

[3] [Lorini 2020]

Vocabulary profile: $\Gamma = (\Gamma_i)_{i \in Agt}$, where each $\Gamma_i \subseteq \mathcal{L}_0$.

It plays a role analogous to that of awareness[4].

Γ-universal model: contains all states at which agent i's explicit beliefs are built from Γ_i.

The model (S, Cxt) is Γ-universal if $S \in Cxt = \mathbf{S}_\Gamma$, with:

$$\mathbf{S}_\Gamma = \{((B'_i)_{i \in Agt}, V') \in \mathbf{S} \mid B'_i \subseteq \Gamma_i, \text{ for all } i \in Agt\}$$

If $\mathbf{S}_\Gamma = \mathbf{S}$, then $(S, \mathbf{S}_\Gamma)$ is a model with maximum ignorance, i.e., it contains only the information provided by S.

[4] [Fagin and Halpern, 1987]

Model checking problem:

input:
- a finite vocabulary profile Γ
- a finite state $S_0 \in \mathbf{S}_\Gamma$
- a formula $\varphi_0 \in \mathcal{L}$

output:
- 1, if $(S_0, \mathbf{S}_\Gamma) \models \varphi_0$
- 0, otherwise

Additional variables: $X_k = \{x_{\alpha,k} \mid \alpha \in \mathcal{L}_0\}$, for each $k \in \mathbb{N}$.

$x_{\alpha,k}$ corresponds to α at modal depth k.

Translation: $\mathrm{tr}(\varphi_0) = \mathrm{tr}_0(\varphi_0)$ with:

$$
\begin{aligned}
\mathrm{tr}_k(p) &= x_{p,k} \\
\mathrm{tr}_k(\neg\varphi) &= \neg\,\mathrm{tr}_k(\varphi) \\
\mathrm{tr}_k(\varphi \land \psi) &= \mathrm{tr}_k(\varphi) \land \mathrm{tr}_k(\psi) \\
\mathrm{tr}_k(\triangle_i \alpha) &= x_{\triangle_i \alpha,k} \\
\mathrm{tr}_k(\Box_i \varphi) &= \forall X_{k+1}(R_{i,k} \to \mathrm{tr}_{k+1}(\varphi)) \\
\mathrm{tr}_k(\Box_i^{\complement} \varphi) &= \forall X_{k+1}(\neg R_{i,k} \to \mathrm{tr}_{k+1}(\varphi))
\end{aligned}
$$

where:

$$
R_{i,k} = \bigwedge_{\alpha \in \Gamma_i} x_{\triangle_i \alpha,k} \to \mathrm{tr}_{k+1}(\alpha)
$$

Theorem

Let $\varphi_0 \in \mathcal{L}$ and $S_0 = ((B_i)_{i \in Agt}, V)$. The following two statements are equivalent:

- $(S_0, \mathbf{S}_\Gamma) \models \varphi_0$
- $\exists X_0 (D_0 \wedge \mathrm{tr}_0(\varphi_0))$ is QBF-true.

where:

$$D_0 = \bigwedge_{i \in Agt} \left(\bigwedge_{\alpha \in B_i} x_{\triangle_i \alpha, 0} \wedge \bigwedge_{\alpha \in \Gamma_i \setminus B_i} \neg x_{\triangle_i \alpha, 0} \right) \wedge \bigwedge_{p \in V} x_{p,0} \wedge \bigwedge_{p \notin V} \neg x_{p,0}$$

Theorem

Model checking $\mathcal{L}$-formulas is PSPACE-complete.

Proof.

Membership: via the translation to QBF (theorem above).

Hardness: via a translation from QBF (already in [Lorini 2019]). $\square$

Available at: `https://src.koda.cnrs.fr/tiago.de.lima/lda/`

Encodes the QBF into a binary decision diagram (BDD).

Made in Haskell, using HasCacBDD[5].

Experiment:

- Selection committee problem: Does $(S_0, \mathbf{S}_\Gamma) \models \Box_1^{\mathrm{O}} \psi_1 \wedge \bigwedge_{i \in \{2,3\}} \Box_i^{\mathrm{O}} \psi_2$?
- Selection committee variant problem: Does $(S_0', \mathbf{S}_\Gamma) \models \Box_2^{\mathrm{O}} \psi_2 \wedge \Box_1 \Box_2 \psi_1 \wedge \neg \Box_1 \Box_2^{\mathrm{O}} \psi_2$?

Meanings:

- ψ_1: the actual ballot by agents 1, 2 and 3.
- ψ_2: for whom agent i voted and 1 did not vote for c_2 or c_3.

[5] [Gattinger, 2023]

Performance in a MacBook Air with 1.6 GHz Dual-Core Intel Core i5, 16 GB RAM, macOS Ventura 13.3.1.

Selection committee example

$\lvert Agt \rvert$	3	4	5	6	7	8	9	10
$\lvert Atm \rvert$	9	16	25	36	49	64	81	100
ratoms	100	164	244	340	452	580	724	884
Execution time (sec.)	0.076	0.015	0.026	0.047	0.066	0.101	0.157	0.248

Variant with higher-order beliefs

$\lvert Agt \rvert$	3	4	5	6	7	8	9	10
$\lvert Atm \rvert$	9	16	25	36	49	64	81	100
ratoms	133	210	305	418	549	698	865	1050
Execution time (sec.)	0.081	0.063	0.334	3.066	17.588	90.809	KO	KO

ratoms (relevant atoms):

- one x_α for each $\alpha \in \Gamma$ and each $\alpha \in \mathrm{sub}(\varphi_0)$
- one x_p for each atomic proposition in Γ and in φ_0
- one $x_{\triangle_i \alpha}$ for each $\alpha \in \Gamma$

Search space = #states $\approx 2^{|Atm|} \times (2^{ratoms})^{|Agt|}$

Conclusion

First to automate model checking for multi-agent only believing (or knowing).

A few results on computation time showing its feasibility.

Results can be improved (very few optimisations are done).

Possible application to epistemic planning.
(The compactness of this semantics can ease the specification.)

Aucher, G. and Belle, V. (2015).
Multi-agent only knowing on planet kripke.
In: Yang, Q. and Wooldridge, M. G. (ed.) Proc. of IJCAI 2015, pp. 2713–2719.

Belle, V. and Lakemeyer, G. (2010)
Multi-agent only-knowing revisited.
In: Lin, F., Sattler, U. and Truszczynski, M. (eds.) Proc. of KR 2010.

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

Gargov, G. and Passy, S. (1990).
A note on Boolean modal logic.
In: Petkov, P.P. (ed.) Mathematical Logic, pp. 299–309.

Gattinger, M. (2023).
HasCacBDD – Haskell bindings for CacBDD, a Binary Decision Diagram (BDD) package with dynamic cache management.
`https://github.com/m4lvin/HasCacBDD`. Version 0.1.0.4.

Halpern, J. and Lakemeyer, G. (2001).
Multi-agent only knowing.
Journal of Logic and Computation, 11(1):41–70.

Lorini, E.
Exploiting belief bases for building rich epistemic structures.
In: Moss, L. S. (ed.) Proc. of TARK'19, pp. 332–353

Lorini, E.
Rethinking epistemic logic with belief bases.
Artificial Intelligence, 282.

Appendix

Agents in *Agt* are members of a selection committee for an academic position.

They have to choose which candidates among $Cand = \{c_1, \ldots, c_m\}$ to admit to an interview.

Each committee member must vote for exactly one candidate:

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

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

It is forbidden to vote for a co-author:

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

where $\text{ca} : Agt \longrightarrow 2^{Cand}$ associates agents to their co-authors among the candidates.

A candidate c is admitted to the interview iff c received at least one vote:

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

We suppose:

- The number of voters and candidates is the same, and it is greater than 2 ($|Agt| = |Cand| > 2$).
- There is exactly one candidate co-author for each voter ($\mathrm{ca}(i) = \{c_i\}$, for all $i \in Agt$).

Moreover, each agent knows:

- her own vote
- the result of the selection
- the rules of the voting procedure

That is, we have:

- $S_0 = \big((B_i)_{i \in Agt}, V\big)$ such that, for every $1 \leq i < n$:

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

  and

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

Each agent is aware of all these possibilities, i.e., for every $i \in Agt$:

$$\Gamma_i = B_i \cup \neg B_i$$

where $\neg B_i = \{\neg\alpha \mid \alpha \in B_i\}$.

Let $|Agt| = 3$. We have:

$$(S_0, \mathbf{S}_\Gamma) \models \triangle_1(\text{vote}(1,c_2) \wedge \neg\text{vote}(1,c_3) \wedge \text{adm}(2) \wedge \alpha_1 \wedge \alpha_2 \wedge \alpha_3)$$
$$(S_0, \mathbf{S}_\Gamma) \models \Box_1(\text{vote}(1,c_2) \wedge \neg\text{vote}(1,c_3) \wedge \text{adm}(2) \wedge \alpha_1 \wedge \alpha_2 \wedge \alpha_3)$$
$$(S_0, \mathbf{S}_\Gamma) \models \Box_1(\neg\text{vote}(3,c_3))$$
$$(S_0, \mathbf{S}_\Gamma) \models \Box_1(\text{vote}(2,c_3))$$

Similarly:

$$(S_0, \mathbf{S}_\Gamma) \models \Box_1(\text{vote}(3,c_2))$$

Agent 1 only knows for whom each agent voted, i.e.:

$$(S_0, \mathbf{S}_\Gamma) \models \Box_1^{\mathrm{O}} \psi_1$$

where:

$$\begin{aligned}\psi_1 \stackrel{\text{def}}{=} \ & \mathrm{vote}(1,c_2) \wedge \neg\mathrm{vote}(1,c_1) \wedge \neg\mathrm{vote}(1,c_3) \\ & \wedge \mathrm{vote}(2,c_3) \wedge \neg\mathrm{vote}(2,c_1) \wedge \neg\mathrm{vote}(2,c_2) \\ & \wedge \mathrm{vote}(3,c_2) \wedge \neg\mathrm{vote}(3,c_1) \wedge \neg\mathrm{vote}(3,c_3),\end{aligned}$$

But that is not the case for the other agents.

In fact, 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 c_2 or for c_3:

$$(S_0, \mathbf{S}_\Gamma) \models \Box_2^{\mathrm{O}} \psi_2 \wedge \Box_3^{\mathrm{O}} \psi_2$$

$$\begin{aligned}
\psi_2 \stackrel{\text{def}}{=} & \neg\text{vote}(1,c_1) \wedge (\text{vote}(1,c_2) \not\leftrightarrow \text{vote}(1,c_3)) \\
& \wedge \text{vote}(2,c_3) \wedge \neg\text{vote}(2,c_1) \wedge \neg\text{vote}(2,c_2) \\
& \wedge \text{vote}(3,c_2) \wedge \neg\text{vote}(3,c_1) \wedge \neg\text{vote}(3,c_3).
\end{aligned}$$

Now, committee member 1 explicitly knows that committee member 2 explicitly knows the rules of the game as well as the results of the selection. i.e., $S_0' = ((B_i')_{i \in Agt}, V')$ such that,

$$B_1' = B_1 \cup \{\triangle_2 \neg \mathrm{adm}(c_1), \triangle_2 \mathrm{adm}(c_2), \ldots, \triangle_2 \mathrm{adm}(c_n), \triangle_2 \alpha_1, \triangle_2 \alpha_2, \triangle_2 \alpha_3\}$$

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

$$B_i' = B_i,$$
$$V' = V,$$

where B_i and V are defined as above.

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

$$(S_0', B) \models \Box_2^{\mathrm{O}} \psi_2 \wedge \Box_1 \Box_2 \psi_2 \wedge \neg \Box_1 \Box_2^{\mathrm{O}} \psi_2$$