A Computationally Grounded Framework for Cognitive Attitudes

Tiago de Lima[1], Emiliano Lorini[2],
Elise Perrotin[3], François Schwarzentruber[4]

[1]CRIL, Université d'Artois, France
[2]IRIT, CNRS, Toulouse University, France
[3]National Institute of Advanced Industrial Science and Technology, Japan
[4]ENS Lyon, France

AAAI 2025

Formal semantics for logics of cognitive attitudes

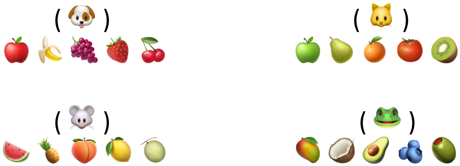
- Two types of cognitive attitudes
  - Epistemic: belief, knowledge, acceptance,...
  - Motivational: desire, goal, preference,...

Formal semantics for logics of cognitive attitudes

- Two types of cognitive attitudes
  - Epistemic: belief, knowledge, acceptance,...
  - Motivational: desire, goal, preference,...
- Standard approach: multi-relational Kripke structures

Formal semantics for logics of cognitive attitudes

- Two types of cognitive attitudes
  - Epistemic: belief, knowledge, acceptance,...
  - Motivational: desire, goal, preference,...
- Standard approach: multi-relational Kripke structures
- Main limitation: they are not succinct
  - Number of possible worlds is huge in real applications
  - $\binom{20}{5} \times \binom{15}{5} \times \binom{10}{5} \times \binom{5}{5} = 1.1732745024 \times 10^{10}$ possible equitable distributions of 20 different fruits between 4 agents



Our contribution

- Novel approach to cognitive attitudes and their interrelations relying on the notion of belief base
- Idea of using belief bases as a semantics for multi-agent epistemic logic has been elaborated in previous work:

Lorini, E. (2020). Rethinking Epistemic Logic with Belief Bases. *Artificial Intelligence*, 282.

Lorini, E. (2018). In Praise of Belief Bases: Doing Epistemic Logic Without Possible Worlds. *AAAI-2018*.

Our contribution

- Novel approach to cognitive attitudes and their interrelations relying on the notion of belief base
- Idea of using belief bases as a semantics for multi-agent epistemic logic has been elaborated in previous work:

Lorini, E. (2020). Rethinking Epistemic Logic with Belief Bases. *Artificial Intelligence*, 282.

Lorini, E. (2018). In Praise of Belief Bases: Doing Epistemic Logic Without Possible Worlds. *AAAI-2018*.

- Advantages:
  - Succinct semantics: agents' accessibility relations computed from their belief bases
  - Well-suited for model checking and epistemic/cognitive planning

Semantics: state

agents' belief bases $\subseteq \mathcal{L}_0$

state

valuation $\subseteq Atm$

$$S = \big((B_i)_{i \in Agt}, V\big)$$

$\mathcal{L}_0 \stackrel{\text{def}}{=} \alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \triangle_i \alpha$ $\quad \triangle_i \alpha$: "agent i explicitly believes that α"

with p ranging over Atm

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

Semantics: desire bases

$$\text{Appetitive desire base } D_i^+(S) = \{\alpha \in \mathcal{L}_0 : \alpha \rightarrow \mathsf{good}_i \in B_i\}$$

$$\text{Aversive desire base } D_i^-(S) = \{\alpha \in \mathcal{L}_0 : \alpha \rightarrow \mathsf{bad}_i \in B_i\}$$

Non-Humean view: desire-as-belief (DAB) thesis [Lewis, 1988]

Semantics: accessibility relations

Let $S = \big((B_i)_{i \in Agt}, V\big), S' = \big((B'_i)_{i \in Agt}, V'\big)$ be two states. Then,

Epistemic relation: $S\mathcal{E}_iS'$ if and only if $\forall\alpha \in B_i, S' \models \alpha$

Attraction relation: $S\mathcal{A}_iS'$ if and only if $\exists\alpha \in D_i^+(S)$ s.t. $S' \models \alpha$

Repulsion relation: $S\mathcal{R}_iS'$ if and only if $\exists\alpha \in D_i^-(S)$ s.t. $S' \models \alpha$

Language

$$\mathcal{L} \quad \stackrel{\text{def}}{=} \quad \varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i\varphi \mid ☺_i\varphi \mid ☹_i\varphi \mid [☺]_i\varphi \mid [☹]_i\varphi$$

with α ranging over $\mathcal{L}_0$ and i ranging over *Agt*

Language

$$\mathcal{L} \quad \stackrel{\text{def}}{=} \quad \varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i\varphi \mid ☺_i\varphi \mid ☹_i\varphi \mid [☺]_i\varphi \mid [☹]_i\varphi$$

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

$\Box_i\varphi$: "agent i implicitly believes that φ"

$☺_i\varphi$: "agent i is completely attracted to the fact that φ"

$☹_i\varphi$: "agent i is completely repelled by the fact that φ"

$[☺]_i\varphi$: "agent i is realistically attracted to the fact that φ"

$[☹]_i\varphi$: "agent i is realistically repelled by the fact that φ"

Interpretation of formulas

Wrt a state S and a set of states (*context*) U:

$$(S, U) \models \alpha \quad \text{if} \quad S \models \alpha$$

Interpretation of formulas

Wrt a state S and a set of states (*context*) U:

$$
\begin{array}{rll}
(S, U) \models \alpha & \text{if} & S \models \alpha \\
(S, U) \models \Box_i \varphi & \text{if} & \forall S' \in U,\ \text{if } S \mathcal{E}_i S' \text{ then } (S', U) \models \varphi
\end{array}
$$

Interpretation of formulas

Wrt a state S and a set of states (*context*) U:

$$
\begin{array}{rll}
(S,U) \models \alpha & \text{if} & S \models \alpha \\
(S,U) \models \Box_i\varphi & \text{if} & \forall S' \in U,\ \text{if } S\mathcal{E}_iS' \text{ then } (S',U) \models \varphi \\
(S,U) \models ☺_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ then } S\mathcal{A}_iS' \\
(S,U) \models ☹_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ then } S\mathcal{R}_iS'
\end{array}
$$

Interpretation of formulas

Wrt a state S and a set of states (*context*) U:

$$
\begin{array}{rll}
(S,U) \models \alpha & \text{if} & S \models \alpha \\
(S,U) \models \Box_i \varphi & \text{if} & \forall S' \in U,\ \text{if } S\mathcal{E}_i S' \text{ then } (S',U) \models \varphi \\
(S,U) \models ☺_i \varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ then } S\mathcal{A}_i S' \\
(S,U) \models ☹_i \varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ then } S\mathcal{R}_i S' \\
(S,U) \models [☺]_i \varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ and } S\mathcal{E}_i S' \text{ then } S\mathcal{A}_i S' \\
(S,U) \models [☹]_i \varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ and } S\mathcal{E}_i S' \text{ then } S\mathcal{R}_i S'
\end{array}
$$

Interpretation of formulas

Wrt a state S and a set of states (*context*) U:

$$
\begin{array}{rll}
(S,U) \models \alpha & \text{if} & S \models \alpha \\
(S,U) \models \Box_i\varphi & \text{if} & \forall S' \in U,\ \text{if } S\mathcal{E}_iS' \text{ then } (S',U) \models \varphi \\
(S,U) \models ☺_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ then } S\mathcal{A}_iS' \\
(S,U) \models ☹_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ then } S\mathcal{R}_iS' \\
(S,U) \models [☺]_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ and } S\mathcal{E}_iS' \text{ then } S\mathcal{A}_iS' \\
(S,U) \models [☹]_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ and } S\mathcal{E}_iS' \text{ then } S\mathcal{R}_iS'
\end{array}
$$

Theorem

The operators $☺_i$, $☹_i$, $[☺]_i$ and $[☹]_i$ are not expressible with the other modalities or each other.

Interpretation of formulas

Wrt a state S and a set of states (*context*) U:

$$
\begin{array}{rll}
(S,U) \models \alpha & \text{if} & S \models \alpha \\
(S,U) \models \Box_i\varphi & \text{if} & \forall S' \in U,\ \text{if } S\mathcal{E}_iS' \text{ then } (S',U) \models \varphi \\
(S,U) \models ☺_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ then } S\mathcal{A}_iS' \\
(S,U) \models ☹_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ then } S\mathcal{R}_iS' \\
(S,U) \models [☺]_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ and } S\mathcal{E}_iS' \text{ then } S\mathcal{A}_iS' \\
(S,U) \models [☹]_i\varphi & \text{if} & \forall S' \in U,\ \text{if } (S',U) \models \varphi \text{ and } S\mathcal{E}_iS' \text{ then } S\mathcal{R}_iS'
\end{array}
$$

Theorem

The operators $☺_i$, $☹_i$, $[☺]_i$ and $[☹]_i$ are not expressible with the other modalities or each other.

⇒ A sound and complete axiomatization is given in the paper

Cognitive positions and preferences

$\stackrel{\text{def}}{=}$	$☺_i\varphi$	$\neg☺_i\varphi$
$☹_i\varphi$	$\mathsf{A}_i\varphi$ (ambivalent)	$\mathsf{M}_i^{\downarrow}\varphi$ (demotivated)
$\neg☹_i\varphi$	$\mathsf{M}_i^{\uparrow}\varphi$ (motivated)	$\mathsf{I}_i\varphi$ (indifferent)

Table: Cognitive attitudes

$\stackrel{\text{def}}{=}$	$[☺]_i\varphi$	$\neg[☺]_i\varphi$
$[☹]_i\varphi$	$\mathsf{RA}_i\varphi$ (realistically ambivalent)	$\mathsf{RM}_i^{\downarrow}\varphi$ (realistically (demotivated)
$\neg[☹]_i\varphi$	$\mathsf{RM}_i^{\uparrow}\varphi$ (realistically motivated)	$\mathsf{RI}_i\varphi$ (realistically (indifferent)

Table: Realistic cognitive attitudes

Cognitive positions and preferences

$\stackrel{\text{def}}{=}$	$☺_i\varphi$	$\neg☺_i\varphi$
$☹_i\varphi$	$\mathsf{A}_i\varphi$ (ambivalent)	$\mathsf{M}_i^{\downarrow}\varphi$ (demotivated)
$\neg☹_i\varphi$	$\mathsf{M}_i^{\uparrow}\varphi$ (motivated)	$\mathsf{I}_i\varphi$ (indifferent)

Table: Cognitive attitudes

$\stackrel{\text{def}}{=}$	$[☺]_i\varphi$	$\neg[☺]_i\varphi$
$[☹]_i\varphi$	$\mathsf{RA}_i\varphi$ (realistically ambivalent)	$\mathsf{RM}_i^{\downarrow}\varphi$ (realistically (demotivated)
$\neg[☹]_i\varphi$	$\mathsf{RM}_i^{\uparrow}\varphi$ (realistically motivated)	$\mathsf{RI}_i\varphi$ (realistically (indifferent)

Table: Realistic cognitive attitudes

$$\psi \prec_i \varphi \stackrel{\text{def}}{=} (\mathsf{M}_i^{\uparrow}\varphi \wedge \neg\mathsf{M}_i^{\uparrow}\psi) \vee (\mathsf{M}_i^{\downarrow}\psi \wedge \neg\mathsf{M}_i^{\downarrow}\varphi)$$

$$\psi \prec_i^{\mathsf{real}} \varphi \stackrel{\text{def}}{=} (\mathsf{RM}_i^{\uparrow}\varphi \wedge \neg\mathsf{RM}_i^{\uparrow}\psi) \vee (\mathsf{RM}_i^{\downarrow}\psi \wedge \neg\mathsf{RM}_i^{\downarrow}\varphi)$$

For $\blacktriangleleft \in \{\prec_i, \prec_i^{\mathsf{real}}\}$:

$$\models \neg(\varphi \blacktriangleleft \varphi)$$

$$\models (\psi \blacktriangleleft \varphi) \rightarrow \neg(\varphi \blacktriangleleft \psi)$$

$$\models \big((\varphi_1 \blacktriangleleft \varphi_2) \wedge (\varphi_2 \blacktriangleleft \varphi_3)\big) \rightarrow (\varphi_1 \blacktriangleleft \varphi_3)$$

$\mathcal{L}_{\mathsf{dyn}}$: extension with dynamic operators $[\pi]$

belief expansion

belief forgetting

$$\mathcal{L}_{\mathsf{prog}} \quad \stackrel{\text{def}}{=} \quad \pi \quad ::= \quad +_i\alpha \mid -_i\alpha \mid \pi;\pi \mid \pi \cup \pi \mid ?\varphi$$

$[\pi]\varphi$: "φ holds after program π has been executed"

$\mathcal{L}_{\mathsf{dyn}}$: extension with dynamic operators $[\pi]$

belief expansion

belief forgetting

$$\mathcal{L}_{\mathsf{prog}} \quad \overset{\text{def}}{=} \quad \pi \quad ::= \quad +_i\alpha \mid -_i\alpha \mid \pi;\pi \mid \pi \cup \pi \mid ?\varphi$$

$[\pi]\varphi$: "φ holds after program π has been executed"

$$(S,U) \models [\pi]\varphi \quad \text{if} \quad \forall S' \in U, \text{ if } S\mathcal{P}^U_\pi S' \text{ then } (S',U) \models \varphi$$

with:

$$S\mathcal{P}^U_{+_i\alpha}S' \quad \text{iff} \quad V = V', B_i^{+_i\alpha} = B_i \cup \{\alpha\} \text{ and } \forall j \neq i, B_j^{+_i\alpha} = B_j$$

$$S\mathcal{P}^U_{-_i\alpha}S' \quad \text{iff} \quad V = V', B_i^{-_i\alpha} = B_i \setminus \{\alpha\} \text{ and } \forall j \neq i, B_j^{-_i\alpha} = B_j$$

$$S\mathcal{P}^U_{\pi_1;\pi_2}S' \quad \text{iff} \quad \exists S'' \in U \text{ such that } S\mathcal{P}^U_{\pi_1}S'' \text{ and } S''\mathcal{P}^U_{\pi_2}S'$$

$$S\mathcal{P}^U_{\pi_1 \cup \pi_2}S' \quad \text{iff} \quad S\mathcal{P}^U_{\pi_1}S' \text{ or } S\mathcal{P}^U_{\pi_2}S'$$

$$S\mathcal{P}^U_{?\varphi}S' \quad \text{iff} \quad S' = S \text{ and } (S,U) \models \varphi$$

Model checking

context induced by agent vocabulary profile $\Gamma = (\Gamma_i)_{i \in Agt}$

finite state

formula in $\mathcal{L}_{\mathsf{dyn}}$

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

$$\mathbf{S}_\Gamma = \Big\{ S = \big((B_i)_{i \in Agt}, V \big) \in \mathbf{S} : \forall i \in Agt, B_i \subseteq \Gamma_i \Big\} \text{ with } \Gamma_i \subseteq \mathcal{L}_0 \text{ finite}$$

Model checking

context induced by agent vocabulary profile $\Gamma = (\Gamma_i)_{i \in Agt}$

finite state — formula in $\mathcal{L}_{\mathsf{dyn}}$

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

$$\mathbf{S}_\Gamma = \Big\{ S = \big((B_i)_{i \in Agt}, V \big) \in \mathbf{S} : \forall i \in Agt, B_i \subseteq \Gamma_i \Big\} \text{ with } \Gamma_i \subseteq \mathcal{L}_0 \text{ finite}$$

Theorem

The model checking problem for $\mathcal{L}_{\mathsf{dyn}}$ is PSPACE-complete.

Poly-size reduction into TQBF given in the paper

Example: Bob and the messy room

Example: Bob and the messy room

$$
\begin{aligned}
S_0 &= (B_{Bob}, V_0) \\
B_{Bob} &= \{td_{Bob} \rightarrow ti_{Bob}, \\
&\quad ti_{Bob} \rightarrow \mathsf{bad}_{Bob}, \\
&\quad cr_{Bob} \rightarrow \mathsf{good}_{Bob}, \\
&\quad \neg tv_{Bob} \rightarrow \mathsf{bad}_{Bob}\} \\
V_0 &= \emptyset
\end{aligned}
$$

$$(S_0, \mathbf{S}_\Gamma) \models (td_{Bob} \prec^{\mathsf{real}}_{Bob} \neg td_{Bob})$$

with $\Gamma = (B_{Bob})$

Example: Bob and the messy room

$$
\begin{aligned}
S_0 &= (B_{Bob}, V_0)\\
B_{Bob} &= \{td_{Bob} \to ti_{Bob},\\
&\qquad ti_{Bob} \to \mathsf{bad}_{Bob},\\
&\qquad cr_{Bob} \to \mathsf{good}_{Bob},\\
&\qquad \neg tv_{Bob} \to \mathsf{bad}_{Bob}\}\\
V_0 &= \emptyset
\end{aligned}
$$

$$(S_0, \mathbf{S}_\Gamma) \models (td_{Bob} \prec^{\mathsf{real}}_{Bob} \neg td_{Bob})$$

with $\Gamma = (B_{Bob})$

$$A_1 \stackrel{\text{def}}{=} +_{Bob}\,(\neg td_{Bob} \to \neg tv_{Bob})$$

$$A_2 \stackrel{\text{def}}{=} +_{Bob}\,(td_{Bob} \to cr_{Bob})$$

$$A_3 \stackrel{\text{def}}{=} -_{Bob}\,(td_{Bob} \to ti_{Bob})$$

$$\pi_{talk} \stackrel{\text{def}}{=} \bigcup_{\substack{A, A' \in \{A_1, A_2, A_3\}:\\ A \neq A'}} A; A'$$

$$(S_0, \mathbf{S}_\Gamma) \models [\pi_{talk}](\neg td_{Bob} \prec^{\mathsf{real}}_{Bob} td_{Bob})$$

Example: Bob and the messy room

$$S_0 = (B_{Bob}, V_0)$$

$$B_{Bob} = \{td_{Bob} \rightarrow ti_{Bob},\ ti_{Bob} \rightarrow \mathsf{bad}_{Bob},\ cr_{Bob} \rightarrow \mathsf{good}_{Bob},\ \neg tv_{Bob} \rightarrow \mathsf{bad}_{Bob}\}$$

$$V_0 = \emptyset$$

$$(S_0, \mathbf{S}_\Gamma) \models (td_{Bob} \prec^{\mathsf{real}}_{Bob} \neg td_{Bob})$$

with $\Gamma = (B_{Bob})$

$$A_1 \stackrel{\text{def}}{=} +_{Bob}\ (\neg td_{Bob} \rightarrow \neg tv_{Bob})$$

$$A_2 \stackrel{\text{def}}{=} +_{Bob}\ (td_{Bob} \rightarrow cr_{Bob})$$

$$A_3 \stackrel{\text{def}}{=} -_{Bob}\ (td_{Bob} \rightarrow ti_{Bob})$$

$$\pi_{talk} \stackrel{\text{def}}{=} \bigcup_{\substack{A, A' \in \{A_1, A_2, A_3\}: \\ A \neq A'}} A; A'$$

$$(S_0, \mathbf{S}_\Gamma) \models [\pi_{talk}](\neg td_{Bob} \prec^{\mathsf{real}}_{Bob} td_{Bob})$$

Generalized to *k* children in the paper!

Experimental analysis

$\lvert Agt \rvert$	1	10	20	40	60
$\lvert Atm \rvert$	6	60	120	240	360
$\lvert \Gamma_i \rvert$	4	4	4	4	4
$\lvert \mathbf{S}_\Gamma \rvert$	2^{18}	2^{180}	2^{360}	2^{720}	2^{1080}
exec. time (sec.)	0.07	0.19	0.57	2.38	5.67

Table: Model checker performance on the example.