The Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25)

A Computationally Grounded Framework for Cognitive Attitudes

Tiago de Lima[1], Emiliano Lorini[2], Elise Perrotin[3], François Schwarzentruber[4]
[1]CRIL, Univ Artois and CNRS, Lens, France
[2]IRIT, CNRS, Toulouse University, France
[3]National Institute of Advanced Industrial Science and Technology, Tokyo, Japan
[4]ENS Lyon, LIP, France
delima@cril-lab.fr, Emiliano.Lorini@irit.fr, elise.perrotin@aist.go.jp, francois.schwarzentruber@ens-lyon.fr

Abstract

We introduce a novel language for reasoning about agents' cognitive attitudes of both epistemic and motivational type. We interpret it by means of a computationally grounded semantics using belief bases. Our language includes five types of modal operators for implicit belief, complete attraction, complete repulsion, realistic attraction and realistic repulsion. We give an axiomatization and show that our operators are not mutually expressible and that they can be combined to represent a large variety of psychological concepts including ambivalence, indifference, being motivated, being demotivated and preference. We present a dynamic extension of the language that supports reasoning about the effects of belief change operations. Finally, we provide a succinct formulation of model checking for our languages and a PSPACE model checking algorithm relying on a reduction into TQBF. We present some experimental results for the implemented algorithm on computation time in a concrete example.

Code — https://gitlab.in2p3.fr/tiago.delima/cognitive-attitudes-source-code

Extended version — https://arxiv.org/abs/2412.14073

Introduction

An agent's cognitive state encompasses its epistemic attitudes (e.g., beliefs) and motivational (or conative) attitudes (e.g., desires and preferences). Their relationships as well as their influence on the agent's behavior are objects of study in cognitive psychology (Albarracín et al. 2018) and philosophy of mind (Searle 2001; Humberstone 1992). They play a prominent role in the explanation of others' behaviours and of our own behaviours through the so-called intentional stance (Dennett 1987). Cognitive (or mental) attitudes have also been studied by logicians, both in philosophy and in AI. Several logics dealing with epistemic and practical reasoning of rational agents have been proposed. These include epistemic logics (Hintikka 1962; Fagin et al. 1995), logics of preferences (Wright 1963, 1972; Liu 2011; van Benthem, Girard, and Roy 2009), logics of beliefs and preferences (Boutilier 1994; Lorini 2021; Azimipour and Naumov 2021), logics of desires and pro-attitudes (Dubois, Lorini, and Prade 2017; Su et al. 2007), logics of intention (Shoham 2009; Icard, Pacuit, and Shoham 2010; Lorini and Herzig 2008), BDI (belief, desire, intention) logics (Cohen and Levesque 1990; Herzig and Longin 2004; Meyer, van der Hoek, and van Linder 1999; Wooldridge 2000). The idea of describing rational agents in terms of their epistemic and motivational attitudes is shared with classical decision theory and game theory, according to which rational agents are assumed to make decisions on the basis of their beliefs and preferences.

Most logics of cognitive attitudes rely on extensional semantics based on so-called Kripke models: sets of possible worlds (or states) supplemented with one or more binary accessibility relations for each agent representing, e.g., the agent's epistemic uncertainty or preference ordering over the states. Multi-relational Kripke models are general and mathematically elegant. Nonetheless, they are limited from a modeling point of view as they are not succinct. Even a simple situation like a card game (e.g., Hanabi) with 4 players having 5 cards each among a set of 50 cards require a Kripke model with 5.5×10^{23} states to represent all alternatives on which an agent's uncertainty and preferences bear. For this reason, it is hard, if not unfeasible, to implement model checking for a logic of cognitive attitudes using a Kripke semantics since the model cannot be explicitly constructed. Some alternatives to solve this succinctness problem have been proposed in the area of epistemic logic including semantics based on BDDs (van Benthem et al. 2018), Boolean formulas and programs (Charrier and Schwarzentruber 2015, 2017, 2018; Charrier, Pinchinat, and Schwarzentruber 2019), and the notion of visibility (Cooper et al. 2021; van der Hoek, Troquard, and Wooldridge 2011). However, to our knowledge no general succinct semantics for a logic of cognitive attitudes, combining epistemic ones with motivational ones, has been proposed up to now.

In this paper, we tackle this problem by relying on a semantics for cognitive attitudes based on the notion of belief base that has been shown to provide an interesting computationally grounded alternative to Kripke models in representing epistemic concepts in the single-agent case (Konolige 1986) as well as in the multi-agent case (Lorini 2018, 2019, 2020; Lorini and Rapion 2022). Unlike standard Kripke semantics for logics of cognitive attitudes in which possible states and agents' accessibility relations are given as primitive, in our semantics they are defined from and grounded on the primitive concept of belief base. We will use belief bases to define three types of accessibility relation capturing, respectively, the states that an agent considers possible, those it finds attractive and those it finds repulsive. The first belongs to the epistemic sphere, while the second and the third belong to the motivational sphere. By means of this semantics, we will interpret a modal language including five modal operators for implicit belief, complete attraction, complete repulsion, realistic attraction and realistic repulsion as well as a dynamic extension of it that supports reasoning about the effects of belief change operations. We will show that in our semantics model checking can be formulated in a more succinct way than in the standard Kripke semantics, which opens up the possibility of using it in practice. We will provide a PSPACE model checking procedure relying on a reduction into TQBF and some experimental results for the implemented procedure on computation time.

Related work Our belief base approach has some aspects in common with the body of literature on preference representation based on priority graphs (de Jongh and Liu 2009; Liu 2011; Souza and Moreira 2021; Souza 2016), that was recently extended to deontic logic in (van Benthem, Grossi, and Liu 2014). We share with these works the idea i) that there are two alternative approaches to the representation of preferences and more generally of mental attitudes, namely the extensional approach based, e.g., on multi-relational Kripke models and the syntactic approach based, e.g., on belief bases or priority graphs, and ii) that the two approaches must coexist and be integrated into the same logical framework. But there are also some important differences. These works are mostly focused on the single-agent case while our approach is multi-agent. Secondly, they mainly focus on preference representation, while our emphasis is on the bipolar aspect of motivational attitudes (i.e., attraction vs repulsion, motivation vs demotivation). Thirdly, our modal language of cognitive attitudes is profoundly different from existing ones, including from the one presented in (Souza 2016, Chapters 4 and 5) which, like ours, covers both epistemic and motivational attitudes. Souza's language includes "betterness" (S4) and "strict betterness" modalities, while our language includes primitive modalities for (complete and realistic) attraction and repulsion. Our language also includes explicit belief modalities and the axiomatization we will provide has specific axioms relating explicit belief to implicit mental attitudes. Souza's language has no explicit belief involved.

All proofs and additional content can be found in the extended version of the paper (see URL on page 1).

Belief Base Semantics

Following (Lorini 2018, 2020), in this section we present a formal semantics for cognitive attitudes exploiting belief bases, where elements of an agent's belief base are explicit beliefs of the agent. Basic notions of appetitive and aversive desire are directly defined from a belief base. We will use belief bases to define three types of accessibility relations for belief, attraction and repulsion.

States

Assume a countably infinite set of atomic propositions Atm and a finite set of agents $Agt = \{1, \ldots, n\}$. We suppose the set Atm includes special atomic formulas of type good_i and bad_i for every $i \in Agt$ meaning respectively that "agent i gets a reward" (or "agent i feels pleasure") and "agent i gets a punishment" (or "agent i feels pain"). We define the language $\mathcal{L}_0$ for explicit belief by the following grammar in Backus-Naur Form (BNF):

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

where p ranges over Atm and i ranges over Agt. $\mathcal{L}_0$ is the language used to represent explicit beliefs. The formula $\triangle_i \alpha$ reads "agent i has the explicit belief that α".

In our semantics, a state is not a primitive notion but is decomposed into different elements: one belief base per agent and a valuation of propositional atoms. An agent's belief base is analogous to the agent's local state and the propositional valuation is analogous to the local state of the environment in interpreted systems (Fagin et al. 1995).

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

The following definition specifies truth conditions for formulas in $\mathcal{L}_0$ relative to states.

Definition 2 (Satisfaction relation)**.** *Let $S = \big((B_i)_{i \in Agt}, V\big) \in \mathbf{S}$. We define $S \models \alpha$ by:*

$$\begin{aligned} S \models p \quad & \textit{if} \quad p \in V, \\ S \models \neg\alpha \quad & \textit{if} \quad S \not\models \alpha, \\ S \models \alpha_1 \wedge \alpha_2 \quad & \textit{if} \quad S \models \alpha_1 \textit{ and } S \models \alpha_2, \\ S \models \triangle_i \alpha \quad & \textit{if} \quad \alpha \in B_i. \end{aligned}$$

The explicit belief operator $\triangle_i$ has a set-theoretic interpretation: agent i has the explicit belief that α at state S if and only if α is included in its belief base at S.

We define the notion of appetitive desire (resp. aversive desire) from explicit belief combined with reward (resp. punishment). Specifically, we call agent i's appetitive desire base (resp. agent i's aversive desire base) at a state S the set of all facts that, according to agent i's explicit beliefs at S, entail a reward (resp. a punishment). The special atoms good_i and bad_i are key elements of the definition.

Definition 3 (Appetitive and aversive desire base)**.** *Let $i \in Agt$ and $S \in \mathbf{S}$. We define:*

$$\begin{aligned} D_i^+(S) &= \{\alpha \in \mathcal{L}_0 : \alpha \to \mathsf{good}_i \in B_i\}, \\ D_i^-(S) &= \{\alpha \in \mathcal{L}_0 : \alpha \to \mathsf{bad}_i \in B_i\}. \end{aligned}$$

$D_i^+(S)$ and $D_i^-(S)$ are, respectively, agent i's appetitive and aversive desire bases at state S.

Note that we could have taken appetitive and aversive desire bases as primitive concepts. We prefer to define them from the primitive concept of belief base for the sake of minimality. The solution we adopt uses only one primitive concept (belief base) instead of three. This point is reminiscent of the debate in philosophy opposing the Humean to the non-Humean view of desire (Karlsson 2000). While according to the Humean view desires and aversions are distinguished from beliefs, according to the non-Humean view, also called the *desire-as-belief (DAB) thesis*, they are reducible to them.[1] In this respect, our semantics is in line with the non-Humean view.

Accessibility Relations

Three types of accessibility relation, for epistemic possibility, attraction and repulsion, can be directly computed from the agents' belief bases that, following Definition 1, are included in the state description. The following definition introduces the epistemic accessibility relation.

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

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

$S\mathcal{E}_i S'$ means that at state S agent i considers state S' (epistemically) possible. According to the definition, the latter is the case if and only if S' satisfies all facts that agent i explicitly believes at S. The set $\mathcal{E}_i(S) = \{S' \in \mathbf{S} : S\mathcal{E}_i S'\}$ is agent i's epistemic state at S.

The relation $\mathcal{E}_i$ as defined above is not necessarily reflexive. As explained in (Lorini 2020), reflexivity can be obtained by restricting it to the subclass of "belief correct" states in which the information in an agent's belief base is actually true. We do not make this restriction since we model belief instead of knowledge. Unlike knowledge, an agent's belief can be incorrect (i.e., it might be the case that an agent believes that φ is true whereas φ is actually false). We also do not consider introspection for belief. However, there is a natural way to modify the accessibility relation $\mathcal{E}_i$ to make implicit belief introspective. It would be sufficient to add the condition that, for a state S' to be accessible from state S, agent i's belief base at state S' is identical to agent i's belief base at state S (i.e., $B_i = B'_i$). By adding this extra condition, the relation $\mathcal{E}_i$ would become transitive and Euclidean. Transitivity corresponds to positive introspection, while Euclidianity corresponds to negative introspection.

Analogously, we compute an agent's accessibility relations for attraction and repulsion from its appetitive and aversive desire bases, that as shown in Definition 3 are defined from its belief base. Specifically, we suppose that an agent is attracted to a state (or finds it attractive) if and only if at least one of its appetitive desires is satisfied at this state, i.e., the state makes something the agent wishes to achieve true. Conversely, we suppose that an agent is repelled by a state (or finds it repulsive) if and only if at least one fact for which the agent has aversion comes true at this state, i.e., the state makes something the agent wishes to avoid true.

[1] An issue of debate is the violation by the DAB thesis of the independence requirement between desire and belief change (see (Lewis 1988, 1996)). In (Bradley and List 2009) it is shown that under some conditions of separation between evaluative and non-evaluative formulas, the requirement is not violated. We leave the analysis of analogous conditions in our framework to future work.

Definition 5 (Attractive and repulsive alternatives). *Let $i \in Agt$. $\mathcal{A}_i$ and $\mathcal{R}_i$ are the binary relations on $\mathbf{S}$ such that, for all $S = \big((B_i)_{i\in Agt}, V\big), S' = \big((B'_i)_{i\in Agt}, V'\big) \in \mathbf{S}$:*

$$S\mathcal{A}_i S' \textit{ if and only if } \exists \alpha \in D_i^+(S) \textit{ s.t. } S' \models \alpha,$$
$$S\mathcal{R}_i S' \textit{ if and only if } \exists \alpha \in D_i^-(S) \textit{ s.t. } S' \models \alpha.$$

$S\mathcal{A}_i S'$ means that at state S agent i is attracted to state S', whereas $S\mathcal{R}_i S'$ means that at S it is repelled by S'.

Models

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 they commonly believe to be the case (Stalnaker 2002).

Definition 6 (Model). *A model is a pair (S, U) such that $S \in U \subseteq \mathbf{S}$. The class of models is denoted by $\mathbf{M}$.*

Note that we suppose the actual state to be included in the agents' common ground. However, validities would not change even if we did not suppose this, due to the fact that we model belief instead of knowledge. If we modeled knowledge instead of belief, supposing that $S \in U$ would become a necessary requirement. Note also that U can be any (possibly infinite) set of states with no restriction on the information that is included in belief bases. Nonetheless, in some cases it is useful to restrict to models in which an agent's belief base is constructed from a finite repository of information, or vocabulary. The agent's vocabulary plays a role analogous to that of awareness in (Fagin and Halpern 1987). This observation leads to the following definition.

Definition 7 (Γ-model). *Let $\Gamma = (\Gamma_i)_{i\in Agt}$ where, for every $i \in Agt$, $\Gamma_i \subseteq \mathcal{L}_0$ represents agent i's vocabulary. The model (S, U) in $\mathbf{M}$ is said to be a Γ-model if $S \in U = \mathbf{S}_\Gamma$, with $\mathbf{S}_\Gamma = \Big\{\big((B'_i)_{i\in Agt}, V'\big) \in \mathbf{S} : \forall i \in Agt, B'_i \subseteq \Gamma_i\Big\}$. The class of Γ-models is denoted by $\mathbf{M}(\Gamma)$.*

$\Gamma = (\Gamma_i)_{i\in Agt}$ in Definition 7 is also called agent vocabulary profile. Clearly, when $\Gamma_i = \mathcal{L}_0$ for every $i \in Agt$, we have $\mathbf{S}_\Gamma = \mathbf{S}$. A model $(S, \mathbf{S})$ is a model with maximal ignorance: it only contains the information provided by the actual state S.

Language

We define a general modal language $\mathcal{L}$ extending the language $\mathcal{L}_0$ by five types of modal operators for cognitive attitudes: implicit belief ($\Box_i$), complete attraction ($☺_i$), complete repulsion ($☹_i$), realistic attraction ($[☺]_i$) and realistic repulsion ($[☹]_i$). It is defined by the following grammar:

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

where α ranges over $\mathcal{L}_0$ and i ranges over Agt. The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from α, $\neg$ and $\wedge$ as usual. The formula $\Box_i\varphi$ is read "agent i implicitly believes that φ". The formula $☺_i\varphi$ is read "agent i is completely attracted to the fact that φ", whereas $☹_i\varphi$ is read "agent i is completely repelled by the fact that φ". Furthermore, the formula $[\smiley]_i\varphi$ is read "agent i is realistically attracted to the fact that φ", whereas $[\frownie]_i\varphi$ is read "agent i is realistically repelled by the fact that φ". The satisfaction relation between models and $\mathcal{L}$-formulas is defined as follows. (Boolean cases are omitted since they are defined as usual.)

Definition 8 (Satisfaction relation, cont.)**.** *Let $(S, U) \in \mathbf{M}$. We define $(S, U) \models \varphi$ by:*

$$
\begin{array}{rll}
(S,U) \models \alpha & \textit{if} & S \models \alpha, \\
(S,U) \models \Box_i\varphi & \textit{if} & \forall S' \in U, \textit{ if } S\mathcal{E}_iS' \textit{ then } \\
 & & (S',U) \models \varphi, \\
(S,U) \models \smiley_i\varphi & \textit{if} & \forall S' \in U, \textit{if } (S',U) \models \varphi \textit{ then } \\
 & & S\mathcal{A}_iS', \\
(S,U) \models \frownie_i\varphi & \textit{if} & \forall S' \in U, \textit{ if } (S',U) \models \varphi \textit{ then } \\
 & & S\mathcal{R}_iS', \\
(S,U) \models [\smiley]_i\varphi & \textit{if} & \forall S' \in U, \textit{ if } (S',U) \models \varphi \textit{ and } \\
 & & S\mathcal{E}_iS' \textit{ then } S\mathcal{A}_iS', \\
(S,U) \models [\frownie]_i\varphi & \textit{if} & \forall S' \in U, \textit{ if } (S',U) \models \varphi \textit{ and } \\
 & & S\mathcal{E}_iS' \textit{ then } S\mathcal{R}_iS'.
\end{array}
$$

Interpretations of the different modalities are restricted to the actual context U. The semantic interpretations of the three modalities $\Box_i$, $\smiley_i$ and $\frownie_i$ highlight their conceptual symmetry. The implicit belief operator is interpreted in the usual way: agent i implicitly believes that φ, i.e. $\Box_i\varphi$, if and only if φ is true at all states that it considers epistemically possible. Conversely, agent i is completely attracted to (resp. repelled by) φ, i.e. $\smiley_i\varphi$ (resp. $\frownie_i\varphi$), if and only if every state satisfying φ is attractive (resp. repulsive) to it. Modalities $\smiley_i$ and $\frownie_i$ are instances of the so-called "window" modality (Humberstone 1983; Gargov, Passy, and Tinchev 1987; Gargov and Goranko 1979; Goranko 1983). The idea of using a "window" modality for representing desire was defended in (Dubois, Lorini, and Prade 2017). Realistic attraction (resp. repulsion) is an agent's attraction (resp. repulsion) relative to its epistemic state: agent i is realistically attracted to (resp. repelled by) φ, i.e. $[\smiley]_i\varphi$ (resp. $[\frownie]_i\varphi$), if every φ-state in its epistemic state is attractive (resp. repulsive).

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 $(S, U) \in \mathbf{M}$ such that $(S, U) \models \varphi$, and φ is valid (denoted $\models \varphi$) if $\neg\varphi$ is not satisfiable. We now establish that the operators $\smiley_i$, $\frownie_i$, $[\smiley]_i$ and $[\frownie]_i$ are not definable using the other operators, including each other.

Theorem 1. *The operators $\smiley_i$, $\frownie_i$, $[\smiley]_i$ and $[\frownie]_i$ are not expressible with the other modalities or each other.*

This result still holds in the context of the dynamic extension we will present later in the paper.

Axiomatics

In this section we provide a sound and complete axiomatization for the class of models $\mathbf{M}$.

Definition 9. *We define the logic* LCA *(Logic of Cognitive Attitudes) to be the extension of classical propositional logic given by the following axioms and rules of inference:*

$$\big(\Box_i\varphi \land \Box_i(\varphi \to \psi)\big) \to \Box_i\psi \qquad \textbf{(A1)}$$

$$\big(\circledcirc\varphi \land \circledcirc(\neg\varphi \land \psi)\big) \to \circledcirc\psi \qquad \textbf{(A2)}$$

$$\triangle_i\alpha \to \Box_i\alpha \qquad \textbf{(A3)}$$

$$\triangle_i(\alpha \to \mathsf{good}_i) \to \smiley_i\alpha \qquad \textbf{(A4)}$$

$$\triangle_i(\alpha \to \mathsf{bad}_i) \to \frownie_i\alpha \qquad \textbf{(A5)}$$

$$\smiley_i\varphi \to [\smiley]_i\varphi \qquad \textbf{(A6)}$$

$$\frownie_i\varphi \to [\frownie]_i\varphi \qquad \textbf{(A7)}$$

$$\Box_i\varphi \to [\smiley]_i\neg\varphi \qquad \textbf{(A8)}$$

$$\Box_i\varphi \to [\frownie]_i\neg\varphi \qquad \textbf{(A9)}$$

$$[\smiley]_i\varphi \to \Box_i(\varphi \to \mathsf{good}_i) \qquad \textbf{(A10)}$$

$$[\frownie]_i\varphi \to \Box_i(\varphi \to \mathsf{bad}_i) \qquad \textbf{(A11)}$$

$$\frac{\varphi}{\Box_i\varphi} \qquad \textbf{(R1)}$$

$$\frac{\varphi}{\circledcirc\neg\varphi} \qquad \textbf{(R2)}$$

$$\frac{\varphi \leftrightarrow \psi}{\circledcirc\varphi \leftrightarrow \circledcirc\psi} \qquad \textbf{(R3)}$$

for $\circledcirc \in \{\smiley_i, \frownie_i, [\smiley]_i, [\frownie]_i\}$.

Most of these axioms and rules are natural to consider and directly follow from the semantics of the operators. Let us highlight the four that may not be immediately intuitive: **A2**, **A8**, **A9**, and **R2**. **A2** can be interpreted as follows. Suppose an agent is completely attracted to the fact that φ (i.e., $\smiley_i\varphi$). This means it finds attractive all situations at which φ holds, which implies that it finds attractive all situations at which $\varphi \land \psi$ holds. Moreover, suppose the agent is completely attracted to the fact that $\neg\varphi \land \psi$ (i.e., $\smiley_i(\neg\varphi \land \psi)$). This means that it finds attractive all situations at which $\neg\varphi \land \psi$ holds. But, finding attractive all situations at which $\varphi \land \psi$ holds and finding attractive all situations at which $\neg\varphi \land \psi$ holds clearly implies finding attractive all situations at which ψ holds (since the union of the set of $\varphi \land \psi$-situations and the set of $\neg\varphi \land \psi$-situations equals the set of ψ-situations). The latter means being completely attracted to the fact that ψ (i.e., $\smiley_i\psi$). The corresponding versions of **A2** for complete repulsion, realistic attraction and realistic repulsion are justified in an analogous way. **A8**, **A9**, and **R2** say that if φ is believed then its negation is both realistically attractive and realistically repulsive, and that the negation of a valid fact is both completely and realistically attractive as well as repulsive. This can be understood as follows: if φ is true in all considered situations, then all considered situations in which it is false (of which there are none) are both attractive and repulsive. Put differently, when an agent believes that φ, it has no realistic concern about $\neg\varphi$. Thus, it is willing to realistically appreciate and despise $\neg\varphi$ at the same time.

Checking soundness of this logic w.r.t. our class of models is straightforward. The proof of the completeness follows the general lines of the proofs of Theorem 1, Theorem 2 and Corollary 1 in (Lorini 2020).

Theorem 2. *The logic* LCA *is sound and complete for the class* $\mathbf{M}$.

Cognitive Positions

It is interesting to study the language $\mathcal{L}$ from a combinatorial perspective. Specifically, it is worth combining the four primitive modalities $☺_i$, $☹_i$, $[☺]_i$ and $[☹]_i$ so as to define eight cognitive positions of motivational type. In Table 1 we define the four notions of "being motivated by φ" ($\mathsf{M}_i^{\uparrow}\varphi$), "being demotivated by φ" ($\mathsf{M}_i^{\downarrow}\varphi$), "being indifferent about φ" ($\mathsf{I}_i\varphi$) and "being ambivalent about φ" ($\mathsf{A}_i\varphi$). The definiendum (e.g., $\mathsf{M}_i^{\uparrow}\varphi$) is defined by the conjunction of the two definiens (e.g., $☺_i\varphi \wedge \neg☹_i\varphi$).

$\stackrel{\text{def}}{=}$	$☺_i\varphi$	$\neg☺_i\varphi$
$☹_i\varphi$	$\mathsf{A}_i\varphi$	$\mathsf{M}_i^{\downarrow}\varphi$
$\neg☹_i\varphi$	$\mathsf{M}_i^{\uparrow}\varphi$	$\mathsf{I}_i\varphi$

Table 1: Cognitive attitudes

In Table 2, we define their realistic counterparts: "being realistically motivated by φ" ($\mathsf{RM}_i^{\uparrow}\varphi$), "being realistically demotivated by φ" ($\mathsf{RM}_i^{\downarrow}\varphi$), "being realistically indifferent about φ" ($\mathsf{RI}_i\varphi$) and "being realistically ambivalent about φ" ($\mathsf{RA}_i\varphi$).

$\stackrel{\text{def}}{=}$	$[☺]_i\varphi$	$\neg[☺]_i\varphi$
$[☹]_i\varphi$	$\mathsf{RA}_i\varphi$	$\mathsf{RM}_i^{\downarrow}\varphi$
$\neg[☹]_i\varphi$	$\mathsf{RM}_i^{\uparrow}\varphi$	$\mathsf{RI}_i\varphi$

Table 2: Realistic cognitive attitudes

From the definitions, it is easy to prove that being motivated (resp. realistically motivated) implies not being demotivated (resp. realistically demotivated) .

Proposition 1. *Let $i \in Agt$. Then,*

$$\models \mathsf{M}_i^{\uparrow}\varphi \rightarrow \neg\mathsf{M}_i^{\downarrow}\varphi, \quad (1)$$

$$\models \mathsf{RM}_i^{\uparrow}\varphi \rightarrow \neg\mathsf{RM}_i^{\downarrow}\varphi. \quad (2)$$

We leverage the notions of being motivated/demotivated and their realistic variants to define two notions of dyadic preference between facts:

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

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

The abbreviation $\psi \prec_i \varphi$ is read "agent i prefers φ to ψ", while $\psi \prec_i^{\mathsf{real}} \varphi$ is read "agent i realistically prefers φ to ψ". An agent prefers φ to ψ if either i) it is motivated by φ without being motivated by ψ, or ii) it is demotivated by ψ without being demotivated by φ. The realistic preference relation is defined analogously. It is straightforward to show that both preference relations are strict partial orders.

Proposition 2. *Let $\blacktriangleleft \in \{\prec_i, \prec_i^{\mathsf{real}}\}$. Then,*

$$\models \neg(\varphi \blacktriangleleft \varphi), \quad (3)$$

$$\models (\psi \blacktriangleleft \varphi) \rightarrow \neg(\varphi \blacktriangleleft \psi), \quad (4)$$

$$\models \big((\varphi_1 \blacktriangleleft \varphi_2) \wedge (\varphi_2 \blacktriangleleft \varphi_3)\big) \rightarrow (\varphi_1 \blacktriangleleft \varphi_3). \quad (5)$$

Validities (3), (4) and (5) capture irreflexivity, asymmetry and transitivity for preference, respectively.

Dynamic Extension

In this section we move from a static to a dynamic perspective on cognitive attitudes by presenting an extension of the language $\mathcal{L}$ that supports reasoning about the consequences of belief change operations. We see the latter as atomic programs on an agent's cognitive state and use the constructs of dynamic logic for representing complex programs. Our extension is defined by the following grammar:

$$\begin{aligned}
\mathcal{L}_{\mathsf{prog}} \stackrel{\text{def}}{=} \pi \quad &::= \quad +_i\alpha \mid -_i\alpha \mid \pi;\pi \mid \pi \cup \pi \mid ?\varphi,\\
\mathcal{L}_{\mathsf{dyn}} \stackrel{\text{def}}{=} \varphi \quad &::= \quad \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i\varphi \mid ☺_i\varphi \mid \\
& \qquad ☹_i\varphi \mid [☺]_i\varphi \mid [☹]_i\varphi \mid [\pi]\varphi,
\end{aligned}$$

where α ranges over $\mathcal{L}_0$ and i ranges over Agt. $\mathcal{L}_{\mathsf{prog}}$ is the language of programs. It includes atomic programs for private belief expansion ($+_i\alpha$) and private forgetting ($-_i\alpha$) as well as complex programs of sequential composition (;), non-deterministic choice ($\cup$) and test (?). The formula $[\pi]\varphi$ is read "φ holds after program π has been executed". The dual of the operator $[\pi]$ is defined in the usual way: $\langle\pi\rangle\varphi \stackrel{\text{def}}{=} \neg[\pi]\neg\varphi$. The formula $\langle\pi\rangle\varphi$ is read "there exists an execution of the program π at the end of which φ holds".

To be able to interpret the language $\mathcal{L}_{\mathsf{dyn}}$ we extend the satisfaction relation of Definition 8 as follows.

Definition 10 (Satisfaction relation, cont.)**.** *Let $(S, U) \in \mathbf{M}$. We define:*

$$\begin{aligned}
(S, U) \models [\pi]\varphi \quad & \textit{if} \quad \forall S' \in U, \textit{ if } S\mathcal{P}_{\pi}^{U}S' \textit{ then}\\
& \qquad (S', U) \models \varphi; \textit{ with}\\
S\mathcal{P}_{+_i\alpha}^{U}S' \quad & \textit{iff} \quad V = V', B_i^{+_i\alpha} = B_i \cup \{\alpha\} \textit{ and}\\
& \qquad \forall j \neq i, B_j^{+_i\alpha} = B_j,\\
S\mathcal{P}_{-_i\alpha}^{U}S' \quad & \textit{iff} \quad V = V', B_i^{-_i\alpha} = B_i \setminus \{\alpha\} \textit{ and}\\
& \qquad \forall j \neq i, B_j^{-_i\alpha} = B_j,\\
S\mathcal{P}_{\pi_1;\pi_2}^{U}S' \quad & \textit{iff} \quad \exists S'' \in U \textit{ such that } S\mathcal{P}_{\pi_1}^{U}S'' \textit{ and}\\
& \qquad S''\mathcal{P}_{\pi_2}^{U}S',\\
S\mathcal{P}_{\pi_1\cup\pi_2}^{U}S' \quad & \textit{iff} \quad S\mathcal{P}_{\pi_1}^{U}S' \textit{ or } S\mathcal{P}_{\pi_2}^{U}S',\\
S\mathcal{P}_{?\varphi}^{U}S' \quad & \textit{iff} \quad S' = S \textit{ and } (S, U) \models \varphi.
\end{aligned}$$

The dynamic modality $[\pi]$ is interpreted in the expected way: φ holds after program π is executed if φ holds at every state which is accessible from the actual one by executing program π. The atomic program $+_i\alpha$ for private belief expansion expands agent i's belief base with the formula α, while $-_i\alpha$ for private forgetting removes formula α from agent i's belief base. They are private operations since they keep the belief bases of all agents different from i unchanged. Sequential composition, non-deterministic choice and test are interpreted in the usual way. Note that the binary relation $\mathcal{P}_{\pi}^{U}$ is parameterized with U to recall the context with respect to which the formulas have to be evaluated. Let us illustrate the dynamic language $\mathcal{L}_{\mathsf{dyn}}$ with an example.

Example 1. *A mother enters the room of her child Bob after hearing a loud noise coming from there. She sees that there is a big mess in the room. The mother's goal is to motivate Bob to tidy up the room. She has to choose the right* *combination of speech acts to achieve her goal. We suppose the mother's repertoire of speech acts includes the following three speech acts: SA1: "You won't be allowed to watch TV this afternoon, if you do not tidy up your room!"; SA2: "I'll prepare some good chocolate crepes for you, if you tidy up your room!"; SA3: "Don't worry, you won't necessarily get tired from tidying up the room, it will only take a few minutes!". Moreover, we suppose Bob has the following information in his belief base: i) tidying up the room (td_{Bob}) is a tiring activity (ti_{Bob}), ii) being tired is a bad thing, iii) a crepe-based snack (cr_{Bob}) is a good thing, iv) being deprived of TV ($\neg tv_{Bob}$) is a bad thing. In formal terms, let $Agt = \{Bob\}$. The initial state is $S_0 = \big(B_{Bob}, V_0\big)$ with*

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

and $V_0 = \emptyset$. The speech acts SA1, SA2 and SA3 are represented formally by the atomic programs $+_{Bob}\alpha_1$, $+_{Bob}\alpha_2$ and $-_{Bob}\alpha_3$ with $\alpha_1 \stackrel{def}{=} \neg td_{Bob} \rightarrow \neg tv_{Bob}$, $\alpha_2 \stackrel{def}{=} td_{Bob} \rightarrow cr_{Bob}$ and $\alpha_3 \stackrel{def}{=} td_{Bob} \rightarrow ti_{Bob}$. It is routine to verify that:

$$(S_0, \mathbf{S}_\Gamma) \models (td_{Bob} \prec^{\mathsf{real}}_{Bob} \neg td_{Bob}), \tag{6}$$

with $\Gamma = (B_{Bob})$. This means that at the initial state S_0 Bob realistically prefers not tidying up his room to tidying it up. Moreover, it is routine to verify that:

$$(S_0, \mathbf{S}_\Gamma) \models [\pi_{talk}](\neg td_{Bob} \prec^{\mathsf{real}}_{Bob} td_{Bob}), \tag{7}$$

with $\pi_{talk} \stackrel{def}{=} \bigcup_{\epsilon,\epsilon' \in \{+_{Bob}\alpha_1, +_{Bob}\alpha_2, -_{Bob}\alpha_3\}: \epsilon \neq \epsilon'} \epsilon; \epsilon'$. This means any sequence of two different speech acts from the mother's speech act repertoire is sufficient to reverse Bob's realistic preference and, consequently, to make the mother achieve her goal.

Note that the example can be generalized to an arbitrary set of children $Agt = \{1, \dots, n\}$ by replacing the information "tidying up the room is a tiring activity" by the information "tidying up the room without the help of the others *is a tiring activity" in the belief base of each child and keeping everything else the same as in the single-agent case. In particular, in the general case we should consider the initial state $S_0 = \big((B_i)_{i \in Agt}, V_0\big)$ such that $B_i = \big\{(\bigwedge_{j \neq i} \neg td_j \wedge td_i) \rightarrow ti_i, ti_i \rightarrow \mathsf{bad}_i, cr_i \rightarrow \mathsf{good}_i, \neg tv_i \rightarrow \mathsf{bad}_i\big\}$ and $V_0 = \emptyset$. Then, we should check the following statement instead of the statement (6) given above:*

$$(S_0, \mathbf{S}_\Gamma) \models \bigwedge_{i \in Agt} \big((\bigwedge_{j \neq i} \neg td_j \wedge td_i) \prec^{\mathsf{real}}_i (td_i \rightarrow \bigvee_{j \neq i} td_j)\big), \tag{8}$$

with $\Gamma = (B_i)_{i \in Agt}$. The formula to be checked expresses the fact that at the initial state every child realistically prefers being helped by someone to tidy up the room to tidying up the room on his own.

Note that the set $\mathbf{S}_\Gamma$ in the Example 1 corresponds to the state space in a Kripke model used to represent the scenario. It is clearly exponential in the size of Γ. But $\mathbf{S}_\Gamma$ and the children' accessibility relations do not need to be explicitly represented in the formulation of the model checking problem. We only need to represent the initial state and the set Γ. So, we can represent the example in an exponentially more succinct way than in the traditional Kripke semantics of epistemic logic.

The next section is devoted to exploring the model checking problem for the language $\mathcal{L}_{\mathsf{dyn}}$ from a complexity and algorithmic perspective. The generalization of the example to an arbitrary set of agents will turn out to be useful since we will use it to test the performance of our model checking algorithm as a function of the number of children.

Model Checking

The model checking problem for $\mathcal{L}_{\mathsf{dyn}}$ is defined as follows: given 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 $\mathbf{S}_\Gamma$, and a formula $\varphi_0 \in \mathcal{L}_{\mathsf{dyn}}$, decide whether $(S_0, \mathbf{S}_\Gamma) \models \varphi_0$ holds.

Theorem 3. *The model checking problem for $\mathcal{L}_{\mathsf{dyn}}$ is PSPACE-complete.*

PSPACE-hardness comes from the PSPACE-hardness of the model checking for the single-agent fragment of $\mathcal{L}$ with only the implicit belief operator $\Box_i$ (Lorini 2019). PSPACE-membership comes from the poly-time reduction into TQBF (true quantified binary formulas) that follows.

Reduction into TQBF

We use abstract symbols s (the reader may think of them as integers) to denote *depths* in the translation of a $\mathcal{L}$-formula into a QBF-formula. We introduce TQBF propositional variables $x_{\alpha,s}$ for all $\alpha \in \mathcal{L}_0$ and for all symbols s. The variables indexed by s are said to be of level s.

Let X_s be the set containing exactly all formulas $x_{\triangle_i\alpha,s}$ $x_{\triangle_i(\alpha \rightarrow \mathsf{good}_i),s}$ and $x_{\triangle_i(\alpha \rightarrow \mathsf{bad}_i),s}$ with $\alpha \in \Gamma_i$ for any agent i, and all $x_{p,s}$ with p appearing in Γ or φ_0.

Let us give some macros that will be used in the reduction.

- The valuations of the two states are equal:

$$\mathrm{eq}_{\mathsf{prop}}(s, s') := \bigwedge_p (x_{p,s} \leftrightarrow x_{p,s'}).$$

- Agent j's belief bases in the two states are equal:

$$\mathrm{eq}_j(s, s') := \bigwedge_{\alpha \in \Gamma_j} (x_{\triangle_j\alpha,s} \leftrightarrow x_{\triangle_j\alpha,s'}).$$

- Agent i's belief bases in the two states are equal, except for α which is added in the second state:

$$\mathrm{eq}_i^{+\alpha}(s, s') := x_{\triangle_i\alpha,s'} \wedge \bigwedge_{\beta \in \Gamma_i : \beta \neq \alpha} (x_{\triangle_i\beta,s} \leftrightarrow x_{\triangle_i\beta,s'}).$$

- Agent i's belief bases in the two states are equal, except for α which is removed from the second state:

$$\mathrm{eq}_i^{-\alpha}(s, s') := \neg x_{\triangle_i\alpha,s'} \wedge \bigwedge_{\beta \in \Gamma_i : \beta \neq \alpha} (x_{\triangle_i\beta,s} \leftrightarrow x_{\triangle_i\beta,s'}).$$

We now give the translation from $\mathcal{L}$ to QBF. In this definition, we write $\forall X_s$ for the sequence $\forall x_1 \ldots \forall x_m$ where $x_1, \ldots, x_m$ is any[2] enumeration of the variables in X_s.

Definition 11. *For all state symbols s, s', we define a function tr_s that maps any formula of $\mathcal{L}$ to a QBF-formula, and a function $tr^{prog}_{s,s'}$ that maps any program π to a a QBF-formula by mutual induction as follows:*

$$\begin{aligned}
tr_s(p) &:= x_{p,s}\\
tr_s(\neg\varphi) &:= \neg tr_s(\varphi)\\
tr_s(\varphi \wedge \psi) &:= tr_s(\varphi) \wedge tr_s(\psi)\\
tr_s(\triangle_i \alpha) &:= \begin{cases} x_{\triangle_i \alpha, s}, & \textit{if } \alpha \in \Gamma_i\\ \bot, & \textit{otherwise}\end{cases}\\
tr_s(\Box_i \varphi) &:= \forall X_{s''}(E_{i,s,s''} \to tr_{s''}(\varphi))\\
tr_s(☺_i \varphi) &:= \forall X_{s''}(tr_{s''}(\varphi) \to A_{i,s,s''})\\
tr_s(☹_i \varphi) &:= \forall X_{s''}(tr_{s''}(\varphi) \to R_{i,s,s''})\\
tr_s([☺]_i \varphi) &:= \forall X_{s''}(tr_{s''}(\varphi) \wedge E_{i,s,s''} \to A_{i,s,s''})\\
tr_s([☹]_i \varphi) &:= \forall X_{s''}(tr_{s''}(\varphi) \wedge E_{i,s,s''} \to R_{i,s,s''})\\
tr_s([\pi]\varphi) &:= \forall X_{s''}(tr^{prog}_{s,s''}(\pi) \to tr_{s''}(\varphi))
\end{aligned}$$

with

$$E_{i,s,s''} := \bigwedge_{\alpha \in \Gamma_i} x_{\triangle_i \alpha, s} \to tr_{s''}(\alpha),$$

$$A_{i,s,s''} := \bigvee_{(\alpha \to \mathsf{good}_i) \in \Gamma_i} x_{\triangle_i(\alpha \to \mathsf{good}_i), s} \wedge tr_{s''}(\alpha),$$

$$R_{i,s,s''} := \bigvee_{(\alpha \to \mathsf{bad}_i) \in \Gamma_i} x_{\triangle_i(\alpha \to \mathsf{bad}_i), s} \wedge tr_{s''}(\alpha),$$

and $tr^{prog}_{s,s'}$ is given by

$$\begin{aligned}
tr^{prog}_{s,s'}(+_i \alpha) &:= \bigwedge_{j \neq i} eq_j(s, s')\\
&\quad \wedge eq_i^{+\alpha}(s, s') \wedge eq_{prop}(s, s')\\
tr^{prog}_{s,s'}(-_i \alpha) &:= \bigwedge_{j \neq i} eq_j(s, s')\\
&\quad \wedge eq_i^{-\alpha}(s, s') \wedge eq_{prop}(s, s')\\
tr^{prog}_{s,s'}(\pi_1; \pi_2) &:= \exists X_{s''}(tr^{prog}_{s,s''}(\pi_1) \wedge tr^{prog}_{s'',s'}(\pi_2))\\
tr^{prog}_{s,s'}(\pi_1 \cup \pi_2) &:= tr^{prog}_{s,s'}(\pi_1) \vee tr^{prog}_{s,s'}(\pi_2)\\
tr^{prog}_{s,s'}(?\varphi) &:= \bigwedge_j eq_j(s, s') \wedge eq_{prop}(s, s') \wedge tr_s(\varphi)
\end{aligned}$$

where s'' in the clauses $tr_s(\Box_i\varphi)$, $tr_s(☺_i\varphi)$, $tr_s(☹_i\varphi)$, $tr_s([☺]_i\varphi)$, $tr_s([☹]_i\varphi)$, $tr_s([\pi]\varphi)$ and $tr^{prog}_{s,s'}(\pi_1;\pi_2)$ is each time a fresh abstract symbol (fresh means that this symbol was not used before).

State S (resp. S') is represented by (the truth values of variables in) X_s (resp. $X_{s''}$). Given a state symbol s, given a formula φ of $\mathcal{L}_{\mathsf{dyn}}$, the intended meaning of $tr_s(\varphi)$ is that formula φ holds in the state S represented by the truth values of variables in X_s. The translation of $\triangle_i \alpha$ is $\bot$ when α is not in the corresponding belief base. The translation $tr_s(\Box_i\varphi)$ directly reflects the semantics of $\Box_i\varphi$; the same for the other operators $☺_i$, $☹_i$, $[☺]_i$, $[☹]_i$. Formula $E_{i,s,s''}$ reformulates $S\mathcal{E}_i S'$, and similarly $A_{i,s,s''}$ and $R_{i,s,s''}$ reformulate $S\mathcal{A}_i S'$ and $S\mathcal{R}_i S'$ respectively.

For the translation $tr_s([\pi]\varphi)$, we use $tr^{\mathrm{prog}}_{s,s''}(\pi)$ which is true iff $S\mathcal{P}^U_\pi S'$ where S and S' are respectively represented by X_s and $X_{s'}$. The definition of $tr^{\mathrm{prog}}_{s,s''}(\pi)$ reflects Definition 10.

In the reduction, we also use formula $\mathrm{desc}_{S_s}(X_s)$ which expresses that variables in X_s represent a given state S_0:

$$\mathrm{desc}_{S_0}(X_s) := \bigwedge_{i \in Agt} \left(\bigwedge_{\alpha \in B_i} x_{\triangle_i \alpha, s} \wedge \bigwedge_{\alpha \in \Gamma_i \setminus B_i} \neg x_{\triangle_i \alpha, s} \right) \wedge \bigwedge_{p \in V} x_{p,s} \wedge \bigwedge_{p \notin V} \neg x_{p,s}.$$

The reduction from the model checking for $\mathcal{L}_{\mathsf{dyn}}$ into TQBF is given in the following proposition.

Proposition 3. *Let $\varphi_0 \in \mathcal{L}_{\mathsf{dyn}}$ and S_0 be a state. We have $(S_0, \mathbf{S}_\Gamma) \models \varphi_0$ iff $\exists X_s(\mathrm{desc}_{S_0}(X_s) \wedge tr_s(\varphi_0))$ is QBF-true.*

Implementation

In order to verify its feasibility, we implemented a symbolic model checker for the static part of the language in Haskell (see Code URL on page 1) which uses the reduction to TQBF. The resulting QBF-formula is solved using the binary decision diagram (BDD) library HasCacBDD (Gattinger 2023). It was compiled with GHC 9.4.8 in a MacBook Air with a 1.6 GHz Dual-Core Intel Core i5 processor and 16,537 GB of RAM, running macOS Sonoma 14.4.1.

$\lvert Agt \rvert$	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 3: Model checker performance on Example 1.

Table 3 shows the performance of the implementation on different instances of Example 1. The execution times correspond to the elapsed time to model check Equation (8). The number of states in the model $|\mathbf{S}_\Gamma|$ gives an idea of the size of the task, and also the reason why a naive implementation would not be feasible. The number of states corresponds to 2 to the power of the number of variables in the set X_s defined above. The latter corresponds to the number of propositional variables plus 3 times the number of formulas in each agent's belief base, i.e., $|Atm| + 3 \times \left|\bigcup_{i \in Agt} \Gamma_i\right|$.

Conclusion

We have presented a modal logic framework for cognitive attitudes of agents whose semantics relies on belief bases and in which models are represented succinctly. It allows us to reason about mental attitudes of both epistemic and motivational type including complete attraction and repulsion,

[2]The semantics does not depend on the particular order.

realistic attraction and repulsion, and the derived concepts of motivation, demotivation, indifference, ambivalence and preference. We have supported our contribution with a number of results about axiomatics and expressiveness, and with a theoretical and experimental analysis of model checking based on a reduction into TQBF.

We did not assume an agent's belief base to be globally consistent due to the presence of belief expansion operations in the dynamic setting. Indeed, a belief expansion operation could make an agent's belief base inconsistent. For example, if an agent explicitly believes that α and that $\alpha \rightarrow \beta$, after expanding its belief base with $\neg\beta$, its belief base will become inconsistent. To be able to handle belief dynamics while preserving the requirement that belief bases are globally consistent, we should replace belief expansion with revision. However, this replacement would make the overall semantics more complex and convoluted. Indeed, as shown in (Lorini and Schwarzentruber 2021), in order to model belief revision in the belief base semantics we must compute the intersection of all maximal consistent subsets of the belief base which contain the input formula. The interested reader can find the technical details of the extension with belief revision as well as the reduction of its model checking problem into TQBF in the extended version of the paper (see Extended version URL on page 1). Interestingly, the PSPACE model checking procedure extends to revision in a natural way.

The attraction and repulsion modalities ☺_i and ☹_i are symmetric with respect to the "good"/"bad" dimension. A further interesting notion we plan to consider in future work is 'strong' motivation. Being 'strongly' motivated by φ is captured by the condition that every state satisfying φ is both attractive and not repulsive. In other words, for an agent to be strongly motivated by φ, it must be the case that the presence of φ makes a situation attractive and prevents a situation from being repulsive. Note that this notion of 'strong' motivation is not definable from the other modalities.

The notion of preference we have considered in the paper is defined from purely qualitative notions of "being motivated" and "being demotivated". In future work we also plan to consider graded notions of being motivated/demotivated (i.e, the agent is motivated/demotivated by φ with a certain strength $k \in \mathbb{N}$). This would allow us to refine the notion of preference, as follows: an agent prefers φ to ψ iff it is motivated by φ more than it is motivated by ψ, or it is demotivated by ψ more than it is demotivated by φ.

Acknowledgments

Emiliano Lorini and François Schwarzentruber are supported by the ANR EpiRL project ANR-22-CE23-0029. Elise Perrotin has benefited from the support of the JSPS KAKENHI Grant Number JP21H04905. Tiago de Lima is supported by the ANR AI Chair project "Responsible AI" (grant number ANR-19-CHIA-0008).

References

Albarracín, D.; Sunderrajan, A.; Lohmann, S.; and M. Sally Chan, D. J. 2018. The Psychology of Attitudes, Motivation, and Persuasion. In Albarracín, D.; and Johnson, B. T., eds., *The Handbook of Attitudes, Volume 1: Basic Principles, 2nd Edition*, 1–42. Routledge.

Azimipour, S.; and Naumov, P. 2021. If You're Happy, Then You Know It: The Logic of Happiness... and Sadness. *CoRR*, abs/2101.00485.

Boutilier, C. 1994. Towards a logic for qualitative decision theory. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, 75–86. AAAI Press.

Bradley, R.; and List, C. 2009. Desire-as-belief revisited. *Mind*, 69(1): 31–37.

Charrier, T.; Pinchinat, S.; and Schwarzentruber, F. 2019. Symbolic model checking of public announcement protocols. *Journal of Logic and Computation*, 29(8): 1211–1249.

Charrier, T.; and Schwarzentruber, F. 2015. Arbitrary Public Announcement Logic with Mental Programs. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015)*, 1471–1479. ACM.

Charrier, T.; and Schwarzentruber, F. 2017. A Succinct Language for Dynamic Epistemic Logic. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2017)*, 123–131. ACM.

Charrier, T.; and Schwarzentruber, F. 2018. Complexity of Dynamic Epistemic Logic with Common Knowledge. In *Proceedings of the 12th conference on Advances in Modal Logic 12*, 103–122. College Publications.

Cohen, P. R.; and Levesque, H. J. 1990. Intention is choice with commitment. *Artificial Intelligence*, 42: 213–261.

Cooper, M. C.; Herzig, A.; Maffre, F.; Maris, F.; Perrotin, E.; and Régnier, P. 2021. A lightweight epistemic logic and its application to planning. *Artif. Intell.*, 298: 103437.

de Jongh, D.; and Liu, F. 2009. Preference, Priorities and Belief. In Gruene-Yanoff, T.; and Hansson, S. O., eds., *Preference Change*, 85–108. Springer.

Dennett, D. C. 1987. *The Intentional Stance*. Cambridge, Massachusetts: MIT Press.

Dubois, D.; Lorini, E.; and Prade, H. 2017. The Strength of Desires: A Logical Approach. *Minds and Machines*, 27(1): 199–231.

Fagin, R.; Halpern, J.; Moses, Y.; and Vardi, M. 1995. *Reasoning about Knowledge*. Cambridge: MIT Press.

Fagin, R.; and Halpern, J. Y. 1987. Belief, Awareness, and Limited Reasoning. *Artificial Intelligence*, 34(1): 39–76.

Gargov, G.; and Goranko, V. 1979. Minimal deontic logics. *Bulletin of the Section of Logic*, 8(1): 36–42.

Gargov, G.; Passy, S.; and Tinchev, T. 1987. Modal Environment for Boolean Speculations. In Skordev, D. G., ed., *Mathematical Logic and Its Applications*, 253–263. Boston, MA: Springer.

Gattinger, M. 2023. HasCacBDD. Version 0.1.0.4, Feb. 3, 2023.

Goranko, V. 1983. Modal Definability in Enriched Languages. *Notre Dame Journal of Formal Logic*, 31(1): 81–105.

Herzig, A.; and Longin, D. 2004. C&L Intention Revisited. In *Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR2004)*, 527–535. AAAI Press.

Hintikka, J. 1962. *Knowledge and Belief*. New York: Cornell University Press.

Humberstone, I. L. 1983. Inaccessible worlds. *Notre Dame Journal of Formal Logic*, 24: 346–352.

Humberstone, I. L. 1992. Direction of fit. *Mind*, 101(401): 59–83.

Icard, T. F.; Pacuit, E.; and Shoham, Y. 2010. Joint Revision of Beliefs and Intention. In *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning (KR 2010)*, 572–574. AAAI Press.

Karlsson, M. M. 2000. Rational Ends: Humean and Non-Humean Considerations. *Sats-Nordic Journal of Philosophy*, 1(2).

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

Lewis, D. 1988. Desire as belief. *Mind*, 97(387): 323–332.

Lewis, D. 1996. Desire as belief II. *Mind*, 105(418): 303–313.

Liu, F. 2011. *Reasoning about preference dynamics*, volume 354. Springer.

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

Lorini, E. 2019. Exploiting Belief Bases for Building Rich Epistemic Structures. In *Proceedings of the Seventeenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2019)*, volume 297 of *EPTCS*, 332–353.

Lorini, E. 2020. Rethinking epistemic logic with belief bases. *Artificial Intelligence*, 282.

Lorini, E. 2021. A Qualitative Theory of Cognitive Attitudes and their Change. *Theory and Practice of Logic Programming*, 21(4): 428–458.

Lorini, E.; and Herzig, A. 2008. A logic of intention and attempt. *Synthese*, 163(1): 45–77.

Lorini, E.; and Rapion, E. 2022. Logical Theories of Collective Attitudes and the Belief Base Perspective. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022)*. ACM.

Lorini, E.; and Schwarzentruber, F. 2021. Multi-Agent Belief Base Revision. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI 2021)*, 1959–1965. ijcai.org.

Meyer, J. J. C.; van der Hoek, W.; and van Linder, B. 1999. A Logical Approach to the Dynamics of Commitments. *Artificial Intelligence*, 113(1-2): 1–40.

Searle, J. 2001. *Rationality in Action*. Cambridge: MIT Press.

Shoham, Y. 2009. Logical Theories of Intention and the Database Perspective. *Journal of Philosophical Logic*, 38(6): 633–647.

Souza, M. 2016. *Choices that make you change your mind: a Dynamic Epistemic Logic approach to the semantics of BDI agent programming languages*. Ph.D. thesis, Universidade Federal Do Rio Grande Do Sul.

Souza, M.; and Moreira, A. 2021. Belief base change as priority change: A study based on dynamic epistemic logic. *Journal of Logical and Algebraic Methods in Programming*, 122.

Stalnaker, R. 2002. Common ground. *Linguistics and Philosophy*, 25(5-6): 701–721.

Su, K.; Sattar, A.; Lin, H.; and Reynolds, M. 2007. A Modal Logic for Beliefs and Pro Attitudes. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07)*, 496–501. AAAI Press.

van Benthem, J.; Girard, P.; and Roy, O. 2009. Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic*, 38: 83–125.

van Benthem, J.; Grossi, D.; and Liu, F. 2014. Priority structures in deontic logic. *Theoria*, 80(2): 116–152.

van Benthem, J.; van Eijck, J.; Gattinger, M.; and Su, K. 2018. Symbolic model checking for Dynamic Epistemic Logic - S5 and beyond. *Journal of Logic and Computation*, 28(2): 367–402.

van der Hoek, W.; Troquard, N.; and Wooldridge, M. 2011. Knowledge and control. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, 719–726. IFAAMAS.

Wooldridge, M. 2000. *Reasoning about rational agents*. Cambridge: MIT Press.

Wright, G. H. V. 1963. *The logic of preference*. Edinburgh University Press.

Wright, G. H. V. 1972. The logic of preference reconsidered. *Theory and Decision*, 3: 140–169.