



**HAL**  
open science

# Reasoning about Cognitive Attitudes in a Qualitative Setting

Emiliano Lorini

► **To cite this version:**

Emiliano Lorini. Reasoning about Cognitive Attitudes in a Qualitative Setting. 16th European Conference on Logics in Artificial Intelligence (JELIA 2019), May 2019, Rende, Italy. pp.726-743. hal-02414905

**HAL Id: hal-02414905**

**<https://hal.science/hal-02414905>**

Submitted on 16 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Reasoning about Cognitive Attitudes in a Qualitative Setting

Emiliano Lorini

IRIT-CNRS, Toulouse University, France

**Abstract.** We present a general logical framework for reasoning about agents' cognitive attitudes of both epistemic type and motivational type. We provide a sound and complete axiomatization for our logic and 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, strong desire, comparative desirability and choice.

## 1 Introduction

Since the seminal work of Hintikka on epistemic logic [18], of Von Wright on the logic of preference [29,30] and of Cohen & Levesque on the logic of intention [11], many formal logics for reasoning about cognitive attitudes of agents such as knowledge and belief [15], preference [22,6], desire [14], intention [28,20] and their combination [24,31] 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 [8,13] 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 maxmin principle (choose the action that will minimize potential loss) and the maxmax principle (choose the action that will maximize potential gain) have been studied and axiomatically characterized [9,10].

The aim of this paper is to present a rich logical framework for representing a variety of agents' cognitive attitudes in a multi-agent setting. In agreement with philosophical theories [26,27,19,23], our logic allows us to distinguish two general categories of cognitive attitudes: *epistemic* attitudes, including belief and knowledge, and *motivational* ones, including desires and preferences. Moreover, in agreement with rational choice

theory, it allows us to capture a notion of choice which depends on what an agent believes and desires as well as on the decision criterion she adopts.

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, respectively, 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, converse and test) to build complex cognitive attitudes from simple ones. Following [25,16], it also uses nominals in order to axiomatize intersection 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, comparative desirability and choice. In Section 4, we present a sound and complete axiomatization for it. In Section 5 we conclude.

## 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.

**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  $v \preceq_{i,\tau} w$ ;
- $V : W \rightarrow 2^{Atm \cup Nom}$  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.

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

$$\begin{aligned} \pi &::= \equiv_i | \preceq_{i,P} | \preceq_{i,D} | \sim_{i,P} | \sim_{i,D} | \pi; \pi' | \pi \cup \pi' | \pi \cap \pi' | \neg \pi | \varphi? \\ \varphi &::= p | x | \neg \varphi | \varphi \wedge \varphi' | [\pi] \varphi \end{aligned}$$

where  $p$  ranges over  $Atm$ ,  $x$  ranges over  $Nom$  and  $i$  ranges over  $Agt$ . The other Boolean constructions  $\top$ ,  $\perp$ ,  $\vee$ ,  $\rightarrow$  and  $\leftrightarrow$  are defined from  $p$ ,  $\neg$  and  $\wedge$  in the standard way.

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

Cognitive programs correspond to the basic constructions of Propositional Dynamic Logic (PDL) [17]: atomic programs of type  $\equiv_i$ ,  $\preceq_{i,P}$ ,  $\preceq_{i,D}$ ,  $\preceq_{i,P}^{\sim}$  and  $\preceq_{i,D}^{\sim}$ , sequential composition ( $;$ ), non-deterministic choice ( $\cup$ ), intersection ( $\cap$ ), converse ( $\bar{\phantom{x}}$ ) and test ( $?$ ). A given cognitive program  $\pi$  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 “ $\varphi$  is true, according to the cognitive program  $\pi$ ”. 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 [15,2].  $[\equiv_i]\varphi$  has to be read “ $\varphi$  is true according to what agent  $i$  knows” or more simply “agent  $i$  knows that  $\varphi$  is true”, which just means that “ $\varphi$  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 “ $\varphi$  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 “ $\varphi$  is true at all states that, according to agent  $i$ , are at least as desirable as the current one”. The atomic programs  $\preceq_{i,P}^{\sim}$  and  $\preceq_{i,D}^{\sim}$  are the complements of the atomic programs  $\preceq_{i,P}$  and  $\preceq_{i,D}$ , respectively. In particular,  $[\preceq_{i,P}^{\sim}]\varphi$  has to be read “ $\varphi$  is true at all states that, according to agent  $i$ , are *not* at least as plausible as the current one”, while  $[\preceq_{i,D}^{\sim}]\varphi$  has to be read “ $\varphi$  is true at all states that, according to agent  $i$ , are *not* at least as desirable as the current one”. The program constructs  $;$ ,  $\cup$ ,  $\cap$ ,  $\bar{\phantom{x}}$  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 “ $\varphi$  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 “ $\varphi$  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  $\mathcal{L}_{DLCA}$ :

**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  $R_\pi$  on  $W$  is inductively defined as follows, with  $\tau \in \{P, D\}$ :

$$\begin{aligned}
wR_{\equiv_i}v &\text{ iff } w \equiv_i v \\
wR_{\prec_{i,\tau}}v &\text{ iff } w \prec_{i,\tau} v \\
wR_{\prec_{i,\tau}^\sim}v &\text{ iff } w \equiv_i v \text{ and } w \not\prec_{i,\tau} v \\
wR_{\pi;\pi'}v &\text{ iff } \exists u \in W : wR_\pi u \text{ and } uR_{\pi'}v \\
wR_{\pi \cup \pi'}v &\text{ iff } wR_\pi v \text{ or } wR_{\pi'}v \\
wR_{\pi \cap \pi'}v &\text{ iff } wR_\pi v \text{ and } wR_{\pi'}v \\
wR_{-\pi}v &\text{ iff } vR_\pi w \\
wR_{\varphi?}v &\text{ iff } w = v \text{ and } M, w \models \varphi
\end{aligned}$$

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:

$$\begin{aligned}
\succ_{i,\tau} &=_{\text{def}} \prec_{i,\tau} \\
\succsim_{i,\tau} &=_{\text{def}} \succ_{i,\tau} \cap \prec_{i,\tau}^\sim \\
\prec_{i,\tau}^\sim &=_{\text{def}} \prec_{i,\tau} \\
\prec_{i,\tau} &=_{\text{def}} \prec_{i,\tau} \cap \succ_{i,\tau}^\sim \\
\approx_{i,\tau} &=_{\text{def}} \prec_{i,\tau} \cap \succ_{i,\tau}
\end{aligned}$$

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  $\varphi$  in  $\mathcal{L}_{\text{DLCA}}$  we say that  $\varphi$  is valid if and only if for every multi-agent cognitive model  $M$  and world  $w$  in  $M$ , we have  $M, w \models \varphi$ . Conversely, we say that  $\varphi$  is satisfiable if  $\neg\varphi$  is not valid.

For a given multi-agent cognitive model  $M = (W, (\prec_{i,P})_{i \in \text{Agt}}, (\prec_{i,D})_{i \in \text{Agt}}, (\equiv_i)_{i \in \text{Agt}}, N, V)$ , we define  $\|\varphi\|_M = \{v \in W : M, v \models \varphi\}$  to be the truth set of  $\varphi$  in  $M$ . Moreover, for every  $w \in W$  and for every  $i \in \text{Agt}$ , 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  $\varphi$  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 Section 1.

#### 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  $\varphi$  if and only if  $\varphi$  is true at all states that the agent considers maximally plausible.

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

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

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

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

It is worth noting that the set  $\text{Best}_{i,P}(w)$  in Definition 3 might be empty, since it is not necessarily the case that the relation  $\preceq_{i,P}$  is conversely well-founded.<sup>1</sup> As a consequence, the belief operator  $\text{B}_i$  does not necessarily satisfy Axiom D, i.e., the formula  $\text{B}_i \varphi \wedge \text{B}_i \neg \varphi$  is satisfiable in the logic DLCA. More details about these aspects will be given at the end of Section 4.

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

**Definition 4 (Strong belief).** Let  $M = (W, (\preceq_{i,P})_{i \in \text{Agt}}, (\preceq_{i,D})_{i \in \text{Agt}}, (\equiv_i)_{i \in \text{Agt}}, V)$  be a MCM and let  $w \in W$ . We say that agent  $i$  strongly believes that  $\varphi$  at  $w$ , denoted by  $M, w \models \text{SB}_i \varphi$ , 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  $M = (W, (\preceq_{i,P})_{i \in \text{Agt}}, (\preceq_{i,D})_{i \in \text{Agt}}, (\equiv_i)_{i \in \text{Agt}}, V)$  be a MCM and let  $w \in W$ . Then, we have

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

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

$$\models (\text{SB}_i \varphi \wedge \langle \equiv_i \rangle \varphi) \rightarrow \text{B}_i \varphi \quad (1)$$

Conditional belief is another notion which has been studied by epistemic logicians given its important role in belief dynamics [5]. We say that an agent believes that  $\varphi$  conditional on  $\psi$ , or she would believe that  $\varphi$  if she learnt that  $\psi$ , if and only if, according to agent  $i$ , all most plausible  $\psi$ -worlds are also  $\varphi$ -worlds.

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

<sup>1</sup> 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  $v$ .

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  $\varphi$  conditional on  $\psi$ , so that the latter can be represented in the language of the logic DLCA.

**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*

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

### 3.2 Motivational attitudes

The first kind of motivational attitude we consider is desire. Following [14], we say that an agent desires that  $\varphi$  if and only if all states that the agent envisages at which  $\varphi$  is true is true are not minimally desirable for the agent. In other words, desiring that  $\varphi$  consists in having some degree of attraction for all situations in which  $\varphi$  is true.

**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  $\varphi$  at  $w$ , denoted by  $M, w \models D_i \varphi$ , 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  $\equiv_i; [\succ_{i,D}] \perp?$ .

**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*

$$M, w \models D_i \varphi \text{ iff } M, w \models [\equiv_i; [\succ_{i,D}] \perp?] \neg \varphi.$$

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.<sup>2</sup> As a consequence, desires are not necessarily consistent, i.e., the formula  $D_i \varphi \wedge D_i \neg \varphi$  is satisfiable in the logic DLCA. As emphasized by [14], this notion of desire satisfies the following property:

$$\models D_i \varphi \rightarrow D_i(\varphi \wedge \psi) \tag{2}$$

Indeed, if an agent has some degree of attraction for all situations in which  $\varphi$  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  $\varphi$ -situations. It is a property that this notion of desire shares with the *open reading* of the concept of permission studied in the area of deontic logic (see, e.g., [1,21]).<sup>3</sup>

<sup>2</sup> 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  $v$ .

<sup>3</sup> According to deontic logicians, there are at least two candidate readings of the statement “ $\varphi$  is permitted”: (i) every instance of  $\varphi$  is OK according to the normative regulation, and (ii) at least one instance of  $\varphi$  (but possibly not all) is OK according to the normative regulation. The former is the so-called *open reading* of permission.

One way of blocking this inference is by strengthening the notion of desire. We say that an agent strongly desires that  $\varphi$  if and only if, according to agent  $i$ , all  $\varphi$ -worlds are strictly more desirable than all  $\neg\varphi$ -worlds.

**Definition 7 (Strong desire).** Let  $M = (W, (\preceq_{i,P})_{i \in \text{Agt}}, (\preceq_{i,D})_{i \in \text{Agt}}, (\equiv_i)_{i \in \text{Agt}}, V)$  be a MCM and let  $w \in W$ . We say that agent  $i$  strongly desires that  $\varphi$  at  $w$ , denoted by  $M, w \models \text{SD}_i\varphi$ , 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  $M = (W, (\preceq_{i,P})_{i \in \text{Agt}}, (\preceq_{i,D})_{i \in \text{Agt}}, (\equiv_i)_{i \in \text{Agt}}, V)$  be a MCM and let  $w \in W$ . Then, we have

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

We have that strong desire implies desire:

$$\models \text{SD}_i\varphi \rightarrow \text{D}_i\varphi \quad (3)$$

Differently from desiring, it is not necessarily the case that strongly desiring that  $\varphi$  implies strongly desiring that  $\varphi \wedge \psi$ , i.e., the formula  $\text{SD}_i\varphi \wedge \neg\text{SD}_i(\varphi \wedge \psi)$  is satisfiable in the logic DLCA. Indeed, strongly desiring that  $\varphi$  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 one.

### 3.3 From comparative desirability to choice

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

Let us first define comparative desirability according to the optimistic view.

**Definition 8 (Comparative desirability: optimistic view).** Let  $M = (W, (\preceq_{i,P})_{i \in \text{Agt}}, (\preceq_{i,D})_{i \in \text{Agt}}, (\equiv_i)_{i \in \text{Agt}}, V)$  be a MCM and let  $w \in W$ . We say that, according to agent  $i$ 's optimistic assessment,  $\varphi$  is at least as desirable as  $\psi$  at  $w$ , denoted by  $M, w \models \text{D}_i^{\text{Opt}}(\psi \preceq \varphi)$ , 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  $\mathcal{L}_{\text{DLCA}}$ .



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

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

The following abbreviation defines *strict* comparative desirability according to the optimistic view:

$$D_i^{\text{Opt}}(\psi \prec \varphi) =_{\text{def}} D_i^{\text{Opt}}(\psi \preceq \varphi) \wedge \neg D_i^{\text{Opt}}(\varphi \preceq \psi)$$

$D_i^{\text{Opt}}(\psi \prec \varphi)$  has to be read “according to  $i$ ’s optimistic assessment,  $\varphi$  is more desirable than  $\psi$ ”.

Let us now define comparative desirability according to the pessimistic view.

**Definition 9 (Comparative desirability: pessimistic view).** Let  $M = (W, (\preceq_{i,P})_{i \in \text{Agt}}, (\preceq_{i,D})_{i \in \text{Agt}}, (\equiv_i)_{i \in \text{Agt}}, V)$  be a MCM and let  $w \in W$ . We say that, according to agent  $i$ ’s pessimistic assessment,  $\varphi$  is at least as desirable as  $\psi$  at  $w$ , denoted by  $M, w \models D_i^{\text{Pess}}(\psi \preceq \varphi)$ , 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  $\mathcal{L}_{\text{DLCA}}$ .

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

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

The following abbreviation defines *strict* comparative desirability according to the pessimistic view:

$$D_i^{\text{Pess}}(\psi \prec \varphi) =_{\text{def}} D_i^{\text{Pess}}(\psi \preceq \varphi) \wedge \neg D_i^{\text{Pess}}(\varphi \preceq \psi)$$

$D_i^{\text{Pess}}(\psi \prec \varphi)$  has to be read “according to  $i$ ’s pessimistic assessment,  $\varphi$  is more desirable than  $\psi$ ”.

The previous notion of (optimistic and pessimistic) comparative desirability does not depend on what the agent believes. This means that, in order to assess whether  $\varphi$  is at least as desirable as  $\psi$ , an agent also takes into account worlds that are implausible (or, more generally, not maximally plausible). *Realistic* comparative desirability requires that an agent compares two formulas  $\varphi$  and  $\psi$  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 [8,9,10].

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

**Definition 10 (Realistic comparative desirability: optimistic view).** Let  $M = (W, (\preceq_{i,P})_{i \in \text{Agt}}, (\preceq_{i,D})_{i \in \text{Agt}}, (\equiv_i)_{i \in \text{Agt}}, V)$  be a MCM and let  $w \in W$ . We say that, according to agent  $i$ ’s optimistic assessment,  $\varphi$  is realistically at least as desirable as  $\psi$  at  $w$ , denoted by  $M, w \models \text{RD}_i^{\text{Opt}}(\psi \preceq \varphi)$ , if and only if  $\forall u \in \text{Best}_{i,P}(w) \cap \|\psi\|_{i,w,M}, \exists v \in \text{Best}_{i,P}(w) \cap \|\varphi\|_{i,w,M} : u \preceq_{i,D} v$ .

The idea is that an “optimistic” agent  $i$  considers  $\varphi$  *realistically* at least as desirable as  $\psi$  if and only if, for every  $\psi$ -situation in agent  $i$ 's belief set there exists a  $\varphi$ -situation in agent  $i$ 's belief set such that the latter is at least as desirable as the former.

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

**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*

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

We define:

$$\text{RD}_i^{Opt}(\psi \prec \varphi) =_{def} \text{RD}_i^{Opt}(\psi \preceq \varphi) \wedge \neg \text{RD}_i^{Opt}(\varphi \preceq \psi)$$

$\text{RD}_i^{Opt}(\psi \prec \varphi)$  has to be read “according to agent  $i$ 's optimistic assessment,  $\varphi$  is realistically more desirable than  $\psi$ ”.

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

**Definition 11 (Realistic comparative desirability: 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,  $\varphi$  is realistically at least as desirable as  $\psi$  at  $w$ , denoted by  $M, w \models \text{RD}_i^{Pess}(\psi \preceq \varphi)$ , if and only if  $\forall v \in \text{Best}_{i,P}(w) \cap \|\varphi\|_{i,w,M}, \exists u \in \text{Best}_{i,P}(w) \cap \|\psi\|_{i,w,M} : u \preceq_{i,D} v$ .*

The idea is that a “pessimistic” agent  $i$  considers  $\varphi$  *realistically* at least as desirable as  $\psi$  if and only if, for every  $\varphi$ -situation in agent  $i$ 's belief set there exists a  $\psi$ -situation in agent  $i$ 's belief set such that the former is at least as desirable as the latter.

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

**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*

$$M, w \models \text{RD}_i^{Pess}(\psi \preceq \varphi) \text{ iff } M, w \models [\equiv_i; [\prec_{i,P}] \perp?; \varphi?] \langle \preceq_{i,D} \cap (\equiv_i; [\prec_{i,P}] \perp?) \rangle \psi.$$

We define:

$$\text{RD}_i^{Pess}(\psi \prec \varphi) =_{def} \text{RD}_i^{Pess}(\psi \preceq \varphi) \wedge \neg \text{RD}_i^{Pess}(\varphi \preceq \psi)$$

$\text{RD}_i^{Pess}(\psi \prec \varphi)$  has to be read “according to agent  $i$ 's pessimistic assessment,  $\varphi$  is realistically more desirable than  $\psi$ ”.

We conclude this section by defining two notions of choice: agent  $i$ 's optimistic choice that  $\varphi$ , denoted by  $C_i^{Opt} \varphi$ , and agent  $i$ 's pessimistic choice that  $\varphi$ , denoted by  $C_i^{Pess} \varphi$ .

$$C_i^{Opt} \varphi =_{def} \text{RD}_i^{Opt}(\neg \varphi \prec \varphi)$$

$$C_i^{Pess} \varphi =_{def} \text{RD}_i^{Pess}(\neg \varphi \prec \varphi)$$

According to these definitions, an optimistic (resp. pessimistic) agent should choose  $\varphi$  if and only if, according to her optimistic (resp. pessimistic) assessment,  $\varphi$  is realistically more desirable than  $\neg \varphi$ .

## 4 Axiomatization

In this section, we provide a sound and complete axiomatization for the logic DLCA. The first step consists in precisely defining this logic.

**Definition 12.** We define DLCA to be the extension of classical propositional logic given by the following axioms and rules with  $\tau \in \{P, D\}$ :

$([\pi]\varphi \wedge [\pi](\varphi \rightarrow \psi)) \rightarrow [\pi]\psi$	<b>(K<sub>π</sub>)</b>
$[\equiv_i]\varphi \rightarrow \varphi$	<b>(T<sub>≡<sub>i</sub></sub>)</b>
$[\equiv_i]\varphi \rightarrow [\equiv_i][\equiv_i]\varphi$	<b>(4<sub>≡<sub>i</sub></sub>)</b>
$\neg[\equiv_i]\varphi \rightarrow [\equiv_i]\neg[\equiv_i]\varphi$	<b>(5<sub>≡<sub>i</sub></sub>)</b>
$[\preceq_{i,\tau}]\varphi \rightarrow \varphi$	<b>(T<sub>⊲<sub>i,τ</sub></sub>)</b>
$[\preceq_{i,\tau}]\varphi \rightarrow [\preceq_{i,\tau}][\preceq_{i,\tau}]\varphi$	<b>(4<sub>⊲<sub>i,τ</sub></sub>)</b>
$[\equiv_i]\varphi \rightarrow [\preceq_{i,\tau}]\varphi$	<b>(Int<sub>⊲<sub>i,τ</sub>,≡<sub>i</sub></sub>)</b>
$(\langle \equiv_i \rangle \varphi \wedge \langle \equiv_i \rangle \psi) \rightarrow (\langle \equiv_i \rangle (\varphi \wedge \langle \preceq_{i,\tau} \rangle \psi) \vee \langle \equiv_i \rangle (\psi \wedge \langle \preceq_{i,\tau} \rangle \varphi))$	<b>(Conn<sub>⊲<sub>i,τ</sub>,≡<sub>i</sub></sub>)</b>
$[\pi; \pi']\varphi \leftrightarrow [\pi][\pi']\varphi$	<b>(Red<sub>;</sub>)</b>
$[\pi \cup \pi']\varphi \leftrightarrow ([\pi]\varphi \wedge [\pi']\varphi)$	<b>(Red<sub>∪</sub>)</b>
$([\pi]\varphi \wedge [\pi']\psi) \rightarrow [\pi \cap \pi'](\varphi \wedge \psi)$	<b>(Add1<sub>∩</sub>)</b>
$\langle \langle \pi \rangle x \wedge \langle \pi' \rangle x \rangle \rightarrow \langle \pi \cap \pi' \rangle x$	<b>(Add2<sub>∩</sub>)</b>
$\varphi \rightarrow [\pi]\langle -\pi \rangle \varphi$	<b>(Conv1<sub>-</sub>)</b>
$\varphi \rightarrow [-\pi]\langle \pi \rangle \varphi$	<b>(Conv2<sub>-</sub>)</b>
$([\preceq_{i,\tau}]\varphi \wedge [\preceq_{i,\tau}^{\sim}]\varphi) \leftrightarrow [\equiv_i]\varphi$	<b>(Comp1<sub>~</sub>)</b>
$\langle \preceq_{i,\tau} \rangle x \rightarrow [\preceq_{i,\tau}^{\sim}]\neg x$	<b>(Comp2<sub>~</sub>)</b>
$[?\varphi]\psi \rightarrow (\varphi \rightarrow \psi)$	<b>(Red<sub>?</sub>)</b>
$\langle \pi \rangle (x \wedge \varphi) \rightarrow [\pi'](x \rightarrow \varphi)$	<b>(Most<sub>x</sub>)</b>
$\frac{\varphi}{[\pi]\varphi}$	<b>(Nec<sub>π</sub>)</b>
$\frac{[\pi]\neg x \text{ for all } x \in Num}{[\pi]\perp}$	<b>(Cov)</b>

For every  $\varphi \in \mathcal{L}_{DLCA}$ , we write  $\vdash \varphi$  to denote the fact that  $\varphi$  is a theorem of DLCA, i.e., there exists an at most countably infinite sequence  $\psi_0, \psi_1, \dots$  such that  $\psi_0 = \varphi$  and for all  $k \geq 0$ ,  $\psi_k$  is an instance of some axiom or  $\psi_k$  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 multi-agent cognitive models.

Soundness, namely checking that the axioms are valid and the 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 [25,16].

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

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

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

**Definition 13.** A set of formulas  $\Sigma$  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  $\perp \notin \Sigma$ . It is said to be a maximal consistent theory (MCT) if it is a consistent theory and, for each consistent theory  $\Sigma'$ , we have that if  $\Sigma \subseteq \Sigma'$  then  $\Sigma = \Sigma'$ .

We have the following property for theories.

**Proposition 10.** Let  $\Sigma$  be a theory and let  $\varphi \in \mathcal{L}_{\text{DLCA}}$ . Then,  $\Sigma + \varphi$  is a theory. Moreover, if  $\Sigma$  is consistent then either  $\Sigma + \varphi$  is consistent or  $\Sigma + \neg\varphi$  is consistent.

The following proposition highlights some standard properties of MCTs.

**Proposition 11.** Let  $\Sigma$  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 [25, Lemma 4.15].

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

The following lemma highlights a fundamental properties of MCTs.

**Lemma 2.** Let  $\Sigma$  be a MCT. Then, there exists  $x \in \text{Num}$  such  $x \in \Sigma$ .

Let us now define the canonical model for our logic.

**Definition 14.** The canonical model is the tuple  $M^c = (W^c, (\preceq_{i,P}^c)_{i \in \text{Agt}}, (\preceq_{i,D}^c)_{i \in \text{Agt}}, (\equiv_i^c)_{i \in \text{Agt}}, V^c)$  such that:

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

Let us now define the canonical relations for the complex programs  $\pi$ .

**Definition 15.** Let  $M^c = (W^c, (\preceq_{i,P}^c)_{i \in \text{Agt}}, (\preceq_{i,D}^c)_{i \in \text{Agt}}, (\equiv_i^c)_{i \in \text{Agt}}, V^c)$  be the canonical model. Then, for all  $\pi \in \mathcal{P}$  and for all  $w, v \in W^c$ :

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

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

**Lemma 3.** *Let  $M^c = (W^c, (\preceq_{i,P}^c)_{i \in \text{Agt}}, (\preceq_{i,D}^c)_{i \in \text{Agt}}, (\equiv_i^c)_{i \in \text{Agt}}, V^c)$  be the canonical model. Then, for all  $\Sigma, \Sigma' \in W^c$ , for all  $\pi \in \mathcal{P}$  and for all  $x \in \text{Num}$ , if  $x \in \Sigma, x \in \Sigma'$  and  $\Sigma R_\pi^c \Sigma'$  then  $\Sigma = \Sigma'$ .*

The next step consists in proving the following existence lemma.

**Lemma 4.** *Let  $M^c = (W^c, (\preceq_{i,P}^c)_{i \in \text{Agt}}, (\preceq_{i,D}^c)_{i \in \text{Agt}}, (\equiv_i^c)_{i \in \text{Agt}}, V^c)$  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  $w R_\pi^c v$  and  $\varphi \in v$ .*

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

**Lemma 5.** *Let  $M^c = (W^c, (\preceq_{i,P}^c)_{i \in \text{Agt}}, (\preceq_{i,D}^c)_{i \in \text{Agt}}, (\equiv_i^c)_{i \in \text{Agt}}, V^c)$  be the canonical model, let  $w \in W^c$  and let  $\varphi \in \mathcal{L}_{\text{DLCA}}$ . 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  $\mathcal{L}_{\text{DLCA}}$  which turns out to be equivalent to the original semantics based on MCMs.

**Definition 16 (Quasi multi-agent cognitive model).** *A quasi multi-agent cognitive model (quasi-MCM) is a tuple  $M = (W, (\preceq_{i,P})_{i \in \text{Agt}}, (\preceq_{i,D})_{i \in \text{Agt}}, (\equiv_i)_{i \in \text{Agt}}, V)$  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  $V_{\text{Nom}}(w) \cap V_{\text{Nom}}(v) \neq \emptyset$  and  $w R_\pi v$  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  $\mathcal{L}_{\text{DLCA}}$ .

**Proposition 12.** *Let  $\varphi \in \mathcal{L}_{\text{DLCA}}$ . Then,  $\varphi$  is valid relative to the class of MCMs if and only if  $\varphi$  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  $\not\models \neg \varphi$ . 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 11, 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  $\varphi$  is satisfiable relative to the class of quasi-MCMs. Therefore, by Proposition 12,  $\varphi$  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 multi-agent cognitive models.*

We conclude this section by discussing the properties 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  $\neg(B_i\varphi \wedge B_i\neg\varphi)$  and  $\neg(D_i\varphi \wedge D_i\neg\varphi)$  become valid. It turns out that these properties can be easily added to our logical framework.

In particular, let us consider the class of multi-agent cognitive models whose relations  $\preceq_{i,D}$  and  $\preceq_{i,P}$  are, respectively, well-founded and conversely well-founded.

Furthermore, let us consider the following two axioms:

$$\begin{aligned} \langle \equiv_i \rangle \psi &\rightarrow \langle \equiv_i \rangle (\psi \wedge [\prec_{i,P}] \neg \psi) && \text{(CWF}_{\preceq_{i,P}}\text{)} \\ \langle \equiv_i \rangle \psi &\rightarrow \langle \equiv_i \rangle (\psi \wedge [\succ_{i,D}] \neg \psi) && \text{(WF}_{\preceq_{i,D}}\text{)} \end{aligned}$$

Let us define  $\text{DLCA}^{wf}$  to be the extension of the logic DLCA of Definition 12 by these axioms. It is straightforward to verify that the logic  $\text{DLCA}^{wf}$  is sound for the class of multi-agent cognitive models whose relations  $\preceq_{i,D}$  and  $\preceq_{i,P}$  are, respectively, well-founded and conversely well-founded. We conjecture that we can easily adapt the proof of Theorem 1 to show that it is also complete.

## 5 Conclusion

We have presented a logical framework for modelling a rich variety of cognitive attitudes of both epistemic type and motivational type. We have provided a sound and complete axiomatization for our logic.

Directions of future research are manifold. The present paper is devoted to study the proof-theoretic aspects of the logic. In future work, we plan to investigate its computational aspects including decidability of its satisfiability problem and, at a later stage, complexity. In order to prove decidability, we expect to be able to use existing filtration techniques from modal logic. Following the literature on dynamic epistemic logic (DEL) [12], we also plan to study several dynamic extensions of our logic in order to capture a large variety of cognitive dynamics in a multi-agent setting. The latter includes belief change, desire change and choice change. We believe choice change is particularly interesting given the dependence of an agent's choices on her beliefs and desires, as illustrated in Section 3.3. Specifically, since an agent's choices depend on her plausibility and desirability orderings over possible worlds, if these orderings change, then the agent's choices may also change. In other words, choice change can be seen as derivative of belief change and desire change. Another research direction we plan to follow in the future is to connect the notion of choice formalized in Section 3.3 with a notion of action in the sense of STIT logic, the logic of "seeing to it that" by [4]. The interesting aspect of STIT is that agents' choices are explicit in its semantics and agents' actions are conceived as results of their choices.

## Acknowledgments

This work was supported by the ANR project CoPains. I would like to thank Philippe Balbiani for his useful comments on its content.

## References

1. A. J. Anglberger, N. Gratzl and O. Roy. Obligation, free choice, and the logic of weakest permissions. *The Review of Symbolic Logic*, 8:807–827, 2015.
2. R. Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, 28(3):263–300, 1999.
3. A. Baltag and S. Smets. Talking your way into agreement: Belief merge by persuasive communication. In *Proceedings of the Second Multi-Agent Logics, Languages, and Organisations Federated Workshops (MALLOW)*, volume 494 of CEUR Workshop Proceedings, 2009.
4. N. Belnap, M. Perloff and M. Xu. *Facing the future: agents and choices in our indeterminist world*. Oxford University Press, 2001.
5. J. van Benthem. Dynamic logic for belief revision *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007.
6. J. van Benthem, P. Girard, and O. Roy. Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic*, 38:83–125, 2009.
7. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.
8. C. Boutilier. Towards a logic for qualitative decision theory. In *Proceedings of International Conference on Principles of Knowledge Representation and Reasoning (KR' 94)*, pages 75–86. AAAI Press, 1994.
9. R. I. Brafman and M. Tennenholtz. An axiomatic treatment of three qualitative decision criteria. *Journal of the ACM*, 47(3):452–482.
10. R. I. Brafman and M. Tennenholtz. On the foundations of qualitative decision theory. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI'96)*, pages 1291–1296. AAAI Press, 1996.
11. P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
12. H. van Ditmarsch, W. van der Hoek and B. Kooi. *Dynamic epistemic logic*. Synthese Library, vol. 337, Springer, 2007.
13. J. Doyle and R. Thomason. Background to qualitative decision theory. *The AI Magazine*, 20(2):55–68, 1999.
14. D. Dubois, E. Lorini and H. Prade. The strength of desires: a logical approach. *Minds and Machines*, 27(1):199–231, 2017.
15. R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, Massachusetts, 1995.
16. G. Gargov and V. Goranko. Modal logic with names. *Journal of Philosophical Logic*, 22:607–636, 1993.
17. D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, Cambridge, Massachusetts, 2000.
18. J. Hintikka. *Knowledge and belief: an introduction to the logic of the two notions*. Cornell University Press, 1962.
19. I. L. Humberstone. Direction of fit. *Mind*, 101(401):59–83, 1992.
20. T. F. Icard, E. Pacuit, and Y. Shoham. Joint Revision of Beliefs and Intention. In *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning (KR 2010)*, pages 572–574. AAAI Press, 2010.
21. D. Lewis. A problem about permission. In *Essays in honour of Jaakko Hintikka*, pages 163–175. Springer, 1979.
22. F. Liu. *Reasoning about Preference Dynamics*. Springer, 2011.
23. E. Lorini. Logics for Games, Emotions and Institutions. *If-CoLog Journal of Logics and their Applications*, 4(9):3075–3113, 2017.

24. J. J. Ch. Meyer, W. van der Hoek, and B. van Linder. A logical approach to the dynamics of commitments. *Artificial Intelligence*, 113(1-2):1–40, 1999.
25. S. Passy and T. Tinchev. An essay in combinatorial dynamic logic. *Information and Computation*, 93:263–332, 1991.
26. M. Platts. *Ways of meaning*. Routledge and Kegan Paul, 1979.
27. J. Searle. *Expression and meaning*. Cambridge University Press, 1979.
28. Y. Shoham. Logical Theories of Intention and the Database Perspective. *Journal of Philosophical Logic*, 38(6):633–647, 2009.
29. G. H. Von Wright. *The logic of preference*. Edinburgh University Press, 1963.
30. G. H. Von Wright. The logic of preference reconsidered. *Theory and Decision*, 3:140–169, 1972.
31. M. Wooldridge. *Reasoning about rational agents*. MIT Press, Cambridge, 2000.

## A Proofs

We provide a selection of the proofs for the results given in the paper.

### A.1 Proof of Lemma 2

*Proof.* We prove the lemma by reductio ad absurdum. Let  $\Sigma$  be a MCT. Moreover, suppose that, for all  $x \in \text{Nom}$ ,  $x \notin \Sigma$ . By Proposition 11, it follows that, for all  $x \in \text{Nom}$ ,  $\neg x \in \Sigma$ .

By Axiom **Red**<sub>?</sub>, we have  $\neg x \leftrightarrow [?\top]\neg x \in \Sigma$  for all  $x \in \text{Nom}$ . Thus, for all  $x \in \text{Nom}$ ,  $[?\top]\neg x \in \Sigma$ . Hence, since  $\Sigma$  is closed under **Cov**,  $[?\top]\perp \in \Sigma$ . By Axiom **Red**<sub>?</sub>, the latter is equivalent to  $\perp \in \Sigma$ . The latter is contradiction with the fact that  $\Sigma$  is a MCT.  $\square$

### A.2 Proof of Lemma 3

*Proof.* Let us first prove that (i) if  $x \in \Sigma$  and  $\varphi \in \Sigma$  then  $[\pi](x \rightarrow \varphi) \in \Sigma$ . Suppose  $x, \varphi \in \Sigma$ . Thus,  $x \wedge \varphi \in \Sigma$  since  $\Sigma$  is a MCT. Moreover,  $(x \wedge \varphi) \rightarrow [\pi](x \rightarrow \varphi) \in \Sigma$ , because of Axiom **Most**<sub>x</sub>. Hence,  $[\pi](x \rightarrow \varphi) \in \Sigma$ .

Now let us prove by absurdum that (ii) if  $x \in \Sigma$ ,  $\Sigma'$  and  $\Sigma R_\pi^c \Sigma'$  then  $\Sigma = \Sigma'$ . Suppose  $x \in \Sigma$ ,  $\Sigma'$ ,  $\Sigma R_\pi^c \Sigma'$  and  $\Sigma \neq \Sigma'$ . The latter implies that there exists  $\varphi$  such that  $\varphi \in \Sigma$  and  $\varphi \notin \Sigma'$ . By item (i) above, it follows that  $[\pi](x \rightarrow \varphi) \in \Sigma$ . Since  $\Sigma R_\pi^c \Sigma'$ , the latter implies that  $x \rightarrow \varphi \in \Sigma'$ . Since  $x \in \Sigma'$ , it follows that  $\varphi \in \Sigma'$  which leads to a contradiction.  $\square$

### A.3 Proof of Lemma 4

*Proof.* Suppose  $w$  is a MCT and  $\langle \pi \rangle \varphi \in w$ . It follows that  $[\pi]w = \{\psi : [\pi]\psi \in w\}$  is a consistent theory. Indeed, it is easy to check that  $[\pi]w$  contains all theorems of DLCA, is closed under modus ponens and rule **Cov**. Let us prove that it is consistent by reductio ad absurdum. Suppose  $\perp \in [\pi]w$ . Thus,  $[\pi]\perp \in w$ . Hence,  $[\pi]\neg\varphi \in w$ . Since  $\langle \pi \rangle \varphi \in w$ ,  $\perp \in w$ . The latter contradicts the fact that  $w$  is a MCT. Let us distinguish two cases.

Case 1:  $\varphi \in [\pi]w$ . Thus,  $\neg\varphi \notin [\pi]w$  since  $w$  is consistent. Thus, by Lemma 1, there exists MCT  $v$  such that  $[\pi]w \subseteq v$ ,  $\varphi \in v$  and  $\neg\varphi \notin v$ . By definition of  $R_\pi^c$ ,  $w R_\pi^c v$ .

Case 2:  $\varphi \notin [\pi]w$ . By Lemma 10,  $[\pi]w + \varphi$  is a theory since  $[\pi]w$  is a theory.  $[\pi]w + \varphi$  is consistent. Suppose it is not. Thus,  $\varphi \rightarrow \perp \in [\pi]w$  and, consequently,  $\neg\varphi \in [\pi]w$ . Hence,  $[\pi]\neg\varphi \in w$ . It follows that  $\perp \in w$ , since  $\langle \pi \rangle \varphi \in w$ . But this contradicts the fact that  $w$  is a MCT. Thus,  $[\pi]w + \varphi$  is a consistent theory. Moreover,  $\varphi \in [\pi]w + \varphi$ ,  $\neg\varphi \notin [\pi]w + \varphi$  and  $[\pi]w \subseteq [\pi]w + \varphi$ . By Lemma 1, there exists MCT  $v$  such that  $[\pi]w \subseteq v$ ,  $\varphi \in v$  and  $\neg\varphi \notin v$ . By definition of  $R_\pi^c$ ,  $w R_\pi^c v$ .  $\square$



#### A.4 Proof of Lemma 6

*Proof.* The fact that  $M^c$  satisfies Constraints C3 and C4\* follows from Lemma 2 and Lemma 3. To prove that  $\equiv_i$  is an equivalence relation that  $\preceq_{i,D}^c$  and  $\preceq_{i,D}$  are preorders and that  $M^c$  satisfies Constraints C1 and C2 is just a routine exercise. Indeed, Axioms  $\mathbf{T}_{\equiv_i}$ ,  $\mathbf{4}_{\equiv_i}$ ,  $\mathbf{5}_{\equiv_i}$ ,  $\mathbf{T}_{\preceq_{i,\tau}}$ ,  $\mathbf{4}_{\preceq_{i,\tau}}$ ,  $\mathbf{Int}_{\preceq_{i,\tau},\equiv_i}$  and  $\mathbf{Conn}_{\preceq_{i,\tau},\equiv_i}$  are canonical for these semantic conditions.

To conclude, we need to prove that the following six conditions hold, for  $i \in \text{Agt}$  and  $\tau \in \{P, D\}$ :

$$\begin{aligned}
(w, v) \in R_{\preceq_{i,\tau}}^c &\text{ iff } (w, v) \in R_{\equiv_i}^c \text{ and } (w, v) \notin R_{\preceq_{i,\tau}}^c \\
(w, v) \in R_{\pi;\pi'}^c &\text{ iff } \exists u \in W^c : (w, u) \in R_{\pi}^c \text{ and } (u, v) \in R_{\pi'}^c \\
(w, v) \in R_{\pi \cup \pi'}^c &\text{ iff } (w, v) \in R_{\pi}^c \text{ or } (w, v) \in R_{\pi'}^c \\
(w, v) \in R_{\pi \cap \pi'}^c &\text{ iff } (w, v) \in R_{\pi}^c \text{ and } (w, v) \in R_{\pi'}^c \\
(w, v) \in R_{-\pi}^c &\text{ iff } (v, w) \in R_{\pi}^c \\
wR_{\varphi?}^c v &\text{ iff } w = v \text{ and } M^c, w \models \varphi
\end{aligned}$$

We only prove the second and fourth conditions which are the most difficult ones to prove.

Let us start with the proof of the second condition. The right-to-left direction is standard. We only prove the left-to-right direction. Suppose  $(w, v) \in R_{\pi;\pi'}^c$ . Let  $[\pi]w = \{\psi : [\pi]\psi \in w\}$ . Moreover, let  $\langle \pi' \rangle v = \{\langle \pi' \rangle \psi : \psi \in v\}$ . Finally, let  $\langle \pi' \rangle \psi_1, \langle \pi' \rangle \psi_2, \dots$  be an enumeration of the elements of  $\langle \pi' \rangle v$ . We define  $\Sigma^1 = [\pi]w + \langle \pi' \rangle \psi_1$  and, for all  $k > 1$ ,  $\Sigma^k = \Sigma^{k-1} + \langle \pi' \rangle \psi_k$ . By Lemma 10 and the fact that  $[\pi]w$  is a theory, it can be shown that every  $\Sigma^k$  is a theory. Moreover, by induction on  $k$ , it can be shown that every  $\Sigma^k$  is consistent. Since  $\Sigma^{k-1} \subseteq \Sigma^k$  for all  $k > 1$ , it follows that  $\Sigma = \bigcup_{k>1} \Sigma^k$  is a consistent theory. By Lemma 1 and the definition of  $\Sigma$ , there exists  $u \in W^c$  such that  $\Sigma \subseteq u$ ,  $(w, u) \in R_{\pi}^c$  and  $(u, v) \in R_{\pi'}^c$ .

Let us now prove the fourth condition. Suppose  $(w, v) \in R_{\pi \cap \pi'}^c$ . By Definition 15 and Proposition 11, it follows that, for all  $\varphi$ , if  $\varphi \in v$  then  $\langle \pi \cap \pi' \rangle \varphi \in w$ . The latter implies that for all  $\varphi$ , if  $\varphi \in v$  then  $\langle \pi \cap \pi' \rangle (\varphi \vee \perp) \in w$  since  $\vdash \langle \pi \cap \pi' \rangle \varphi \rightarrow \langle \pi \cap \pi' \rangle (\varphi \vee \perp)$ . By Axiom  $\mathbf{K}_{\pi}$ , it follows that, for all  $\varphi$ , if  $\varphi \in v$  then  $\langle \pi \rangle \varphi \vee \langle \pi' \rangle \perp \in w$ . Thus, for all  $\varphi$ , if  $\varphi \in v$  then  $\langle \pi \rangle \varphi \in w$ , since  $\vdash (\langle \pi \rangle \varphi \vee \langle \pi' \rangle \perp) \rightarrow \langle \pi \rangle \varphi$ . In a similar way, we can prove that, for all  $\varphi$ , if  $\varphi \in v$  then  $\langle \pi' \rangle \varphi \in w$ . By Definition 15 and Proposition 11, it follows that  $(w, v) \in R_{\pi}^c$  and  $(w, v) \in R_{\pi'}^c$ .

Now suppose  $(w, v) \in R_{\pi}^c$  and  $(w, v) \in R_{\pi'}^c$ . Thus, by Definition 15 and Proposition 11, (i) for all  $\varphi$ , if  $\varphi \in v$  then  $\langle \pi \rangle \varphi \in w$  and  $\langle \pi' \rangle \varphi \in w$ . By Proposition 11 and Lemma 2, we have that (ii) there exists  $x \in \text{Num}$  such that, for all  $\varphi$ ,  $\varphi \in v$  iff  $x \wedge \varphi \in v$ . Item (i) and item (ii) together imply that (iii) there exists  $x \in \text{Num}$  such that, for all  $\varphi$ , if  $\varphi \in v$  then  $\langle \pi \rangle (x \wedge \varphi) \in w$  and  $\langle \pi' \rangle (x \wedge \varphi) \in w$ . We are going to prove the following theorem:

$$\vdash (\langle \pi \rangle (x \wedge \varphi) \wedge \langle \pi' \rangle (x \wedge \varphi)) \rightarrow \langle \pi \cap \pi' \rangle (x \wedge \varphi)$$

By Axiom  $\mathbf{K}_{\pi}$ ,  $\langle \pi \rangle (x \wedge \varphi) \wedge \langle \pi' \rangle (x \wedge \varphi)$  implies  $\langle \pi \rangle x \wedge \langle \pi' \rangle x$ . By Axiom  $\mathbf{Add2}_{\cap}$ , the latter implies  $\langle \pi \cap \pi' \rangle x$ . Moreover, by Axiom  $\mathbf{Int}_{\preceq_{i,\tau},\equiv_i}$  and Axiom  $\mathbf{Most}_x$ ,  $\langle \pi \rangle (x \wedge \varphi)$  implies  $[\equiv_{\emptyset}](x \rightarrow \varphi)$ . By Axiom  $\mathbf{Int}_{\preceq_{i,\tau},\equiv_i}$ , the latter implies  $[\pi \cap \pi'](x \rightarrow \varphi)$ . By Axiom  $\mathbf{K}_{\pi}$ ,  $[\pi \cap \pi'](x \rightarrow \varphi)$  and  $\langle \pi \cap \pi' \rangle x$  together imply  $\langle \pi \cap \pi' \rangle (x \wedge \varphi)$ . Thus,  $\langle \pi \rangle (x \wedge \varphi) \wedge \langle \pi' \rangle (x \wedge \varphi)$  implies  $\langle \pi \cap \pi' \rangle (x \wedge \varphi)$ .

From previous item (iii) and the previous theorem it follows that there exists  $x \in \text{Num}$  such that, for all  $\varphi$ , if  $\varphi \in v$  then  $\langle \pi \cap \pi' \rangle (x \wedge \varphi)$ . The latter implies that, for all  $\varphi$ , if  $\varphi \in v$  then  $\langle \pi \cap \pi' \rangle \varphi$ . The latter implies that  $(w, v) \in R_{\pi \cap \pi'}^c$ .  $\square$