TPLP **21** (4): 428–458, 2021. © The Author(s), 2021. Published by Cambridge University Press
doi:10.1017/S1471068421000053 First published online 6 April 2021

A Qualitative Theory of Cognitive Attitudes and their Change

EMILIANO LORINI
IRIT-CNRS, Toulouse University, Toulouse, France
(*e-mail:* Emiliano.Lorini@irit.fr)

submitted 13 January 2020; revised 16 February 2021;accepted 23 February 2021

Abstract

We present a general logical framework for reasoning about agents' cognitive attitudes of both epistemic type and motivational type. We show that it allows us to express a variety of relevant concepts for qualitative decision theory including the concepts of knowledge, belief, strong belief, conditional belief, desire, conditional desire, strong desire, and preference. We also present two extensions of the logic, one by the notion of choice and the other by dynamic operators for belief change and desire change, and we apply the former to the analysis of single-stage games under incomplete information. We provide sound and complete axiomatizations for the basic logic and for its two extensions.

KEYWORDS: knowledge representation, nonmonotonic reasoning, modal logic, cognitive agents

1 Introduction

Since the seminal work of Hintikka on epistemic logic (Hintikka 1962), of Von Wright on the logic of preference (Wright 1963; 1972) , and of Cohen & Levesque on the logic of intention (Cohen and Levesque 1990), many formal logics for reasoning about cognitive attitudes of agents such as knowledge and belief (Fagin *et al*. 1995), preference (Liu 2011; van Benthem *et al*. 2009), desire (Dubois *et al*. 2017), intention (Shoham 2009; Icard *et al*. 2010), and their combination (Meyer *et al*. 1999; Wooldridge 2000) have been proposed. Generally speaking, these logics are nothing but formal models of rational agency relying on the idea that an agent endowed with cognitive attitudes makes decisions on the basis of what she believes and of what she desires or prefers.

The idea of describing rational agents in terms of their epistemic and motivational attitudes is something that these logics share with classical decision theory and game theory. Classical decision theory and game theory provide a quantitative account of individual and strategic decision-making by assuming that agents' beliefs and desires can be, respectively, modeled by subjective probabilities and utilities. Qualitative approaches to individual and strategic decision-making have been proposed in AI (Boutilier 1994; Doyle and Thomason 1999) to characterize criteria that a rational agent should adopt for making decisions when she cannot build a probability distribution over the set of possible events and her preference over the set of possible outcomes cannot be expressed by a utility function but only by a qualitative ordering over the outcomes. For example, going beyond expected utility maximization, qualitative criteria such as the maximin principle (choose the action that will minimize potential loss) and the maxmax principle

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

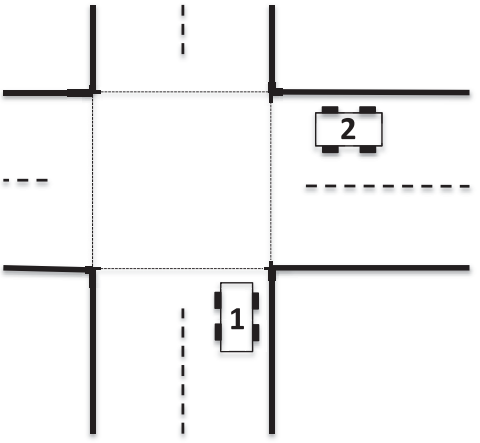


Fig. 1. Crossroad game.

(choose the action that will maximize potential gain) have been studied and axiomatically characterized (Brafman and Tennenholtz 2000, 1996).

The aim of this paper is to present an expressive logical framework for representing both the static and the dynamic aspects of a rich variety of agents' cognitive attitudes in a multi-agent setting. In agreement with philosophical theories (Platts 1979; Searle 1979; Humberstone 1992; Lorini 2017), our logic allows us to distinguish two general categories of cognitive attitudes: *epistemic* attitudes, including belief and knowledge, and *motivational* ones, including desire and preference. Moreover, in agreement with rational choice theory, it allows us to capture a notion of choice which depends on what an agent believes and prefers.[1]

The example depicted in Figure 1 brings to the fore the epistemic and motivational attitudes that are involved in everyday situations whereby artificial agents are supposed to interact. There are two autonomous agents meeting at a crossroad: agent 1 and agent 2. The two agents could be either two mobile robots or two autonomous vehicles. Each agent can decide either to stop or to continue. If an agent stops, then it will lose time. If both agents decide to continue, they will collide and, consequently, each of them will lose time. Therefore, for an agent not to lose time, it has to continue, while the other agent decides to stop.

In this situation, each agent is identified with the set of cognitive attitudes it endorses. For instance, it is reasonable to suppose that the two agents know that in the situation they face, necessarily some of them will lose time and that if one of them loses time by letting the other pass, there will be no collision. On the motivational side, it is reasonable to suppose that each agent is strongly motivated by two desires, namely, the desire not to lose time and the desire to avoid a collision.

On the dynamic side, we consider two basic forms of cognitive attitude change, namely, belief change and desire change. While belief change has been extensively studied in the area of belief revision (Alchourrón *et al*. 1985; Boutilier 1993; Darwiche and Pearl 1997;

[1] Rational choice theory (RCT) is a umbrella term for a family of theories prescribing that an agent should choose the course of action that, according to her beliefs, leads to the most desirable (or most preferred) consequences. In other words, RCT relies on the general assumption that agents make optimal choices in the light of her beliefs, desires, and preferences. See Paternotte (2011) for more details on RCT.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

Spohn 1988; Rott 2009; Benferhat *et al*. 2002) and dynamic epistemic logic (DEL) (van Benthem 2007; Baltag and Smets 2008; van Ditmarsch 2005; Aucher 2005), desire change is far less studied and understood. We will study two basic forms of cognitive attitude revision, namely, *radical* attitude revision and *conservative* attitude revision. While the distinction between radical and conservative belief revision has been drawn before (see, e.g., van Benthem 2007), the distinction between radical and conservative desire revision is new. Radical belief revision by an input φ makes all states at which φ is true more plausible than all states at which φ is false, whereas conservative belief revision by φ simply promotes the most plausible states in which φ is true to the highest plausibility rank, but apart from that, it keeps the old plausibility ordering. For example, suppose in the crossroad game of Figure 1, agent 1 and agent 2 can communicate. Agent 1 informs agent 2 that "if they both lose time, then there will no collision" and agent 2 trusts what agent 1 says. Then, by performing a conservative belief revision, agent 2 will promote the most plausible situations in which the formula announced by 1 is true to the highest plausibility rank. As a consequence, agent 2 will start to believe what 1 has just said.

Symmetrically, radical desire revision by φ makes all states at which φ is true more desirable than all states at which φ is false, whereas conservative desire revision by φ simply demotes the least desirable states in which φ is false to the lowest desirability rank, but apart from that, it keeps the old desirability ordering. For example, suppose in the crossroad game, agent 1 has just learnt that agent 2 is an ambulance which has to transport a patient to the hospital as quickly as possible. Consequently, 1 starts to be altruistically motivated by the fact that 2 does not lose time. Thus, by performing a radical desire revision, agent 1 will start to consider all situations in which 2 does not lose time more desirable than the situations in which it does. This radical desire revision operation leads agent 1 to strongly desire that agent 2 does not lose time.

The paper is organized as follows. In Section 2, we present the semantics and syntax of our logic, called Dynamic Logic of Cognitive Attitudes (DLCA). At the semantic level, it exploits two orderings that capture an agent's comparative plausibility and comparative desirability over states. At the syntactic level, it uses program constructs of dynamic logic (sequential composition, non-deterministic choice, intersection, complement, converse, and test) to build complex cognitive attitudes from simple ones. Following Passy and Tinchev (1991), and Gargov and Goranko (1993), it also exploits nominals in order to axiomatize intersection and complement of programs. In Section 3, we illustrate the expressive power of our logic by using it to formalize a variety of cognitive attitudes of agents including knowledge, belief, strong belief, conditional belief, desire, strong desire, conditional desire, and preference. We instantiate some of these concepts in the crossroad game depicted in Figure 1. In Section 4, we present a sound and complete axiomatization for our logic. In Section 5, we present the first extension of our logic by the notion of choice and apply it to the analysis of single-stage games under incomplete information. Section 6 presents the second extension of our logic by dynamic operators for belief and desire change. In Section 7, we conclude. Formal proofs are given in a technical annex published as supplementary material.

2 Dynamic logic of cognitive attitudes

Let Atm be a countable infinite set of atomic propositions, let Nom be a countable infinite set of nominals disjoint from Atm, and let Agt be a finite set of agents.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

Definition 1 (*Multi-agent cognitive model*)
A multi-agent cognitive model (MCM) is a tuple $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ where:

- W is a set of worlds or states;
- for every $i \in Agt$, $\preceq_{i,P}$ and $\preceq_{i,D}$ are preorders on W and $\equiv_i$ is an equivalence relation on W such that for all $\tau \in \{P, D\}$ and for all $w, v \in W$:

  (C1) $\preceq_{i,\tau} \subseteq \equiv_i$,
  (C2) if $w \equiv_i v$, then $w \preceq_{i,\tau} v$ or $v \preceq_{i,\tau} w$;

- $V : W \longrightarrow 2^{Atm\cup Nom}$ is a valuation function such that for all $w, v \in W$:

  (C3) $V_{Nom}(w) \neq \emptyset$,
  (C4) if $V_{Nom}(w) \cap V_{Nom}(v) \neq \emptyset$, then $w = v$;

  where $V_{Nom}(w) = Nom \cap V(w)$.

$w \preceq_{i,P} v$ means that, according to agent i, v is at least as plausible as w, whereas $w \preceq_{i,D} v$ means that, according to agent i, v is at least as desirable as w. Finally, $w \equiv_i v$ means that w and v are indistinguishable for agent i. For every $w \in W$, $\equiv_i(w)$ is also called agent i's information set at state w. According to Constraint C1, an agent can only compare the plausibility (resp. desirability) of two states in her information set. According to Constraint C2, the plausibility (resp. desirability) of two states in an agent's information set are always comparable. Constraints C3 and C4 capture the two basic properties of nominals: every state is associated with at least one nominal, and there are no different states associated with the same nominal.

Note that there is no connection between binary relations $\preceq_{i,P}$ and $\preceq_{i,D}$. In accord with classical decision and game theory in which an agent's subjective probability and utility function do not interact, we adopt a normative view of epistemic and motivational attitudes according to which an agent's epistemic plausibility and desirability are assumed to be independent.[2] Therefore, we do not consider cognitive biases typical of human reasoning such as wishful thinking, as the tendency to form beliefs according to what is desired in the absence of a clear evidence against it (Marsh and Wallace 2005). Nonetheless, as we will show in Section 3.3, the primitive relations $\preceq_{i,P}$ and $\preceq_{i,D}$ can be combined to obtain a notion of realistic preference which is essential for elucidating the connection between an agent's beliefs and desires and her choices.

We introduce the following modal language $\mathcal{L}_{\mathrm{DLCA}}(Atm, Nom, Agt)$, or simply $\mathcal{L}_{\mathrm{DLCA}}$, for the DLCA:

$$\begin{array}{lll} \pi & ::= & \equiv_i | \preceq_{i,P} | \preceq_{i,D} | \widetilde{\preceq}_{i,P} | \widetilde{\preceq}_{i,D} | \ \pi;\pi' \mid \pi \cup \pi' \mid \pi \cap \pi' \mid -\pi \mid \varphi? \\ \varphi & ::= & p \mid x \mid \neg\varphi \mid \varphi \wedge \varphi' \mid [\pi]\varphi, \end{array}$$

where p ranges over Atm, x ranges over Nom, and i ranges over Agt. The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$, and $\leftrightarrow$ are defined from p, $\neg$, and $\wedge$ in the standard way. The propositional language built from the set of atomic propositions Atm is noted $\mathcal{L}_{\mathsf{PL}}(Atm)$. Note that the sets Atm, Nom, and Agt define the signature of the language $\mathcal{L}_{\mathrm{DLCA}}$. They

[2] The normative view is usually opposed to the descriptive view. The normative view is aimed at describing the reasoning and decision-making of ideal agents conforming to standards of rationality, while the descriptive view is concerned with psychologically realistic cognitive agents who systematically violate standards of rationality and exhibit different types of cognitive bias.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

are not part of the model since every atomic proposition p, nominal x, and modal formula $[\pi]\varphi$ should be interpretable relative to any MCM.

Elements π are called *cognitive programs* or, more shortly, *programs*. The set of all programs is noted $\mathcal{P}(Atm, Nom, Agt)$, or simply, $\mathcal{P}$.

Cognitive programs correspond to the basic constructions of Propositional Dynamic Logic (PDL) (Harel *et al*. 2000): atomic programs of type $\equiv_i$, $\preceq_{i,P}$, $\preceq_{i,D}$, $\preceq^{\sim}_{i,P}$, and $\preceq^{\sim}_{i,D}$, sequential composition (;), non-deterministic choice ($\cup$), intersection ($\cap$), converse ($^-$), and test (?). A given cognitive program π corresponds to a specific configuration of the agents' cognitive states including their epistemic states and their motivational states.

The formula $[\pi]\varphi$ has to be read "φ is true, according to the cognitive program π". As usual, we define $\langle\pi\rangle$ to be the dual operator of $[\pi]$, that is, $\langle\pi\rangle\varphi =_{def} \neg[\pi]\neg\varphi$.

The atomic program $\equiv_i$ represents the standard S5, partition-based, and fully introspective notion of knowledge (Fagin *et al*. 1995; Aumann 1999). $[\equiv_i]\varphi$ has to be read "φ is true according to what agent i knows" or more simply "agent i knows that φ is true", which just means that "φ is true in all worlds that agent i envisages".

The atomic programs $\preceq_{i,P}$ and $\preceq_{i,D}$ capture, respectively, agent i's plausibility ordering and agent i's desirability ordering over facts. In particular, $[\preceq_{i,P}]\varphi$ has to be read "φ is true at all states that, according to agent i, are at least as plausible as the current one", while $[\preceq_{i,D}]\varphi$ has to be read "φ is true at all states that, according to agent i, are at least as desirable as the current one". The atomic programs $\preceq^{\sim}_{i,P}$ and $\preceq^{\sim}_{i,D}$ are the complements of the atomic programs $\preceq_{i,P}$ and $\preceq_{i,D}$, respectively. In particular, $[\preceq^{\sim}_{i,P}]\varphi$ has to be read "φ is true at all states that, according to agent i, are *not* at least as plausible as the current one", while $[\preceq^{\sim}_{i,D}]\varphi$ has to be read "φ is true at all states that, according to agent i, are *not* at least as desirable as the current one". The program constructs ;, $\cup$, $\cap$, $-$, and ? are used to define complex cognitive programs from the atomic cognitive programs. For example, the formula $[\preceq_{i,P} \cup \preceq_{i,D}]\varphi$ has to be read "φ is true at all states that, according to agent i, are either at least as plausible *or* at least as desirable as the current one", whereas the formula $[\preceq_{i,P} \cap \preceq_{i,D}]\varphi$ has to be read "φ is true at all states that, according to agent i, are at least as plausible *and* at least as desirable as the current one".

The following definition provides truth conditions for formulas in $\mathcal{L}_{\mathrm{DLCA}}$:

Definition 2 (*Truth conditions*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then:

$$\begin{aligned}
M, w \models p &\iff p \in V(w),\\
M, w \models x &\iff x \in V(w),\\
M, w \models \neg\varphi &\iff M, w \not\models \varphi,\\
M, w \models \varphi \wedge \psi &\iff M, w \models \varphi \text{ and } M, w \models \psi,\\
M, w \models [\pi]\varphi &\iff \forall v \in W : \text{ if } wR_\pi v \text{ then } M, v \models \varphi,
\end{aligned}$$

where the binary relation R_π on W is inductively defined as follows, with $\tau \in \{P, D\}$:

$$\begin{aligned}
wR_{\equiv_i} v &\text{ iff } w \equiv_i v,\\
wR_{\preceq_{i,\tau}} v &\text{ iff } w \preceq_{i,\tau} v,\\
wR_{\preceq^{\sim}_{i,\tau}} v &\text{ iff } w \equiv_i v \text{ and } w \not\preceq_{i,\tau} v,
\end{aligned}$$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

$$
\begin{aligned}
wR_{\pi;\pi'}v &\text{ iff } \exists u \in W : wR_{\pi}u \text{ and } uR_{\pi'}v,\\
wR_{\pi\cup\pi'}v &\text{ iff } wR_{\pi}v \text{ or } wR_{\pi'}v,\\
wR_{\pi\cap\pi'}v &\text{ iff } wR_{\pi}v \text{ and } wR_{\pi'}v,\\
wR_{-\pi}v &\text{ iff } vR_{\pi}w,\\
wR_{\varphi?}v &\text{ iff } w = v \text{ and } M, w \models \varphi.
\end{aligned}
$$

For notational convenience, we use $wR_{\pi}v$ and $(w, v) \in R_{\pi}$ as interchangeable notations.

We can build a variety of cognitive programs capturing different types of plausibility and desirability relations between possible worlds. For instance, for every $\tau \in \{P, D\}$, we can define:

$$
\begin{aligned}
\succeq_{i,\tau} &=_{def} -\preceq_{i,\tau},\\
\succ_{i,\tau} &=_{def} \succeq_{i,\tau} \cap \preceq^{\sim}_{i,\tau},\\
\succeq^{\sim}_{i,\tau} &=_{def} -\preceq^{\sim}_{i,\tau},\\
\prec_{i,\tau} &=_{def} \preceq_{i,\tau} \cap \succeq^{\sim}_{i,\tau},\\
\approx_{i,\tau} &=_{def} \preceq_{i,\tau} \cap \succeq_{i,\tau}.
\end{aligned}
$$

The five definitions denote, respectively, "at most as plausible (resp. desirable) as", "less plausible (resp. desirable) than", "not at most as plausible (resp. desirable) as", "more plausible (resp. desirable) than", and "equally plausible (resp. desirable) as".

For every formula φ in $\mathcal{L}_{\mathrm{DLCA}}$, we say that φ is valid, noted $\models_{MCM} \varphi$, if and only if for every MCM M and world w in M, we have $M, w \models \varphi$. Conversely, we say that φ is satisfiable if $\neg\varphi$ is not valid.

For a given MCM $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, N, V)$, we define $||\varphi||_M = \{v \in W : M, v \models \varphi\}$ to be the truth set of φ in M. Moreover, for every $w \in W$ and for every $i \in Agt$, we define $||\varphi||_{i,w,M} = \{v \in W : M, v \models \varphi \text{ and } w \equiv_i v\}$ to be the truth set of φ from i's point of view at state w in M.

3 Formalization of cognitive attitudes

In this section, we show how the logic DLCA can be used to model the variety of cognitive attitudes of agents that we have briefly discussed in the introduction.

3.1 Epistemic attitudes

We start with the family of epistemic attitudes by defining a standard notion of belief. We say that an agent believes that φ if and only if φ is true at all states that the agent considers maximally plausible.

Definition 3 (*Belief*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that agent i believes that φ at w, noted $M, w \models \mathsf{B}_i\varphi$, if and only if $Best_{i,P}(w) \subseteq ||\varphi||_M$ where $Best_{i,P}(w) = \{v \in W : w \equiv_i v \text{ and } \forall u \in W, \text{ if } w \equiv_i u, \text{ then } u \preceq_{i,P} v\}$.

As the following proposition highlights, the previous notion of belief is expressible in the logic DLCA by means of the cognitive program $\equiv_i; [\prec_{i,P}]\bot?$.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

Proposition 1
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{B}_i\varphi \text{ iff } M, w \models \big[\equiv_i; [\prec_{i,P}]\bot? \big]\varphi.$$

It is worth noting that the set $Best_{i,P}(w)$ in Definition 3 might be empty, since it is not necessarily the case that the relation $\preceq_{i,P}$ is conversely well-founded.[3] As a consequence, the belief operator B_i does not necessarily satisfy Axiom D, that is, the formula $\mathsf{B}_i\varphi \wedge \mathsf{B}_i\neg\varphi$ is satisfiable in the logic DLCA.

In the literature on epistemic logic (Baltag and Smets 2009), mere belief of Definition 3 is usually distinguished from strong belief. Specifically, we say that an agent strongly believes that φ if and only if, according to agent i, all φ-worlds are strictly more plausible than all $\neg\varphi$-worlds.

Definition 4 (*Strong belief*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that agent i strongly believes that φ at w, noted $M, w \models \mathsf{SB}_i\varphi$, if and only if $\forall v \in ||\varphi||_{i,w,M}$ and $\forall u \in ||\neg\varphi||_{i,w,M} : u \prec_{i,P} v$.

As the following proposition highlights, the previous notion of strong belief is expressible in the logic DLCA by means of the cognitive program $\equiv_i; \varphi?; \preceq_{i,P}$.

Proposition 2
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{SB}_i\varphi \text{ iff } M, w \models \big[\equiv_i; \varphi?; \preceq_{i,P} \big]\varphi.$$

Strong belief that φ implies belief that φ, if the agent envisages at least one state in which φ is true. This property is expressed by the following validity:

$$\models_{MCM} \big(\mathsf{SB}_i\varphi \wedge \langle\equiv_i\rangle\varphi\big) \rightarrow \mathsf{B}_i\varphi \tag{1}$$

Conditional belief is another notion which has been studied by epistemic logicians given its important role in belief dynamics (van Benthem 2007). We say that an agent believes that φ conditional on ψ, or she would believe that φ if she learnt that ψ, if and only if, according to the agent, all most plausible ψ-worlds are also φ-worlds.

Definition 5 (*Conditional belief*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that agent i would believe that φ if she learnt that ψ at w, noted $M, w \models \mathsf{B}_i(\psi, \varphi)$, if and only if $Best_{i,P}(\psi, w) \subseteq ||\varphi||_M$, where $Best_{i,P}(\psi, w) = \{v \in ||\psi||_{i,w,M} : \forall u \in ||\psi||_{i,w,M}, u \preceq_{i,P} v\}$.

Note that $Best_{i,P}(\top, w) = Best_{i,P}(w)$.

As for belief and strong belief, we have a specific cognitive program $\equiv_i; (\psi \wedge [\prec_{i,P}]\neg\psi)?$ corresponding to the belief that φ conditional on ψ, so that the latter can be represented in the language of the logic DLCA.

[3] This means that there could be a world v such that $w \equiv_i v$ and there is a $\preceq_{i,P}$-infinite ascending chain from v.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

Proposition 3
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{B}_i(\psi, \varphi) \text{ iff } M, w \models \big[\equiv_i; (\psi \wedge [\prec_{i,P}]\neg\psi)?\big]\varphi.$$

3.2 Motivational attitudes I: Desires

The first kind of motivational attitude we consider is desire. Following Dubois *et al.* (2017), we say that an agent desires that φ if and only if all states that the agent envisages at which φ is true are not minimally desirable for her. In other words, desiring that φ consists in having some degree of attraction for all situations in which φ is true, since minimally desirable states are those to which the agent is not attracted at all.

Definition 6 (*Desire*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that agent i desires that φ at w, noted $M, w \models \mathsf{D}_i\varphi$, if and only if $Worst_{i,D}(w) \cap ||\varphi||_M = \emptyset$, where $Worst_{i,D}(w) = \{v \in W : w \equiv_i v \text{ and } \forall u \in W, \text{ if } w \equiv_i u \text{ then } v \preceq_{i,D} u\}$.

As the following proposition highlights, the previous notion of desire is characterized by the cognitive program $\equiv_i; [\succ_{i,D}]\bot?$.

Proposition 4
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{D}_i\varphi \text{ iff } M, w \models \big[\equiv_i; [\succ_{i,D}]\bot?\big]\neg\varphi.$$

Similarly to the set $Best_{i,P}(w)$ in Definition 3, the set $Worst_{i,D}(w)$ in Definition 6 might be empty, since it is not necessarily the case that the relation $\preceq_{i,D}$ is well-founded.[4] As a consequence, desires are not necessarily consistent and an agent may desire the tautology, that is, the formulas $\mathsf{D}_i\varphi \wedge \mathsf{D}_i\neg\varphi$ and $\mathsf{D}_i\top$ are satisfiable in the logic DLCA. As emphasized by Dubois *et al.* (2017), this notion of desire satisfies the following property:

$$\models_{MCM} \mathsf{D}_i\varphi \rightarrow \mathsf{D}_i(\varphi \wedge \psi). \tag{2}$$

Indeed, if an agent has some degree of attraction for all situations in which φ is true then, clearly, it should have some degree of attraction for all situations in which $\varphi \wedge \psi$ is true, since all $\varphi \wedge \psi$-situations are also φ-situations.

Note that there is no counterpart of this property for belief, as the formula $\mathsf{B}_i\varphi \rightarrow \mathsf{B}_i(\varphi \wedge \psi)$ is clearly not valid.[5]

It is a property that the notion of desire shares with the *open reading* of the concept of permission studied in the area of deontic logic (see, e.g. Anglberger *et al.* 2015; Lewis 1979).[6] One way of blocking this inference is by strengthening the notion of desire. We

[4] This means that there could be a world v such that $w \equiv_i v$ and there is a $\preceq_{i,D}$-infinite descending chain from v.

[5] See Dubois *et al.* (2017) for more details about the differences between the notion of belief and the notion of desire.

[6] According to deontic logicians, there are at least two candidate readings of the statement "φ is permitted": (i) every instance of φ is OK according to the normative regulation, and (ii) at least one instance of φ (but possibly not all) is OK according to the normative regulation. The former is the so-called *open reading* of permission.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

say that an agent strongly desires that φ if and only if, according to agent i, all φ-worlds are strictly more desirable than all $\neg\varphi$-worlds.

Definition 7 (*Strong desire*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that agent i strongly desires that φ at w, noted $M, w \models \mathsf{SD}_i\varphi$, if and only if $\forall v \in ||\varphi||_{i,w,M}$ and $\forall u \in ||\neg\varphi||_{i,w,M} : u \prec_{i,D} v$.

As for desire, there exists a cognitive program which characterizes strong desire, namely, the program $\equiv_i; \varphi?; \preceq_{i,D}$.

Proposition 5
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{SD}_i\varphi \text{ iff } M, w \models \left[\equiv_i; \varphi?; \preceq_{i,D} \right]\varphi.$$

We have that strong desire implies desire, when the agent envisages at least one state in which φ is false:

$$\models_{MCM} (\mathsf{SD}_i\varphi \wedge \langle\equiv_i\rangle\neg\varphi) \rightarrow \mathsf{D}_i\varphi. \tag{3}$$

Unlike desire, it is not necessarily the case that strongly desiring that φ implies strongly desiring that $\varphi \wedge \psi$, that is, $\mathsf{SD}_i\varphi \wedge \neg\mathsf{SD}_i(\varphi \wedge \psi)$ is satisfiable in the logic DLCA. Indeed, strongly desiring that φ is compatible with envisaging a situation in which $\varphi \wedge \psi$ holds and another situation in which $\varphi \wedge \neg\psi$ holds such that the first situation is less desirable than the second.

The last motivational attitude we consider is conditional desire which parallels the notion of conditional belief of Definition 5. We say that an agent desires that φ conditional on ψ, or she would desire that φ if she started to desire that ψ, if and only if, according to agent i, there is no least desirable $\neg\psi$-world which is also a φ-world. The idea behind this notion is the following. If the agent started to desire that ψ, all ψ-worlds would start to have some degree of attraction for her and the least desirable $\neg\psi$-worlds would become the minimally desirable worlds. Therefore, the fact that there is no least desirable $\neg\psi$-world which is also a φ-world guarantees that, if the agent started to desire that ψ, no φ-world would be included in the set of minimally desirable worlds for the agent. The latter means that, if the agent started to desire that ψ, all φ-worlds would have some degree of attraction for her and she would desire that φ.

Definition 8 (*Conditional desire*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that agent i would desire that φ if she started to desire that ψ at w, noted $M, w \models \mathsf{D}_i(\psi, \varphi)$, if and only if $Worst_{i,D}(\neg\psi, w) \cap ||\varphi||_M = \emptyset$, with $Worst_{i,D}(\neg\psi, w) = \{v \in ||\neg\psi||_{i,w,M} : \forall u \in ||\neg\psi||_{i,w,M}, v \preceq_{i,D} u\}$.

As for the other cognitive attitudes, there is a specific cognitive program which characterizes conditional desire.

Proposition 6
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{D}_i(\psi, \varphi) \text{ iff } M, w \models \left[\equiv_i; (\neg\psi \wedge [\succ_{i,D}]\psi)? \right]\neg\varphi.$$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

In Section 2, we emphasized that the relations $\preceq_{i,P}$ and $\preceq_{i,D}$ do not interact since our logic is aimed at modeling ideal rational agents with no wishful thinking and, more generally, with no cognitive biases. We conclude this section by showing how the assumption of independence between epistemic plausibility and desirability could be relaxed and, consequently, how wishful thinking could be modeled in our framework.

A wishful thinker is nothing but an agent who systematically believes what she strongly desires in the absence of a reason to believe the contrary. Such a connection between the agent's beliefs and desires is captured by the following "wishful thinking" (WT) constraint on MCMs:

$$\forall w \in W : Best_{i,P}(w) \subseteq Best_{i,D}(w) \text{ or } Best_{i,P}(w) \subseteq Worst_{i,D}(w),$$

where $Best_{i,P}(w)$ and $Worst_{i,D}(w)$ are defined as in Definitions 3 and 6, and $Best_{i,D}(w) = \{v \in W : w \equiv_i v \text{ and } \forall u \in W, \text{ if } w \equiv_i u \text{ then } u \preceq_{i,D} v\}$. It is routine to verify that if the MCM $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ satisfies the previous constraint WT, then the following holds for every $w \in W$:

$$M, w \models (\mathsf{SD}_i\varphi \wedge \neg \mathsf{B}_i \neg \varphi) \rightarrow \mathsf{B}_i\varphi.$$

We leave for future work an in-depth analysis of the variant of our logic in which wishful thinking is enabled.

3.3 Motivational attitudes II: Preferences

We consider two views about comparative statements between formulas of the form "agent i prefers φ to ψ" or "the state of affairs φ is for agent i at least as good as the state of affairs ψ". According to the optimistic view, when assessing whether φ is at least as good as ψ, an agent focuses on the best φ-situations in comparison with the best ψ-situations. Specifically, an "optimistic" agent i prefers φ to ψ if and only if, for every ψ-situation envisaged by i, there exists a φ-situation envisaged by i such that the latter is at least as desirable as the former. According to the pessimistic view, she focuses on the worst φ-situations in comparison with the worst ψ-situations. Specifically, a "pessimistic" agent i prefers φ to ψ if and only if, for every φ-situation envisaged by i, there exists a ψ-situation envisaged by i such that the former is at least as desirable as the latter.

Let us first define a dyadic operator for preference according to the optimistic view.

Definition 9 (*Preference: optimistic view*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that, according to agent i's optimistic assessment, φ is at least as good as ψ at w, noted $M, w \models \mathsf{P}_i^{Opt}(\psi \preceq \varphi)$, if and only if $\forall u \in ||\psi||_{i,w,M}, \exists v \in ||\varphi||_{i,w,M} : u \preceq_{i,D} v$.

As the following proposition highlights, it is expressible in the language $\mathcal{L}_{\mathrm{DLCA}}$.

Proposition 7
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{P}_i^{Opt}(\psi \preceq \varphi) \text{ iff } M, w \models \left[\equiv_i ; \psi? \right] \langle \preceq_{i,D} \rangle \varphi.$$

Let us now define preference according to the pessimistic view.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

Definition 10 (*Preference: pessimistic view*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that, according to agent i's pessimistic assessment, φ is at least as good as ψ at w, noted $M, w \models \mathsf{P}_i^{Pess}(\psi \preceq \varphi)$, if and only if $\forall v \in ||\varphi||_{i,w,M}, \exists u \in ||\psi||_{i,w,M} : u \preceq_{i,D} v$.

As for the optimistic view, the pessimistic view is also expressible in the language $\mathcal{L}_{\mathrm{DLCA}}$.

Proposition 8
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{P}_i^{Pess}(\psi \preceq \varphi) \text{ iff } M, w \models \big[\equiv_i; \varphi?\big]\langle \succeq_{i,D}\rangle \psi.$$

Thanks to the totality of the relation $\preceq_{i,D}$ (Constraint C2 in Definition 1), dyadic preference over formulas is total too. This fact is illustrated by the following validity. For every $x \in \{Opt, Pess\}$:

$$\models_{MCM} \mathsf{P}_i^x(\psi \preceq \varphi) \vee \mathsf{P}_i^x(\varphi \preceq \psi). \tag{4}$$

To see this suppose $M, w \models \neg \mathsf{P}_i^{Opt}(\psi \preceq \varphi)$ for an arbitrary model M and world w in M. Because of Constraint C2 in Definition 1, the latter implies that $\exists u \in ||\psi||_{i,w,M}, \forall v \in ||\varphi||_{i,w,M} : v \prec_{i,D} u$. Therefore, $\forall v \in ||\varphi||_{i,w,M}, \exists u \in ||\psi||_{i,w,M} : v \preceq_{i,D} u$ which is equivalent to $M, w \models \mathsf{P}_i^{Opt}(\varphi \preceq \psi)$. The case $x = Pess$ can be proved in an analogous way.

The previous notion of (optimistic and pessimistic) preference does not depend on what the agent believes. This means that, in order to assess whether φ is at least as good as ψ, an agent also takes into account worlds that are implausible (or, more generally, not maximally plausible). *Realistic* preference requires that an agent compares two formulas φ and ψ only with respect to the set of most plausible states. This idea has been discussed in the area of qualitative decision theory by different authors (Boutilier 1994; Brafman and Tennenholtz 2000; 1996).

The following definition introduces *realistic* preference according to the optimistic view.

Definition 11 (*Realistic preference: optimistic view*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that, according to agent i's optimistic assessment, φ is realistically at least as good as ψ at w, noted $M, w \models \mathsf{RP}_i^{Opt}(\psi \preceq \varphi)$, if and only if $\forall u \in Best_{i,P}(w) \cap ||\psi||_{i,w,M}, \exists v \in Best_{i,P}(w) \cap ||\varphi||_{i,w,M} : u \preceq_{i,D} v$.

The idea is that an "optimistic" agent i considers φ *realistically* at least as good as ψ if and only if, for every ψ-situation in agent i's belief set, there exists a φ-situation in agent i's belief set such that the latter is at least as good as the former.

The previous notion as well is expressible in the language $\mathcal{L}_{\mathrm{DLCA}}$.

Proposition 9
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{RP}_i^{Opt}(\psi \preceq \varphi) \text{ iff } M, w \models \big[\equiv_i; [\prec_{i,P}]\bot?; \psi?\big]\langle \preceq_{i,D} \cap (\equiv_i; [\prec_{i,P}]\bot?)\rangle \varphi.$$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

The following definition introduces *realistic* preference according to the pessimistic view.

Definition 12 (*Realistic preference: pessimistic view*)
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. We say that, according to agent i's pessimistic assessment, φ is realistically at least as good as ψ at w, noted $M, w \models \mathsf{RP}_i^{Pess}(\psi \preceq \varphi)$, if and only if $\forall v \in Best_{i,P}(w) \cap ||\varphi||_{i,w,M}, \exists u \in Best_{i,P}(w) \cap ||\psi||_{i,w,M} : u \preceq_{i,D} v$.

The idea is that a "pessimistic" agent i considers φ *realistically* at least as good as ψ if and only if, for every φ-situation in agent i's belief set, there exists a ψ-situation in agent i's belief set such that the former is at least as good as the latter.

It is also expressible in the language $\mathcal{L}_{\mathrm{DLCA}}$.

Proposition 10
Let $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and let $w \in W$. Then, we have

$$M, w \models \mathsf{RP}_i^{Pess}(\psi \preceq \varphi) \text{ iff } M, w \models \big[\equiv_i; [\prec_{i,P}]\bot?; \varphi?\big]\langle \succeq_{i,D} \cap (\equiv_i; [\prec_{i,P}]\bot?)\rangle\psi.$$

Like dyadic preference over formulas, realistic dyadic preference over formulas is total. In fact, for every for $x \in \{Opt, Pess\}$, we have:

$$\models_{MCM} \mathsf{RP}_i^x(\psi \preceq \varphi) \vee \mathsf{RP}_i^x(\varphi \preceq \psi). \tag{5}$$

The following abbreviations define *strict* variants of dyadic preference operators:

$$\mathsf{P}_i^{Opt}(\psi \prec \varphi) =_{def} \neg \mathsf{P}_i^{Opt}(\varphi \preceq \psi)$$
$$\mathsf{P}_i^{Pess}(\psi \prec \varphi) =_{def} \neg \mathsf{P}_i^{Pess}(\varphi \preceq \psi)$$
$$\mathsf{RP}_i^{Opt}(\psi \prec \varphi) =_{def} \neg \mathsf{RP}_i^{Opt}(\varphi \preceq \psi)$$
$$\mathsf{RP}_i^{Pess}(\psi \prec \varphi) =_{def} \neg \mathsf{RP}_i^{Pess}(\varphi \preceq \psi)$$

$\mathsf{P}_i^{Opt}(\psi \prec \varphi)$ (resp. $\mathsf{P}_i^{Pess}(\psi \prec \varphi)$) has to be read "according to i's optimistic (resp. pessimistic) assessment, φ is better than ψ". $\mathsf{RP}_i^{Opt}(\psi \prec \varphi)$ (resp. $\mathsf{RP}_i^{Pess}(\psi \prec \varphi)$) has to be read "according to agent i's optimistic (resp. pessimistic) assessment, φ is realistically better than ψ".

We conclude this section by defining two notions of monadic preference and corresponding two notions of *realistic* monadic preference, respectively, noted $\mathsf{P}_i^{Opt}\varphi$, $\mathsf{P}_i^{Pess}\varphi$, $\mathsf{RP}_i^{Opt}\varphi$, and $\mathsf{RP}_i^{Pess}\varphi$:

$$\mathsf{P}_i^{Opt}\varphi =_{def} \mathsf{P}_i^{Opt}(\neg\varphi \prec \varphi)$$
$$\mathsf{P}_i^{Pess}\varphi =_{def} \mathsf{P}_i^{Pess}(\neg\varphi \prec \varphi)$$
$$\mathsf{RP}_i^{Opt}\varphi =_{def} \mathsf{RP}_i^{Opt}(\neg\varphi \prec \varphi)$$
$$\mathsf{RP}_i^{Pess}\varphi =_{def} \mathsf{RP}_i^{Pess}(\neg\varphi \prec \varphi).$$

An optimistic (resp. pessimistic) agent has a preference for φ, noted $\mathsf{P}_i^{Opt}\varphi$ (resp. $\mathsf{P}_i^{Pess}\varphi$), if and only if, according to her optimistic (resp. pessimistic) assessment, φ is better than $\neg\varphi$. An optimistic (resp. pessimistic) agent has a realistic preference for φ, noted $\mathsf{RP}_i^{Opt}\varphi$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

(resp. $\mathsf{RP}_i^{Pess}\varphi$), if and only if, according to her optimistic (resp. pessimistic) assessment, φ is realistically better than $\neg\varphi$.

The following validity illustrates the relationship between the notion of desire defined in Definition 6 and the previous notion of pessimistic monadic preference:

$$\models_{MCM} \neg\mathsf{D}_i\top \rightarrow (\mathsf{D}_i\varphi \leftrightarrow \mathsf{P}_i^{Pess}\varphi). \tag{6}$$

This means that if there exists at least a minimally desirable state for agent i (condition $\neg\mathsf{D}_i\top$), then i desires that φ if and only if, according to her pessimistic assessment, φ is better than $\neg\varphi$.

3.4 Example

In the previous sections, we have defined a variety of cognitive attitudes of epistemic and motivational type. Let us illustrate them with the help of the crossroad scenario sketched in the introduction. For simplicitly, we assume that $Agt = \{1, 2\}$ and that the set of atomic propositions Atm includes the following elements with their corresponding meaning: co ("agent 1 and agent 2 collide"), lo_1 ("agent 1 loses time"), and lo_2 ("agent 2 loses time").

We are going to make different hypotheses about the agents' cognitive attitudes and present a number of conclusions that can be drawn from them. Our initial hypothesis concerns the agents' knowledge:

$$\varphi_1 =_{def} \bigwedge_{i\in\{1,2\}} [\equiv_i]\Big(\big((lo_1 \wedge \neg lo_2) \rightarrow \neg co\big)\wedge \big((\neg lo_1 \wedge lo_2) \rightarrow \neg co\big) \wedge \neg(\neg lo_1 \wedge \neg lo_2)\Big).$$

According to hypothesis φ_1, agents 1 and 2 know (i) that there will be no collision if one of them loses time by letting the other pass and (ii) that necessarily one of them will lose time (since if they both pass, there will be a collision so that they will both lose time).

Our second hypothesis concerns what the agents merely envisage:

$$\varphi_2 =_{def} \bigwedge_{i\in\{1,2\}} \Big(\langle\equiv_i\rangle co \wedge \langle\equiv_i\rangle(lo_1 \wedge \neg lo_2) \wedge \langle\equiv_i\rangle(\neg lo_1 \wedge lo_2) \wedge \langle\equiv_i\rangle(lo_1 \wedge lo_2 \wedge \neg co)\Big).$$

According to hypothesis φ_2, agents 1 and 2 envisage four possible situations: (i) the situations in which they collide, (ii) the two situations in which one of them loses its time, while the other does not, and (iii) the situation in which they both lose time because of a collision.

We conclude with the following hypothesis about the agents' motivations, according to which each agent strongly desires not to collide and strongly desires not to lose time:

$$\varphi_3 =_{def} \bigwedge_{i\in\{1,2\}} \big(\mathsf{SD}_i\neg lo_i \wedge \mathsf{SD}_i\neg co\big).$$

As the following validities highlight, the previous hypotheses lead to different conclusions about the agents' epistemic and motivational attitudes:

$$\models_{MCM} \varphi_1 \rightarrow \bigwedge_{i\in\{1,2\}} \Big([\equiv_i]\big(co \rightarrow (lo_1 \wedge lo_2)\big) \wedge \mathsf{B}_i(\neg lo_1, lo_2) \wedge \mathsf{B}_i(\neg lo_2, lo_1)\Big) \tag{7}$$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

$$\models_{MCM}(\varphi_2 \wedge \varphi_3) \rightarrow \bigwedge_{i\in\{1,2\}} \big(\mathsf{SD}_i(\neg lo_i \wedge \neg co) \wedge \mathsf{D}_i\neg lo_i \wedge \mathsf{D}_i\neg co\big) \tag{8}$$

$$\models_{MCM}(\varphi_1 \wedge \varphi_2 \wedge \varphi_3) \rightarrow \bigwedge_{i\in\{1,2\}} \big(\mathsf{D}_i(lo_1 \wedge \neg lo_2) \wedge \mathsf{D}_i(\neg lo_1 \wedge lo_2)\big) \tag{9}$$

$$\models_{MCM}(\varphi_2 \wedge \varphi_3) \rightarrow \bigwedge_{i\in\{1,2\}} \big(\neg\mathsf{SD}_i(lo_1 \wedge \neg lo_2) \wedge \neg\mathsf{SD}_i(\neg lo_1 \wedge lo_2)\big). \tag{10}$$

The single hypothesis φ_1 leads to the conclusion (i) that the agents know that a collision implies that they both lose time and (ii) that they believe that an agent loses time conditional on the fact that the other does not. Thanks to the set of hypotheses $\{\varphi_2, \varphi_3\}$, we can conclude (i) that each agent strongly desires not to lose time and to avoid a collision and (ii) each agent has both the desire not to lose time and the desire to avoid a collision. Finally, thanks to the set of hypotheses $\{\varphi_1, \varphi_2, \varphi_3\}$, we can conclude that each agent finds desirable the situations in which only one of them loses time by letting the other pass. As the last validity indicates, such situations are merely desirable for the agent but not strongly desirable.

4 Axiomatization

In this section, we provide a sound and complete axiomatization for the DLCA. The first step consists in precisely defining this logic which includes several axioms and rule of necessitation for the modalities $[\pi]$ as well as one non-standard rule of inference for nominals.

Definition 13 (*Logic* DLCA)
We define DLCA to be the extension of classical propositional logic given by the following axioms and rules with $\tau \in \{P, D\}$:

$$([\pi]\varphi \wedge [\pi](\varphi \rightarrow \psi)) \rightarrow [\pi]\psi \qquad (\mathbf{K}_\pi)$$

$$[\equiv_i]\varphi \rightarrow \varphi \qquad (\mathbf{T}_{\equiv_i})$$

$$[\equiv_i]\varphi \rightarrow [\equiv_i][\equiv_i]\varphi \qquad (\mathbf{4}_{\equiv_i})$$

$$\neg[\equiv_i]\varphi \rightarrow [\equiv_i]\neg[\equiv_i]\varphi \qquad (\mathbf{5}_{\equiv_i})$$

$$[\preceq_{i,\tau}]\varphi \rightarrow \varphi \qquad (\mathbf{T}_{\preceq_{i,\tau}})$$

$$[\preceq_{i,\tau}]\varphi \rightarrow [\preceq_{i,\tau}][\preceq_{i,\tau}]\varphi \qquad (\mathbf{4}_{\preceq_{i,\tau}})$$

$$[\equiv_i]\varphi \rightarrow [\preceq_{i,\tau}]\varphi \qquad (\mathbf{Inc}_{\preceq_{i,\tau},\equiv_i})$$

$$\big(\langle\equiv_i\rangle\varphi \wedge \langle\equiv_i\rangle\psi\big) \rightarrow \big(\langle\equiv_i\rangle(\varphi \wedge \langle\preceq_{i,\tau}\rangle\psi) \vee \langle\equiv_i\rangle(\psi \wedge \langle\preceq_{i,\tau}\rangle\varphi)\big) \qquad (\mathbf{Conn}_{\preceq_{i,\tau},\equiv_i})$$

$$[\pi;\pi']\varphi \leftrightarrow [\pi][\pi']\varphi \qquad (\mathbf{Red}_;)$$

$$[\pi \cup \pi']\varphi \leftrightarrow ([\pi]\varphi \wedge [\pi']\varphi) \qquad (\mathbf{Red}_\cup)$$

$$([\pi]\varphi \wedge [\pi']\psi) \rightarrow [\pi \cap \pi'](\varphi \wedge \psi) \qquad (\mathbf{Add1}_\cap)$$

$$(\langle\pi\rangle x \wedge \langle\pi'\rangle x) \rightarrow \langle\pi \cap \pi'\rangle x \qquad (\mathbf{Add2}_\cap)$$

$$\varphi \rightarrow [\pi]\langle-\pi\rangle\varphi \qquad (\mathbf{Conv1}_-)$$

$$\varphi \rightarrow [-\pi]\langle\pi\rangle\varphi \qquad (\mathbf{Conv2}_-)$$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

$$([\preceq_{i,\tau}]\varphi \wedge [\preceq^{\sim}_{i,\tau}]\varphi) \leftrightarrow [\equiv_i]\varphi \qquad (\mathbf{Comp1}_{\sim})$$

$$\langle \preceq_{i,\tau}\rangle x \rightarrow [\preceq^{\sim}_{i,\tau}]\neg x \qquad (\mathbf{Comp2}_{\sim})$$

$$[?\varphi]\psi \rightarrow (\varphi \rightarrow \psi) \qquad (\mathbf{Red}_{?})$$

$$\langle \pi\rangle(x \wedge \varphi) \rightarrow [\pi'](x \rightarrow \varphi) \qquad (\mathbf{Most}_{x})$$

$$\frac{\varphi}{[\pi]\varphi} \qquad (\mathbf{Nec}_{\pi})$$

$$\frac{[\pi]\neg x \text{ for all } x \in \mathit{Nom}}{[\pi]\bot}. \qquad (\mathbf{Cov})$$

Note that the primitive operators $[\preceq_{i,P}]$ and $[\preceq_{i,D}]$ are S4 (or KT4), while $[\equiv_i]$ is S5. The only interaction principles between these three operators are the "inclusion" Axiom $\mathbf{Inc}_{\preceq_{i,\tau},\equiv_i}$ and the "connectedness" Axiom $\mathbf{Conn}_{\preceq_{i,\tau},\equiv_i}$. Operators $[\preceq_{i,P}]$ and $[\preceq_{i,D}]$ do not interact since, as we have emphasized in Section 2, epistemic plausibility and desirability are assumed to be independent notions.

For every $\varphi \in \mathcal{L}_{\mathrm{DLCA}}$, we write $\vdash \varphi$ to denote the fact that φ is a theorem of DLCA, that is, there exists an at most countably infinite sequence $\psi_0, \psi_1, \ldots$ such that $\psi_0 = \varphi$ and for all $k \geq 0$, ψ_k is an instance of some axiom or ψ_k can be obtained from some later members of the sequence by an application of some inference rule.

The rest of this section is devoted to prove that the logic DLCA is sound and complete for the class of MCMs.

Soundness, namely checking that the axioms are valid and the rules of inferences preserve validity, is a routine exercise. Notice that the admissibility of the rule of inference **Cov** is guaranteed by the fact that the set of nominals *Nom* is infinite.

As for completeness, the proof is organized in several steps. We use techniques from dynamic logic and modal logic with names (Passy and Tinchev 1991; Gargov and Goranko 1993).

In the rest of this section, we denote sets of formulas from $\mathcal{L}_{\mathrm{DLCA}}$ by $\Sigma, \Sigma', \ldots$. Let $\varphi \in \mathcal{L}_{\mathrm{DLCA}}$ and $\Sigma \subseteq \mathcal{L}_{\mathrm{DLCA}}$, we define:

$$\Sigma + \varphi = \{\psi \in \mathcal{L}_{\mathrm{DLCA}} : \varphi \rightarrow \psi \in \Sigma\}.$$

Let us start by defining the concepts of theory and maximal consistent theory (MCT).

Definition 14 (*Theory*)
A set of formulas Σ is said to be a theory if it contains all theorems of DLCA and is closed under modus ponens and rule **Cov**. It is said to be a consistent theory if it is a theory and $\bot \notin \Sigma$. It is said to be a MCT if it is a consistent theory and, for each consistent theory Σ', we have that if $\Sigma \subseteq \Sigma'$, then $\Sigma = \Sigma'$.

We have the following property for theories.

Proposition 11
Let Σ be a theory and let $\varphi \in \mathcal{L}_{\mathrm{DLCA}}$. Then, $\Sigma + \varphi$ is a theory. Moreover, if Σ is consistent, then either $\Sigma + \varphi$ is consistent or $\Sigma + \neg\varphi$ is consistent.

The following proposition highlights some standard properties of MCTs.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

Proposition 12
Let Σ be a MCT. Then, for all $\varphi, \psi \in \mathcal{L}_{\text{DLCA}}$:

- $\varphi \in \Sigma$ or $\neg\varphi \in \Sigma$,
- $\varphi \vee \psi \in \Sigma$ iff $\varphi \in \Sigma$ or $\psi \in \Sigma$.

The following variant of the Lindenbaum's lemma is proved in the same way as Passy and Tinchev (1991, Lemma 4.15).

Lemma 1
Let Σ be a consistent theory and let $\varphi \notin \Sigma$. Then, there exists a MCT Σ^+ such that $\Sigma \subseteq \Sigma^+$ and $\varphi \notin \Sigma^+$.

The following lemma highlights a fundamental property of MCTs.

Lemma 2
Let Σ be a MCT. Then, there exists $x \in Nom$ such $x \in \Sigma$.

Let us now define the canonical model for our logic.

Definition 15 (*Canonical model*)
The canonical model is the tuple $M^c = (W^c, (\preceq^c_{i,P})_{i\in Agt}, (\preceq^c_{i,D})_{i\in Agt}, (\equiv^c_i)_{i\in Agt}, V^c)$ such that:

- W^c is the set of all MCTs,
- for all $i \in Agt$, for all $\tau \in \{P, D\}$, for all $w, v \in W^c$, $w \preceq^c_{i,\tau} v$ iff, for all $\varphi \in \mathcal{L}_{\text{DLCA}}$, if $[\preceq_{i,\tau}]\varphi \in w$, then $\varphi \in v$,
- for all $i \in Agt$, for all $w, v \in W^c$, $w \equiv^c_i v$ iff, for all $\varphi \in \mathcal{L}_{\text{DLCA}}$, if $[\equiv_i]\varphi \in w$, then $\varphi \in v$,
- for all $w \in W^c$, $V^c(w) = (Atm \cup Nom) \cap w$.

Let us now define the canonical relations for the complex programs π.

Definition 16 (*Canonical relation*)
Let $M^c = (W^c, (\preceq^c_{i,P})_{i\in Agt}, (\preceq^c_{i,D})_{i\in Agt}, (\equiv^c_i)_{i\in Agt}, V^c)$ be the canonical model. Then, for all $\pi \in \mathcal{P}$ and for all $w, v \in W^c$:

$$wR^c_\pi v \text{ iff, for all } \varphi \in \mathcal{L}_{\text{DLCA}}, \text{ if } [\pi]\varphi \in w \text{ then } \varphi \in v.$$

The following Lemma 3 highlights one fundamental property of the canonical model.

Lemma 3
Let $M^c = (W^c, (\preceq^c_{i,P})_{i\in Agt}, (\preceq^c_{i,D})_{i\in Agt}, (\equiv^c_i)_{i\in Agt}, V^c)$ be the canonical model. Then, for all $\Sigma, \Sigma' \in W^c$, for all $\pi \in \mathcal{P}$, and for all $x \in Nom$, if $x \in \Sigma, x \in \Sigma'$ and $\Sigma R^c_\pi \Sigma'$, then $\Sigma = \Sigma'$.

The next step consists in proving the following existence lemma.

Lemma 4
Let $M^c = (W^c, (\preceq^c_{i,P})_{i\in Agt}, (\preceq^c_{i,D})_{i\in Agt}, (\equiv^c_i)_{i\in Agt}, V^c)$ be the canonical model, let $w \in W^c$, and let $\langle\pi\rangle\varphi \in \mathcal{L}_{\text{DLCA}}$. Then, if $\langle\pi\rangle\varphi \in w$, then there exists $v \in W^c$ such that $wR^c_\pi v$ and $\varphi \in v$.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

The following truth lemma is proved in the usual way by induction on the structure of φ thanks to Lemma 4.

Lemma 5
Let $M^c = (W^c, (\preceq^c_{i,P})_{i\in Agt}, (\preceq^c_{i,D})_{i\in Agt}, (\equiv^c_i)_{i\in Agt}, V^c)$ be the canonical model, let $w \in W^c$, and let $\varphi \in \mathcal{L}_{\mathrm{DLCA}}$. Then, $M^c, w \models \varphi$ iff $\varphi \in w$.

The pre-final stage of the proof consists in introducing an alternative semantics for the language $\mathcal{L}_{\mathrm{DLCA}}$ which turns out to be equivalent to the original semantics based on MCMs.

Definition 17 (*Quasi multi-agent cognitive model*)
A quasi-MCM is a tuple $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ where W, $\preceq_{i,P}$, $\preceq_{i,D}$, $\equiv_i$, and V are as in Definition 1 except that Constraint C4 is replaced by the following weaker constraint. For all $w, v \in W$:

(C4*) if $V_{Nom}(w) \cap V_{Nom}(v) \neq \emptyset$ and $wR_\pi v$ for some $\pi \in \mathcal{P}$ then $w = v$.

By the generated submodel property, it is easy to show that the semantics in terms of MCMs and the semantics in terms of quasi-MCMs are equivalent with respect to the language $\mathcal{L}_{\mathrm{DLCA}}$.

Proposition 13
Let $\varphi \in \mathcal{L}_{\mathrm{DLCA}}$. Then, φ is valid relative to the class of MCMs if and only if φ is valid relative to the class of quasi-MCMs.

The following theorem highlights that the canonical model is indeed a structure of the right type.

Lemma 6
The canonical model M^c is a quasi-MCM.

Let us conclude the proof by supposing $\nvdash \neg\varphi$. Therefore, by Lemma 1 and the fact that the set of DLCA theorems is a consistent theory, there exists a MCT w such that $\neg\varphi \notin w$. Thus, by Proposition 12, we can find a MCT w such that $\varphi \in w$. By Lemma 5, the latter implies $M^c, w \models \varphi$ for some $w \in W^c$. Since, by Lemma 6, M^c is a quasi-MCM, it follows that φ is satisfiable relative to the class of quasi-MCMs. Therefore, by Proposition 13, φ is satisfiable relative to the class of MCMs.

We can finally state the main result of this section.

Theorem 1
The logic DLCA is sound and complete for the class of MCMs.

5 Application to game theory

In this section, we apply our logical framework to the analysis of single-stage games under incomplete information in which agents only play once (i.e. interaction is non-repeated) and may not know some relevant characteristic of others including their preferences, choices, and beliefs.

Let Act be a set of action names with elements noted $a, b, \ldots$ Let a joint action be a function $\delta : Agt \longrightarrow Act$ and the set of joint actions be denoted by $JAct$.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

For every coalition $C \in 2^{Agt}$ and for every $\delta \in JAct$, let δ_C be the C-restriction of δ, that is, the function $\delta_C : C \longrightarrow Act$ such that $\delta_C(i) = \delta(i)$ for all $i \in C$. For notational convenience, we write $-i$ instead of $Agt \setminus \{i\}$, with $i \in Agt$.

In order to model strategic interaction in our setting, we extend MCMs of Definition 1 by agents' choices. We call MCM with choices the resulting models.

Definition 18 (*Multi-agent cognitive model with choices*)
A multi-agent cognitive model with choices (MCMC) is a tuple $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, (C_i)_{i\in Agt}, V)$, where $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ is a MCM and every C_i is a choice function $C_i : W \longrightarrow Act$, which satisfies the following constraint, for each $i \in Agt$ and $\delta \in JAct$:

(C5) if $\forall j \in Agt$, $\exists w_j \in W$ such that $w \equiv_i w_j$ and $C_j(w_j) = \delta(j)$, then $\exists v \in W$ such that $w \equiv_i v$ and, $\forall j \in Agt$, $C_j(v) = \delta(j)$.

For every $w \in W$, $C_i(w)$ denotes agent i's actual choice at w. If $w \equiv_i v$ and $C_j(v) = a$, then a is a potential choice of agent j from agent i's perspective.

According to Constraint C5, agents' choices are subjectively independent, in the sense that every agent i knows that an agent cannot be deprived of her choices due to the choices made by the others. In other words, suppose that, from agent i's perspective, $\delta(j)$ is a potential choice of j for every agent j. Then, from agent i's perspective, there should be a state at which the agents choose the joint action δ. It is a subjective version of the property of choice independence formulated in the "seeing to it that" (STIT) framework (Belnap *et al.* 2001; Lorini 2013; Balbiani *et al.* 2008).

At the syntactic level, we extend the language $\mathcal{L}_{\mathrm{DLCA}}$ by special constants for choices of type $\mathsf{play}(i, a)$, with $i \in Agt$ and $a \in Act$, denoting the fact that "agent i plays (or chooses) action a". The resulting language is noted $\mathcal{L}_{\mathrm{DLCAG}}$, where DLCAG stands for "Dynamic Logic of Cognitive Attitudes in Games" and a constant $\mathsf{play}(i, a)$ is interpreted relative to a MCMC M and a world w in M, as follows:

$$M, w \models \mathsf{play}(i, a) \iff C_i(w) = a.$$

Let $\delta \in JAct$ and $C \in 2^{Agt}$. We define:

$$\mathsf{play}(\delta_C) =_{def} \bigwedge_{i\in C} \mathsf{play}\big(i, \delta_C(i)\big).$$

For every formula φ in $\mathcal{L}_{\mathrm{DLCAG}}$, we say that φ is valid, noted $\models_{MCMC} \varphi$, if and only if for every MCM with choices M and world w in M, we have $M, w \models \varphi$.

Definition 19 (*Logic* DLCAG)
We define DLCAG to be the extension of logic DLCA given by the following axioms:

$$\mathsf{play}(i, a) \rightarrow \neg\mathsf{play}(i, b) \text{ if } a \neq b \qquad (\mathbf{MostAct})$$

$$\bigvee_{a\in Act} \mathsf{play}(i, a) \qquad (\mathbf{LeastAct})$$

$$\Big(\bigwedge_{j\in Agt} \langle\equiv_i\rangle \mathsf{play}\big(j, \delta(j)\big) \Big) \rightarrow \langle\equiv_i\rangle \mathsf{play}\big(\delta_{Agt}\big). \qquad (\mathbf{SIC})$$

Axiom **MostAct** means that an agent chooses at most one action from Act, while, according to Axiom **LeastAct**, an agent chooses at least one action from Act. Axiom

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

SIC is the syntactic counterpart of *subjective* choice independence expressed by Constraint C5.

We can adapt the techniques used for proving Theorem 1 in order to prove the following Theorem 2.

Theorem 2
The logic DLCAG is sound and complete for the class of MCMs with choices.

With the support of the language $\mathcal{L}_{\text{DLCAG}}$, we can define a variety of notions from the theory of games under incomplete information. The first notion we consider is best response, both from the perspective of an optimistic agent and from the perspective of a pessimistic one:

$$\mathsf{BR}_i^{Opt}(a,\delta_{-i}) =_{def} \bigwedge_{b\in Act} \mathsf{RP}_i^{Opt}\Big(\big(\mathsf{play}(i,b)\wedge\mathsf{play}(\delta_{-i})\big)\preceq\big(\mathsf{play}(i,a)\wedge\mathsf{play}(\delta_{-i})\big)\Big),$$

$$\mathsf{BR}_i^{Pess}(a,\delta_{-i}) =_{def} \bigwedge_{b\in Act} \mathsf{RP}_i^{Pess}\Big(\big(\mathsf{play}(i,b)\wedge\mathsf{play}(\delta_{-i})\big)\preceq\big(\mathsf{play}(i,a)\wedge\mathsf{play}(\delta_{-i})\big)\Big).$$

We say that playing action a is for agent i an optimistic (resp. pessimistic) best response to the others' joint action δ_{-i}, noted $\mathsf{BR}_i^{Opt}(a,\delta_{-i})$ (resp. $\mathsf{BR}_i^{Pess}(a,\delta_{-i})$) if and only if for every action b, according to agent i's optimistic (resp. pessimistic) assessment, playing a, while the others play δ_{-i} is realistically at least as good as playing b, while the others play δ_{-i}.

As for best response, we can define two types of *subjective* Nash equilibrium, one for optimistic agents and the other for pessimistic ones. Our notion of subjective Nash equilibrium corresponds to a *qualitative* variant of the notion of Bayesian Nash equilibrium (BNE): a similar qualitative variant of BNE is studied by Amor *et al.* (2019) in the context of possibility theory. The joint action δ is said to be a subjective optimistic (resp. pessimistic) Nash equilibrium, noted $\mathsf{NE}^{Opt}(\delta)$ (resp. $\mathsf{NE}^{Pess}(\delta)$), if no agent i wants to unilaterally deviate from the chosen strategy $\delta(i)$, under that the assumption that i is optimistic (resp. pessimistic):

$$\mathsf{NE}^{Opt}(\delta) =_{def} \bigwedge_{i\in Agt} \mathsf{BR}_i^{Opt}\big(\delta(i),\delta_{-i}\big),$$

$$\mathsf{NE}^{Pess}(\delta) =_{def} \bigwedge_{i\in Agt} \mathsf{BR}_i^{Pess}\big(\delta(i),\delta_{-i}\big).$$

Note that assuming the finiteness of the set of agents Agt is essential for defining Nash equilibrium, since our language is finitary and does not allow universal quantification over infinite sets.

Given the distinction between optimistic and pessimistic agents, two notions of rationality can be defined. Agent i is said to be optimistic (resp. pessimistic) rational, noted Rat_i^{Opt} (resp. Rat_i^{Pess}), if she cannot choose an action that, according to her optimistic (resp. pessimistic) assessment, is better not to choose than to choose:

$$\mathsf{Rat}_i^{Opt} =_{def} \bigwedge_{a\in Act} \Big(\mathsf{play}(i,a)\rightarrow\mathsf{RP}_i^{Opt}\big(\neg\mathsf{play}(i,a)\preceq\mathsf{play}(i,a)\big)\Big),$$

$$\mathsf{Rat}_i^{Pess} =_{def} \bigwedge_{a\in Act} \Big(\mathsf{play}(i,a)\rightarrow\mathsf{RP}_i^{Pess}\big(\neg\mathsf{play}(i,a)\preceq\mathsf{play}(i,a)\big)\Big).$$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

As the following proposition indicates, the action chosen by an optimistic (resp. pessimistic) rational agent is, according to the agent's optimistic (resp. pessimistic) assessment, at least as good as the other actions she may choose.

Proposition 14
Let $i \in Agt$ and $x \in \{Opt, Pess\}$. Then,

$$\models_{MCMC} \big(\mathsf{Rat}_i^x \wedge \mathsf{play}(i,a)\big) \rightarrow \bigwedge_{b \in Act} \mathsf{RP}_i^x\big(\mathsf{play}(i,b) \preceq \mathsf{play}(i,a)\big). \tag{11}$$

The following proposition elucidates the connection between the notions of belief, rationality, and Nash equilibrium: if all agents are optimistic (resp. pessimistic) rational and have a correct belief about the others' actual choices, then the joint action they choose is a subjective optimistic (resp. pessimistic) Nash equilibrium.[7]

Proposition 15
Let $x \in \{opt, pess\}$ and $\delta \in JAct$. Then:

$$\models_{MCMC} \Big(\mathsf{play}(\delta) \wedge \bigwedge_{i \in Agt} \big(\mathsf{Rat}_i^x \wedge \mathsf{B}_i \mathsf{play}(\delta_{-i})\big)\Big) \rightarrow \mathsf{NE}_i^x(\delta). \tag{12}$$

We conclude this section by illustrating the game-theoretic concepts involved in the crossroad game described in Section 3.4. It is a game under incomplete information since an agent does not necessarily know the other agent's beliefs and desires. It is single stage since that interaction is non-repeated and agents are supposed to choose simultaneously.

Example (cont.) *Let us suppose that the set of actions that agents* 1 *and* 2 *can choose is* $Act = \{C, S\}$*, where* C *is the action "to continue" and* S *is the action "to stop". The following hypotheses capture the agents' knowledge and beliefs about actions and their effects:*

$$\begin{aligned}
\varphi_4 =_{def} \bigwedge_{i \in \{1,2\}} [\equiv_i]\Big(&\big((\mathsf{play}(1,C) \wedge \mathsf{play}(2,C)) \rightarrow co\big) \wedge \\
&\big((\mathsf{play}(1,C) \wedge \mathsf{play}(2,S)) \rightarrow (\neg lo_1 \wedge lo_2)\big) \wedge \\
&\big((\mathsf{play}(1,S) \wedge \mathsf{play}(2,C)) \rightarrow (lo_1 \wedge \neg lo_2)\big) \wedge \\
&\big((\mathsf{play}(1,S) \wedge \mathsf{play}(2,S)) \rightarrow (lo_1 \wedge lo_2 \wedge \neg co)\big)\Big), \\
\varphi_5 =_{def} \Big(&\big(\widehat{\mathsf{B}}_1 \mathsf{play}(1 \mapsto C, 2 \mapsto C) \leftrightarrow \widehat{\mathsf{B}}_1 \mathsf{play}(1 \mapsto S, 2 \mapsto C)\big) \wedge \\
&\big(\widehat{\mathsf{B}}_1 \mathsf{play}(1 \mapsto C, 2 \mapsto S) \leftrightarrow \widehat{\mathsf{B}}_1 \mathsf{play}(1 \mapsto S, 2 \mapsto S)\big) \wedge \\
&\big(\widehat{\mathsf{B}}_2 \mathsf{play}(1 \mapsto C, 2 \mapsto C) \leftrightarrow \widehat{\mathsf{B}}_2 \mathsf{play}(1 \mapsto C, 2 \mapsto S)\big) \wedge \\
&\big(\widehat{\mathsf{B}}_2 \mathsf{play}(1 \mapsto S, 2 \mapsto C) \leftrightarrow \widehat{\mathsf{B}}_2 \mathsf{play}(1 \mapsto S, 2 \mapsto S)\big)\Big),
\end{aligned}$$

where $\widehat{\mathsf{B}}_i \varphi =_{def} \neg \mathsf{B}_i \neg \varphi$*. According to the hypothesis* φ_4*, the agents know that (i) if they both continue, they will collide; (ii) if one of them continues while the other stops, then*

[7] A similar epistemic characterization of Nash equilibrium is provided by Aumann & Brandenburger (A&B) (Aumann and Brandenburger 1995) in the context of games with complete information. See also Spohn (1982) for a similar result using a probabilistic approach.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

the first will lose its time while the second will not; and (iii) if they both stop, each of them will lose its time but there will be no collision. According to the hypothesis φ_5*, the fact that an agent considers possible that the other will decide to continue (resp. to stop) does not depend on the agent's choice. This hypothesis is justified by the assumption that an agent's beliefs are* ex ante*, that is, relative to the instant before an agent makes its choice.*

As the following validity indicates, the previous hypotheses φ_4 *and* φ_5 *together with the hypotheses* φ_1 *and* φ_3 *stated in Section 3.4 lead to the conclusion that (i) an agent's action of continuing is both an optimistic and a pessimistic best response to the other agent's action of stopping, and an agent's action of stopping is both an optimistic and a pessimistic best response to the other agent's action of continuing. For every* $x \in \{opt, pess\}$*, we have:*

$$\models_{MCMC}(\varphi_1 \wedge \varphi_3 \wedge \varphi_4 \wedge \varphi_5) \rightarrow \big(\mathsf{BR}_1^x(S, 2 \mapsto C) \wedge \mathsf{BR}_1^x(C, 2 \mapsto S) \wedge \mathsf{BR}_2^x(S, 1 \mapsto C) \wedge \mathsf{BR}_2^x(C, 1 \mapsto S)\big). \quad (13)$$

6 Dynamic extension

The logics we presented so far merely provide a static picture of the cognitive attitudes and choices of multiple agents in interactive situations. Following the tradition of DEL (van Ditmarsch *et al*. 2007), in this section, we move from a static to a dynamic perspective and extend the language $\mathcal{L}_{\text{DLCA}}$ by a variety of dynamic operators for cognitive attitude change. We consider two types of cognitive attitude change, namely, radical attitude and conservative attitude change. Radical attitude change, both in its epistemic and in its motivational form, satisfies a strong form of success postulate. Particularly, if an agent forms the belief that φ, as a consequence of a radical belief revision by φ, then she should also form the strong belief that φ. Analogously, if an agent forms the desire that φ, as a consequence of a radical desire revision by φ, then she should also form the strong desire that φ. On the contrary, after a conservative belief (resp. desire) revision by φ is performed, an agent may form the belief (resp. desire) that φ without forming the strong belief (resp. strong desire) that φ. While radical and conservative belief revision have been studied before in the literature on DEL (van Benthem 2007; Baltag and Smets 2008), we are the first to apply DEL techniques to the analysis of desire revision and to oppose belief revision to desire revision in the DEL setting.[8]

In the rest of this section, we first define the semantics of radical belief revision and desire revision operators (Section 6.1). Then, we turn to conservative attitude change and define the semantics of conservative belief revision and desire revision operators (Section 6.2). Finally, we provide an axiomatics for the dynamic extension of our logic DLCA (Section 6.3).

6.1 Radical attitude revision

Radical attitude revision operators are of the form $[\Uparrow_{i,\tau} \varphi]$, with $\tau \in \{P, D\}$. They describe the consequences of a radical revision operation. In particular, the formula $[\Uparrow_{i,P}$

[8] Research in the DEL area has rather concentrated on preference change (van Benthem and Liu 2007; van Eijck 2008), leaving desire change unexplored.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

$\varphi]\psi$ is meant to stand for "ψ holds, after agent i has radically revised her beliefs with φ", whereas $[\Uparrow_{i,D}\,\varphi]\psi$ is meant to stand for "ψ holds, after agent i has radically revised her desires with φ". We assume that radical revision operations are public, that is, if an agent radically revises her beliefs (resp. desires) with φ, then this is common knowledge among all agents. This assumption could be easily relaxed by using action models as introduced in Baltag *et al.* (1998), Baltag and Moss (2004), which would allow us to model private and semi-private attitude change operations. Radical revision operators are interpreted relative to a MCM $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ and a world w in W, as follows:

$$M, w \models [\Uparrow_{i,\tau}\,\varphi]\psi \iff M^{\Uparrow_{i,\tau}\varphi}, w \models \psi,$$

where

$$M^{\Uparrow_{i,P}\varphi} = (W, (\preceq_{i,P}^{\Uparrow_{i,P}\varphi})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V),$$

$$M^{\Uparrow_{i,D}\varphi} = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D}^{\Uparrow_{i,D}\varphi})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V),$$

$$\preceq_{i,\tau}^{\Uparrow_{i,\tau}\varphi} = \Big\{(w,v) \in W\times W : \big((M,w\models\varphi \text{ iff } M,v\models\varphi) \text{ and } w \preceq_{i,\tau} v\big) \text{ or } (M,w\models\neg\varphi,\, M,v\models\varphi \text{ and } w\equiv_i v)\Big\},$$

and $\preceq_{j,\tau}^{\Uparrow_{i,\tau}\varphi} = \preceq_{j,\tau}$ for all $j \in Agt$ such that $i \neq j$.

Radical belief and desire revision are completely symmetric from the point of view of the plausibility and desirability ordering. Agent i's radical belief revision with φ transforms agent i's plausibility ordering $\preceq_{i,P}$ into the new plausibility ordering $\preceq_{i,P}^{\Uparrow_{i,P}\varphi}$. In particular, it makes all φ-worlds in i's information set more plausible than all $\neg\varphi$-worlds and, within those two zones, it keeps the old plausibility ordering. Analogously, agent i's radical desire revision with φ transforms agent i's desirability ordering $\preceq_{i,D}$ into the new desirability ordering $\preceq_{i,D}^{\Uparrow_{i,D}\varphi}$. It makes all φ-worlds in i's information set more desirable than all $\neg\varphi$-worlds and, within those two zones, it keeps the old desirability ordering.

As emphasized above, radical revision satisfies a strong form of success principle which is formally expressed by the following two validities. Let $\varphi \in \mathcal{L}_{\mathsf{PL}}(Atm)$. Then,

$$\models_{MCM} \langle\equiv_i\rangle\varphi \to [\Uparrow_{i,P}\,\varphi](\mathsf{B}_i\varphi \wedge \mathsf{SB}_i\varphi) \tag{14}$$

$$\models_{MCM} \langle\equiv_i\rangle\neg\varphi \to [\Uparrow_{i,D}\,\varphi](\mathsf{D}_i\varphi \wedge \mathsf{SD}_i\varphi). \tag{15}$$

This means that (i) if φ is compatible with an agent's knowledge, then after she has radically revised her beliefs with φ, the agent will both believe that φ and strongly believe that φ, and (ii) if $\neg\varphi$ is compatible with an agent's knowledge, then after she has radically revised her desires with φ, the agent will both desire that φ and strongly desire that φ.

The two validities highlight that belief and desire behave in a slightly different way under radical revision, despite the fact that the plausibility and desirability ordering are modified in the same way.

We have the following additional validities, for $\varphi \in \mathcal{L}_{\mathsf{PL}}(Atm)$:

$$\models_{MCM} [\Uparrow_{i,P}\,\varphi](\mathsf{B}_i\varphi \to \mathsf{SB}_i\varphi) \tag{16}$$

$$\models_{MCM} [\Uparrow_{i,D}\,\varphi](\mathsf{D}_i\varphi \to \mathsf{SD}_i\varphi). \tag{17}$$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

This means that the formation of a belief (resp. desire) through radical belief (resp. desire) revision necessarily entails the formation of a strong belief (resp. strong desire) with the same content.

Example (cont.) *Let us go back to the crossroad game in order to illustrate the radical desire revision mechanism. Suppose agent* 1 *performs a radical desire revision operation with* $\neg lo_2$*, since it learns that agent* 2 *is an ambulance which has to lose no time at the crossroad. By the previous validity (15), we can prove that, under the hypothesis* φ_2 *stated in Section 3.4,* 1 *will both desire and strongly desire that* $\neg lo_2$*, after the radical desire revision operation with* $\neg lo_2$*:*

$$\models_{MCM} \varphi_2 \rightarrow [\Uparrow_{1,D} \neg lo_2](\mathsf{D}_1 \neg lo_2 \wedge \mathsf{SD}_1 \neg lo_2). \tag{18}$$

Moreover, under the set of hypotheses $\{\varphi_1, \varphi_2, \varphi_3\}$*, after the radical desire revision operation with* $\neg lo_2$*,* 1 *will not strongly desire anymore not to lose time, but it will merely desire it:*

$$\models_{MCM} (\varphi_1 \wedge \varphi_2 \wedge \varphi_3) \rightarrow [\Uparrow_{1,D} \neg lo_2](\mathsf{D}_1 \neg lo_1 \wedge \neg \mathsf{SD}_1 \neg lo_1). \tag{19}$$

As the following proposition indicates, we have reduction axioms which allow us to eliminate radical attitude revision operators from a formula.

Proposition 16
The following equivalences are valid:

$$[\Uparrow_{i,\tau} \varphi]p \leftrightarrow p$$
$$[\Uparrow_{i,\tau} \varphi]x \leftrightarrow x$$
$$[\Uparrow_{i,\tau} \varphi]\neg\psi \leftrightarrow \neg[\Uparrow_{i,\tau} \varphi]\psi$$
$$[\Uparrow_{i,\tau} \varphi](\psi_1 \wedge \psi_2) \leftrightarrow ([\Uparrow_{i,\tau} \varphi]\psi_1 \wedge [\Uparrow_{i,\tau} \varphi]\psi_2)$$
$$[\Uparrow_{i,\tau} \varphi][\pi]\psi \leftrightarrow [F^{\Uparrow_{i,\tau}\varphi}(\pi)][\Uparrow_{i,\tau} \varphi]\psi,$$

where for all $j \in Agt$ and for all $\tau, \tau' \in \{P, D\}$:

$$F^{\Uparrow_{i,\tau}\varphi}(\equiv_j) = \equiv_j$$
$$F^{\Uparrow_{i,\tau}\varphi}(\preceq_{i,\tau}) = (\varphi?; \preceq_{i,\tau}; \varphi?) \cup (\neg\varphi?; \preceq_{i,\tau}; \neg\varphi?) \cup (\neg\varphi?; \equiv_i; \varphi?)$$
$$F^{\Uparrow_{i,\tau}\varphi}(\preceq_{j,\tau'}) = \preceq_{j,\tau'} \text{ if } i \neq j \text{ or } \tau \neq \tau'$$
$$F^{\Uparrow_{i,\tau}\varphi}(\preceq^{\sim}_{i,\tau}) = (\varphi?; \preceq^{\sim}_{i,\tau}; \varphi?) \cup (\neg\varphi?; \preceq^{\sim}_{i,\tau}; \neg\varphi?) \cup (\varphi?; \equiv_i; \neg\varphi?)$$
$$F^{\Uparrow_{i,\tau}\varphi}(\preceq^{\sim}_{j,\tau'}) = \preceq^{\sim}_{j,\tau'} \text{ if } i \neq j \text{ or } \tau \neq \tau'$$
$$F^{\Uparrow_{i,\tau}\varphi}(\pi;\pi') = F^{\Uparrow_{i,\tau}\varphi}(\pi); F^{\Uparrow_{i,\tau}\varphi}(\pi')$$
$$F^{\Uparrow_{i,\tau}\varphi}(\pi \cup \pi') = F^{\Uparrow_{i,\tau}\varphi}(\pi) \cup F^{\Uparrow_{i,\tau}\varphi}(\pi')$$
$$F^{\Uparrow_{i,\tau}\varphi}(\pi \cap \pi') = F^{\Uparrow_{i,\tau}\varphi}(\pi) \cap F^{\Uparrow_{i,\tau}\varphi}(\pi')$$
$$F^{\Uparrow_{i,\tau}\varphi}(-\pi) = -F^{\Uparrow_{i,\tau}\varphi}(\pi)$$
$$F^{\Uparrow_{i,\tau}\varphi}(\psi?) = [\Uparrow_{i,\tau} \varphi]\psi?.$$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

6.2 Conservative attitude revision

Let us move from radical attitude change to conservative attitude change by introducing radical revision operators of type $[\uparrow_{i,\tau} \varphi]$, with $\tau \in \{P, D\}$. The formula $[\uparrow_{i,P} \varphi]\psi$ (resp. $[\uparrow_{i,D} \varphi]\psi$) is meant to stand for "ψ holds, after agent i has conservatively revised her beliefs (resp. desires) with φ". As for radical revision, we assume that conservative revision operations are public, that is, if an agent conservatively revises her beliefs (resp. desires) with φ, then this is common knowledge among all agents. The semantic interpretation of such operators relative to a MCM $M = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V)$ be a MCM and a world w in W is as follows:

$$M, w \models [\uparrow_{i,\tau} \varphi]\psi \iff M^{\uparrow_{i,\tau}\varphi}, w \models \psi,$$

where:

$$M^{\uparrow_{i,P}\varphi} = (W, (\preceq_{i,P}^{\uparrow_{i,P}\varphi})_{i\in Agt}, (\preceq_{i,D})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V),$$
$$M^{\uparrow_{i,D}\varphi} = (W, (\preceq_{i,P})_{i\in Agt}, (\preceq_{i,D}^{\uparrow_{i,D}\varphi})_{i\in Agt}, (\equiv_i)_{i\in Agt}, V),$$

with:

$$\preceq_{i,P}^{\uparrow_{i,P}\varphi} = \Big\{(w,v) \in W \times W : \Big(\big(w \in Best_{i,P}(\varphi, w) \text{ iff } v \in Best_{i,P}(\varphi, w)\big) \text{ and } w \preceq_{i,P} v\Big) \text{ or } \big(w \notin Best_{i,P}(\varphi, w), v \in Best_{i,P}(\varphi, w) \text{ and } w \equiv_i v\big)\Big\},$$

$$\preceq_{i,D}^{\uparrow_{i,D}\varphi} = \Big\{(w,v) \in W \times W : \Big(\big(w \in Worst_{i,D}(\neg\varphi, w) \text{ iff } v \in Worst_{i,D}(\neg\varphi, w)\big) \text{ and } w \preceq_{i,D} v\Big) \text{ or } \big(w \in Worst_{i,D}(\neg\varphi, w), v \notin Worst_{i,D}(\neg\varphi, w) \text{ and } w \equiv_i v\big)\Big\},$$

and $\preceq_{j,\tau}^{\uparrow_{i,\tau}\varphi} = \preceq_{j,\tau}$ for all $j \in Agt$ such that $i \neq j$.

Unlike radical revision, plausibility update and desirability update are asymmetric under conservative revision. Agent i's conservative belief revision with φ replaces the current plausibility ordering $\preceq_{i,P}$ with the new plausibility ordering $\preceq_{i,P}^{\uparrow_{i,P}\varphi}$. It promotes the most plausible φ-worlds to the highest plausibility rank, but apart from that, the old plausibility ordering remains. Agent i's conservative desire revision with φ replaces the current desirability ordering $\preceq_{i,D}$ with the new desirability ordering $\preceq_{i,P}^{\uparrow_{i,P}\varphi}$. In particular, it demotes the least desirable $\neg\varphi$-worlds to the lowest desirability rank, but apart from that, the old desirability ordering remains.

Conservative attitude revision satisfies a weak form of success principle which guarantees the formation of a belief (resp. a desire), after a belief (resp. desire) revision is performed. Let $\varphi\mathcal{L}_{\mathsf{PL}}(Atm)$. Then,

$$\models_{MCM} \neg \mathsf{B}_i(\varphi, \bot) \rightarrow [\uparrow_{i,P} \varphi]\mathsf{B}_i\varphi \tag{20}$$

$$\models_{MCM} \neg \mathsf{D}_i(\varphi, \top) \rightarrow [\uparrow_{i,D} \varphi]\mathsf{D}_i\varphi. \tag{21}$$

According to the previous validities, if an agent does not believe a contradiction conditional on φ, then after she has conservatively revised her beliefs with φ, she will believe that φ. If an agent does not desire a tautology conditional on φ then, after she has conservatively revised her desires with φ, she will desire that φ. But, unlike radical attitude revision, conservative attitude revision does not necessarily guarantee the formation of

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

a strong belief (resp. a strong desire), after a belief (resp. desire) revision is performed. Indeed, we have the following, for $\varphi \in \mathcal{L}_{\mathsf{PL}}(Atm)$:

$$\not\models_{MCM} [\uparrow_{i,P} \varphi](\mathsf{B}_i\varphi \rightarrow \mathsf{SB}_i\varphi) \tag{22}$$

$$\not\models_{MCM} [\uparrow_{i,D} \varphi](\mathsf{D}_i\varphi \rightarrow \mathsf{SD}_i\varphi). \tag{23}$$

This means that the formation of a belief (resp. desire) through conservative belief (resp. desire) revision does not necessarily entail the formation of a strong belief (resp. strong desire) with the same content.

Example (cont.) *Let us illustrate the conservative belief revision mechanism with the help of the crossroad game. Suppose agent* 1 *informs agent* 2 *that "if they both lose time, then there will no collision" and, as a consequence,* 2 *performs a conservative belief revision operation with input* $(lo_1 \wedge lo_2) \rightarrow \neg co$. *By the previous validity (20), we can prove that, under the hypothesis* φ_1 *stated in Section 3.4 and the assumption that* 2 *does not believe a contradiction conditional on* 1*'s assertion,* 2 *believes that there will be no collision, after its conservative belief operation:*

$$\models_{MCM} \Big(\neg \mathsf{B}_2\big((lo_1 \wedge lo_2) \rightarrow \neg co, \bot\big) \wedge \varphi_1\Big) \rightarrow [\uparrow_{2,P} (lo_1 \wedge lo_2) \rightarrow \neg co]\mathsf{B}_2 \neg co. \tag{24}$$

As for radical revision, we have reduction axioms which allow us to eliminate conservative attitude revision operators from a formula.

Proposition 17
The following equivalences are valid:

$$[\uparrow_{i,\tau} \varphi]p \leftrightarrow p$$

$$[\uparrow_{i,\tau} \varphi]x \leftrightarrow x$$

$$[\uparrow_{i,\tau} \varphi]\neg\psi \leftrightarrow \neg[\uparrow_{i,\tau} \varphi]\psi$$

$$[\uparrow_{i,\tau} \varphi](\psi_1 \wedge \psi_2) \leftrightarrow ([\uparrow_{i,\tau} \varphi]\psi_1 \wedge [\uparrow_{i,\tau} \varphi]\psi_2)$$

$$[\uparrow_{i,\tau} \varphi][\pi]\psi \leftrightarrow [F^{\uparrow_{i,\tau}\varphi}(\pi)][\uparrow_{i,\tau} \varphi]\psi,$$

where for all $j \in Agt$ and for all $\tau, \tau' \in \{P, D\}$:

$$F^{\uparrow_{i,\tau}\varphi}(\equiv_j) = \equiv_j$$

$$\begin{aligned} F^{\uparrow_{i,P}\varphi}(\preceq_{i,P}) = {} & \big((\varphi \wedge [\prec_{i,P}]\neg\varphi)?; \preceq_{i,\tau}; (\varphi \wedge [\prec_{i,P}]\neg\varphi)?\big) \cup \\ & \big(\neg(\varphi \wedge [\prec_{i,P}]\neg\varphi)?; \preceq_{i,\tau}; \neg(\varphi \wedge [\prec_{i,P}]\neg\varphi)?\big) \cup \\ & \big(\neg(\varphi \wedge [\prec_{i,P}]\neg\varphi)?; \equiv_i; (\varphi \wedge [\prec_{i,P}]\neg\varphi)?\big) \\ F^{\uparrow_{i,D}\varphi}(\preceq_{i,D}) = {} & \big((\neg\varphi \wedge [\succ_{i,D}]\varphi)?; \preceq_{i,\tau}; (\neg\varphi \wedge [\succ_{i,D}]\varphi)?\big) \cup \\ & \big(\neg(\neg\varphi \wedge [\succ_{i,D}]\varphi)?; \preceq_{i,\tau}; \neg(\neg\varphi \wedge [\succ_{i,D}]\varphi)?\big) \cup \\ & \big((\neg\varphi \wedge [\succ_{i,D}]\varphi)?; \equiv_i; \neg(\neg\varphi \wedge [\succ_{i,D}]\varphi)?\big) \\ F^{\uparrow_{i,\tau}\varphi}(\preceq_{j,\tau'}) = {} & \preceq_{j,\tau'} \text{ if } i \neq j \text{ or } \tau \neq \tau' \\ F^{\uparrow_{i,P}\varphi}(\preceq^{\sim}_{i,P}) = {} & \big((\varphi \wedge [\prec_{i,P}]\neg\varphi)?; \preceq_{i,\tau}; (\varphi \wedge [\prec_{i,P}]\neg\varphi)?\big) \cup \\ & \big(\neg(\varphi \wedge [\prec_{i,P}]\neg\varphi)?; \preceq_{i,\tau}; \neg(\varphi \wedge [\prec_{i,P}]\neg\varphi)?\big) \cup \end{aligned}$$

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

$$
\begin{aligned}
& \big((\varphi \wedge [\prec_{i,P}]\neg\varphi)?; \equiv_i; \neg(\varphi \wedge [\prec_{i,P}]\neg\varphi)?\big) \\
F^{\uparrow_{i,D}\varphi}(\preceq^{\sim}_{i,D}) = & \big((\neg\varphi \wedge [\succ_{i,D}]\varphi)?; \preceq_{i,\tau}; (\neg\varphi \wedge [\succ_{i,D}]\varphi)?\big) \cup \\
& \big(\neg(\neg\varphi \wedge [\succ_{i,D}]\varphi)?; \preceq_{i,\tau}; \neg(\neg\varphi \wedge [\succ_{i,D}]\varphi)?\big) \cup \\
& \big(\neg(\neg\varphi \wedge [\succ_{i,D}]\varphi)?; \equiv_i; (\neg\varphi \wedge [\succ_{i,D}]\varphi)?\big) \\
F^{\uparrow_{i,\tau}\varphi}(\preceq^{\sim}_{j,\tau'}) = & \preceq^{\sim}_{j,\tau'} \quad \text{if } i \neq j \text{ or } \tau \neq \tau' \\
F^{\uparrow_{i,\tau}\varphi}(\pi;\pi') = & F^{\uparrow_{i,\tau}\varphi}(\pi); F^{\uparrow_{i,\tau}\varphi}(\pi') \\
F^{\uparrow_{i,\tau}\varphi}(\pi \cup \pi') = & F^{\uparrow_{i,\tau}\varphi}(\pi) \cup F^{\uparrow_{i,\tau}\varphi}(\pi') \\
F^{\uparrow_{i,\tau}\varphi}(\pi \cap \pi') = & F^{\uparrow_{i,\tau}\varphi}(\pi) \cap F^{\uparrow_{i,\tau}\varphi}(\pi') \\
F^{\uparrow_{i,\tau}\varphi}(-\pi) = & -F^{\uparrow_{i,\tau}\varphi}(\pi) \\
F^{\uparrow_{i,\tau}\varphi}(\psi?) = & [\uparrow_{i,\tau} \varphi]\psi?.
\end{aligned}
$$

6.3 Dynamic logic of cognitive attitudes and their change

The modal language $\mathcal{L}_{\mathrm{DLCAC}}(Atm, Nom, Agt)$, or simply $\mathcal{L}_{\mathrm{DLCAC}}$, for the Dynamic Logic of Cognitive Attitudes and their Change (DLCAC) extends the language $\mathcal{L}_{\mathrm{DLCA}}$ of the logic DLCA by dynamic operators of type $[\Uparrow_{i,\tau} \varphi]$ and $[\uparrow_{i,\tau} \varphi]$. It is defined by the following grammar:

$$\varphi \quad ::= \quad p \mid x \mid \neg\varphi \mid \varphi \wedge \varphi' \mid [\pi]\varphi \mid [\Uparrow_{i,\tau} \varphi]\psi \mid [\uparrow_{i,\tau} \varphi]\psi,$$

where π ranges over the language of cognitive programs $\mathcal{P}$, p ranges over Atm, x ranges over Nom, i ranges over Agt, and τ ranges over $\{P, D\}$.

Definition 20
We define DLCAC to be the extension of DLCA given by the reduction principles of Propositions 16 and 17 and the following rule of replacement of equivalents

$$\frac{\psi_1 \leftrightarrow \psi_2}{\varphi \leftrightarrow \varphi[\psi_1/\psi_2]}, \qquad (\mathbf{REP})$$

where $\varphi[\psi_1/\psi_2]$ is the formula that results from φ by replacing zero or more occurrences of ψ_1, in φ, by ψ_2.

As the rule of replacement of equivalents preserves validity, the equivalences of Propositions 16 and 17 together with this allow to reduce every formula of the language $\mathcal{L}_{\mathrm{DLCAC}}$ to an equivalent formula of the language $\mathcal{L}_{\mathrm{DLCA}}$. Call *red* the mapping which iteratively applies the above equivalences from the left to the right, starting from one of the innermost modal operators. *red* pushes the dynamic operators inside the formula and finally eliminates them when facing an atomic formula.

Proposition 18
Let φ be a formula in the language of $\mathcal{L}_{\mathrm{DLCAC}}$. Then

- $red(\varphi)$ has no dynamic operators $[\Uparrow_{i,\tau} \varphi]$ or $[\uparrow_{i,\tau} \varphi]$, and
- $red(\varphi) \leftrightarrow \varphi$ is valid relative to the class of MCMs.

The first item of Proposition 18 is clear. The second item is proved using the equivalences of Propositions 16 and 17 and the rule of replacement of equivalents.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

The following theorem is a direct consequence of Theorem 1 and Proposition 18.

Theorem 3
The logic DLCAC is sound and complete for the class of MCMs.

7 Conclusion and perspectives

We have presented a logical framework for modeling a rich variety of cognitive attitudes of both epistemic type and motivational type. We have presented two extensions of the basic setting, one by the notion of choice and the other by dynamic operators for belief change and desire change. We have applied the former to the analysis of games under incomplete information. We have provided sound and complete axiomatizations for the basic setting and for its two extensions. Directions of future research are manifold and are briefly discussed in the rest of this section.

Decidability and complexity The present paper is devoted to study the proof-theoretic aspects of the proposed logics. In future work, we plan to investigate their computational aspects including decidability of their satisfiability problems and, at a later stage, complexity. In order to prove decidability, we expect to be able to use existing filtration techniques from modal logic. Note that once we have proved decidability of the static setting DLCA, we can use the reduction axioms of Propositions 16 and 17 to prove decidability of the dynamic setting DLCAC.

We plan to study complexity of the satisfiability problems for interesting fragments of the language $\mathcal{L}_{\text{DLCA}}$ by reducing them to satisfiability problems of existing logics. For instance, consider the following single-agent (sa) fragment of the language $\mathcal{L}_{\text{DLCA}}$ where only atomic programs (ap) are allowed, noted $\mathcal{L}^{sa,ap}_{\text{DLCA}}$:

$$\varphi \quad ::= \quad p \mid x \mid \neg\varphi \mid \varphi \wedge \varphi' \mid [\preceq_{1,P}]\varphi \mid [\preceq_{1,D}]\varphi \mid [\equiv_1]\varphi,$$

where 1 is an arbitrary agent in Agt. We can observe that the satisfiability problem for this fragment is EXPTIME-hard. Indeed, because of Constraint C1 in Definition 1, the modality $[\equiv_1]$ plays the role of the universal modality with respect to the modalities $[\preceq_{1,P}]$ and $[\preceq_{1,D}]$. As shown in Hemaspaandra (1996), adding the universal modality to a multimodal logic with independent modalities, such as the S4-modalities $[\preceq_{1,P}]$ and $[\preceq_{1,D}]$, causes EXPTIME-hardness.

Consider now the following intersection-free (if) and complement-free (cf) fragment of $\mathcal{L}_{\text{DLCA}}$, noted $\mathcal{L}^{if,cf}_{\text{DLCA}}$:

$$\begin{array}{lll} \pi & ::= & \equiv_i | \preceq_{i,P} | \preceq_{i,D} | \ \pi;\pi' \mid \pi \cup \pi' \mid -\pi \mid \varphi? \\ \varphi & ::= & p \mid x \mid \neg\varphi \mid \varphi \wedge \varphi' \mid [\pi]\varphi. \end{array}$$

Our first conjecture is that we can find a polysize reduction of the satisfiability problem for $\mathcal{L}^{if,cf}_{\text{DLCA}}$ to the satisfiability problem of converse PDL with nominals, also called converse combinatory PDL.[9] The latter problem is known to be EXPTIME-complete

[9] The main idea of the polynomial embedding is to exploit the iteration construct * of PDL for the translation tr of the cognitive programs, by stipulating that $tr(\equiv_i) = (any_i \cup -any_i)^*$, $tr(\preceq_{i,P}) = P_i^*$, $tr(\preceq_{i,D}) = D_i^*$, and homomorphic otherwise, where $\mathcal{A}_i$ is agent i's set of atomic programs (or actions), $\mathcal{A} = \bigcup_{i \in Agt} \mathcal{A}_i$ is the set of PDL atomic programs, $any_i = \bigcup_{a_i \in \mathcal{A}_i}$ and, finally, P_i^* and D_i^* are special atomic programs in $\mathcal{A}_i$.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

(De Giacomo 1995). Therefore, if our conjecture is true, we will be able to conclude that the satisfiability problems for the fragments $\mathcal{L}^{sa,ap}_{\mathrm{DLCA}}$ and $\mathcal{L}^{if,cf}_{\mathrm{DLCA}}$ are both EXPTIME-complete.

We also intend to study complexity of the nominal-free (nf) fragment of $\mathcal{L}_{\mathrm{DLCA}}$, noted $\mathcal{L}^{nf}_{\mathrm{DLCA}}$. Nominals play a technical role in the logic DLCA by making it easier the task of axiomatizing intersection and complement of programs (Axioms **Add2**$_\cap$ and **Comp1**$_\sim$ in Definition 13). Our second conjecture is that the language $\mathcal{L}_{\mathrm{DLCA}}$ is strictly more expressive than its nominal-free fragment $\mathcal{L}^{nf}_{\mathrm{DLCA}}$. Our third conjecture is that we can find a polysize reduction of the satisfiability problem for $\mathcal{L}^{nf}_{\mathrm{DLCA}}$ to the satisfiability problem of boolean modal logic with a bounded number of modal parameters which is known to be EXPTIME-complete (Lutz and Sattler 2000). We leave the proof of the previous three conjectures to future work. We leave to future work (i) the proof of the previous three conjectures and (ii) the development of tableau-based automated reasoning procedures for the language $\mathcal{L}_{\mathrm{DLCA}}$ and for its fragments $\mathcal{L}^{sa,ap}_{\mathrm{DLCA}}$, $\mathcal{L}^{if,cf}_{\mathrm{DLCA}}$, and $\mathcal{L}^{nf}_{\mathrm{DLCA}}$ which can be used for programming artificial agents endowed with cognitive attitudes.

Well-foundedness Future work will also be devoted to study a variant of our logic DLCA under the assumption of converse well-foundedness for the relation $\preceq_{i,P}$ and well-foundedness for the relation $\preceq_{i,D}$. As emphasized in Section 3, these properties are required to make agents' beliefs and desires consistent, namely, to guarantee that the formulas $\neg(\mathsf{B}_i\varphi \wedge \mathsf{B}_i\neg\varphi)$, $\neg\mathsf{B}_i\bot$, $\neg(\mathsf{D}_i\varphi \wedge \mathsf{D}_i\neg\varphi)$, and $\neg\mathsf{D}_i\top$ become valid. We will define the logic DLCA^{wf} to be the extension of the logic DLCA of Definition 13 by the following two axioms:

$$\langle\equiv_i\rangle\psi \rightarrow \langle\equiv_i\rangle(\psi \wedge [\prec_{i,P}]\neg\psi) \qquad (\mathbf{CWF}_{\preceq_{i,P}})$$

$$\langle\equiv_i\rangle\psi \rightarrow \langle\equiv_i\rangle(\psi \wedge [\succ_{i,D}]\neg\psi). \qquad (\mathbf{WF}_{\preceq_{i,D}})$$

Such axioms are variants of the so-called Gödel-Löb (**GL**) axiom from provability logic (Boolos 1993). Our conjecture is that the logic DLCA^{wf} so defined is sound and complete for the class of MCMs whose relations $\preceq_{i,D}$ and $\preceq_{i,P}$ are, respectively, well-founded and conversely well-founded.

Ceteris paribus preference We also plan to study a *ceteris paribus* notion of dyadic preference in the sense of Von Wright (Wright 1963), which has been recently formalized in a modal logic setting by van Benthem *et al.* (2009). According to Von Wright, for an agent to have a preference of φ over ψ, she should prefer a situation in which φ is true to a situation in which ψ is true, *all other things being equal*.[10] Our aim is to show that the DLCA framework is expressive enough to capture both the static and the dynamic aspects of this notion of *ceteris paribus* preference.

Acknowledgments

This work was supported by the ANR project CoPains ("Cognitive Planning in Persuasive Multimodal Communication"). Support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute is also gratefully acknowledged.

[10] See also Wellman and Doyle (1991) for a "ceteris paribus" interpretation of the notion of goal.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

References

Alchourrón, C. E., Gardenfors, P. and Makinson, D. 1985. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic 50*, 510–530.

Amor, N. B., Fargier, H., Sabbadin, R. and Trabelsi, M. 2019. Possibilistic games with incomplete information. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI 2019)*, 1544–1550.

Anglberger, A. J., Gratzl, N. and Roy, O. 2015. Obligation, free choice, and the logic of weakest permissions. *The Review of Symbolic Logic 8*, 807–827.

Aucher, G. 2005. A combined system for update logic and belief revision. In *Intelligent Agents and Multi-Agent Systems, 7th Pacific Rim International Workshop on Multi-Agents (PRIMA 2004)*. LNCS, vol. 3371. Springer, 1–17.

Aumann, R. 1999. Interactive epistemology I: Knowledge. *International Journal of Game Theory 28,* 3, 263–300.

Aumann, R. and Brandenburger, A. 1995. Epistemic conditions for Nash equilibrium. *Econometrica 63*, 1161–1180.

Balbiani, P., Herzig, A. and Troquard, N. 2008. Alternative axiomatics and complexity of deliberative stit theories. *Journal of Philosophical Logic 37,* 4, 387–406.

Baltag, A., Moss, L. and Solecki, S. 1998. The logic of public announcements, common knowledge and private suspicions. In *Proceedings of TARK'98*. Morgan Kaufmann, 43–56.

Baltag, A. and Moss, L. S. 2004. Logics for epistemic programs. *Synthese 139,* 2, 165–224.

Baltag, A. and Smets, S. 2008. A qualitative theory of dynamic interactive belief revision. In *Proceedings of LOFT 7*. Texts in Logic and Games, vol. 3. Amsterdam University Press, 13–60.

Baltag, A. and Smets, S. 2009. Talking your way into agreement: Belief merge by persuasive communication. In *Proceedings of the Second Multi-Agent Logics, Languages, and Organisations Federated Workshops (MALLOW)*, vol. 494. CEUR.

Belnap, N., Perloff, M. and Xu, M. 2001. *Facing the Future: Agents and Choices in Our Indeterminist World*. Oxford University Press.

Benferhat, S., Dubois, D., Prade, H. and Williams, M.-A. 2002. A practical approach to revising prioritized knowledge bases. *Studia Logica 70*, 105–130.

Boolos, G. 1993. *The Logic of Provability*. Cambridge University Press.

Boutilier, C. 1993. Revision sequences and nested conditionals. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI 1993)*. Morgan Kaufmann, 519–525.

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

Brafman, R. I. and Tennenholtz, M. 1996. On the foundations of qualitative decision theory. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI'96)*. AAAI Press, 1291–1296.

Brafman, R. I. and Tennenholtz, M. 2000. An axiomatic treatment of three qualitative decision criteria. *Journal of the ACM 47,* 3, 452–482.

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

Darwiche, A. and Pearl, J. 1997. On the logic of iterated belief revision. *Artificial Intelligence 89,* 1–2, 1–29.

De Giacomo, G. 1995. *Decidability of Class-Based Knowledge Representation Formalisms*. Ph.D. thesis, Università di Roma "La Sapienza".

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press

Doyle, J. and Thomason, R. 1999. Background to qualitative decision theory. *The AI Magazine 20,* 2, 55–68.

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*. MIT Press, Cambridge, Massachusetts.

Gargov, G. and Goranko, V. 1993. Modal logic with names. *Journal of Philosophical Logic 22*, 607–636.

Harel, D., Kozen, D. and Tiuryn, J. 2000. *Dynamic Logic*. MIT Press, Cambridge, Massachusetts.

Hemaspaandra, E. 1996. The price of universality. *Notre Dame Journal of Formal Logic 37,* 2, 174–203.

Hintikka, J. 1962. *Knowledge and Belief: an Introduction to the Logic of the Two Notions*. Cornell University Press.

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)*. AAAI Press, 572–574.

Lewis, D. 1979. A problem about permission. In *Essays in Honour of Jaakko Hintikka*. 163–175.

Liu, F. 2011. *Reasoning about Preference Dynamics*. Springer.

Lorini, E. 2013. Temporal STIT logic and its application to normative reasoning. *Journal of Applied Non-Classical Logics 23,* 4, 372–399.

Lorini, E. 2017. Logics for games, emotions and institutions. *If-CoLog Journal of Logics and their Applications 4,* 9, 3075–3113.

Lutz, C. and Sattler, U. 2000. The complexity of reasoning with boolean modal logics. In *Proceedings of the Third Conference on Advances in Modal logic (AiML 3)*. World Scientific, 329–348.

Marsh, K. and Wallace, H. 2005. The influence of attitudes on beliefs: Formation and change. In *The Handbook of Attitudes*, D. Albarracin, B. T. Johnson and M. P. Zanna, Eds. Lawrence Erlbaum Ass., 369–395.

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.

Passy, S. and Tinchev, T. 1991. An essay in combinatorial dynamic logic. *Information and Computation 93*, 263–332.

Paternotte, C. 2011. Rational choice theory. In *SAGE Handbook for the Philosophy of Social Sciences*, I. Jarvie and J. Zamora-Bonilla, Eds. SAGE Publications Inc., 307–321.

Platts, M. 1979. *Ways of Meaning*. Routledge, and Kegan Paul.

Rott, H. 2009. Shifting priorities: Simple representations for 27 iterated theory change operators. In *Towards Mathematical Philosophy: Papers from the Studia Logica conference Trends in Logic IV*, D. Makinson, J. Malinowski, and H. Wansing, Eds. Springer, 269–296.

Searle, J. 1979. *Expression and Meaning*. Cambridge University Press.

Shoham, Y. 2009. Logical theories of intention and the database perspective. *Journal of Philosophical Logic 38,* 6, 633–647.

Spohn, W. 1982. How to make sense of game theory. In *Philosophy of Economics*, vol. 2, 239–270.

Spohn, W. 1988. Ordinal conditional functions: A dynamic theory of epistemic states. In *Causation in Decision, Belief Change and Statistics*. Kluwer, 105–134.

van Benthem, J. 2007. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics 17,* 2, 129–155.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University 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. and Liu, F. 2007. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics 17,* 2, 157–182.

van Ditmarsch, H., van der Hoek, W. and Kooi, B. 2007. *Dynamic Epistemic Logic*, vol. 337. Synthese Library, Springer.

van Ditmarsch, H. P. 2005. Prolegomena to dynamic logic for belief revision. *Synthese 147,* 2, 229–275.

van Eijck, J. 2008. Yet more modal logics of preference change and belief revision. In *New Perspectives on Games and Interaction*. Texts in Logic and Games, vol. 4. Amsterdam University Press, 81–104.

Wellman, M. P. and Doyle, J. 1991. Preferential semantics for goals. In *Proceedings of the Ninth National conference on Artificial intelligence (AAAI'91)*, 698–703.

Wooldridge, M. 2000. *Reasoning about Rational Agents*. MIT Press, Cambridge.

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.

https://doi.org/10.1017/S1471068421000053 Published online by Cambridge University Press