Philippe Balbiani
David Fernández-Duque
Emiliano Lorini

The Dynamics of Epistemic Attitudes in Resource-Bounded Agents

Abstract. The paper presents a new logic for reasoning about the formation of beliefs through perception or through inference in non-omniscient resource-bounded agents. The logic distinguishes the concept of explicit belief from the concept of background knowledge. This distinction is reflected in its formal semantics and axiomatics: (i) we use a non-standard semantics putting together a neighborhood semantics for explicit beliefs and relational semantics for background knowledge, and (ii) we have specific axioms in the logic highlighting the relationship between the two concepts. Mental operations of perceptive type and inferential type, having effects on epistemic states of agents, are primitives in the object language of the logic. At the semantic level, they are modelled as special kinds of model-update operations, in the style of dynamic epistemic logic (DEL). Results about axiomatization, decidability and complexity for the logic are given in the paper.

Keywords: Epistemic logic; cognitive agents; resource-bounded reasoning

1. Introduction

Most existing logical theories of epistemic attitudes developed in the area of epistemic logic assume that agents are omniscient, in the sense that: (i) their beliefs are closed under conjunction and implication, *i.e.*, if φ is believed and ψ is believed then $\varphi \wedge \psi$ is believed and if φ is believed and $\varphi \to \psi$ is believed then ψ is believed; (ii) their beliefs are closed under logical consequence (*alias* valid implication), *i.e.*, if φ is believed and φ logically implies ψ, *i.e.*, $\varphi \to \psi$ is valid, then ψ is believed as well; (iii) they believe all valid sentences or tautologies; (iv) they have introspection over their beliefs, *i.e.*, if φ is believed then it is believed that φ is believed.

As pointed out by [25, 31], relaxing the assumption of logical omniscience allows for a resource-bounded agent who might fail to draw any connection between φ and its logical consequence ψ and, consequently, who might not believe some valid sentences and who might need time to infer and form new beliefs from her existing knowledge and beliefs.

Presented by **Name of Editor**; *Received* December 1, 2005

The aim of this paper is to propose a new logic which helps in clarifying how a non-omniscient resource-bounded agent can form new beliefs either through perception or through inference from her existing knowledge and beliefs. More precisely, the aim of the paper is to introduce a dynamic logic, called DLEK (Dynamic Logic of Explicit Beliefs and Knowledge) in which programs are mental operations, either of perceptive type or of inferential type, having effects on epistemic states of resourced-bounded agents.

DLEK is a logical theory of the relationship between explicit beliefs and background knowledge, both from a static and from a dynamic perspective. This distinction is reflected in its formal semantics and axiomatics: (i) we use a non-standard semantics putting together a neighborhood semantics for explicit beliefs and relational semantics for background knowledge, and (ii) we have specific axioms in the logic highlighting the relationship between the two concepts.

Related work DLEK is not the first logic of epistemic attitudes for non-omniscient agents. Logics of awareness have been studied in the recent years (see, *e.g.*, [36, 24, 22, 1]) starting from the seminal work of Fagin & Halpern [15]. These logics distinguish between awareness, implicit belief and explicit belief. Awareness and implicit belief are primitive and explicit belief is defined as that implicit belief the agent is aware of. (See [28] for a critical discussion on this family of logics.) The crucial difference between DLEK and existing logics of awareness is that the latter make no use of concepts as 'reasoning' or 'inference'. On the contrary, DLEK provides a constructive theory of explicit beliefs, as it accounts for the perceptive and inferential steps leading from an agent's knowledge and beliefs to new beliefs. This 'constructive' aspect of epistemic attitudes is something our theory shares with other approaches in the literature including the dynamic theory of evidence-based beliefs by [35] — that also use a neighborhood semantics for the notion of evidence —, the sentential approach to explicit beliefs and their dynamics by [29], and the dynamic theory of explicit and implicit beliefs by [38].

The logic of inference steps by Velázquez-Quesada [39] and the logical system $\mathrm{DES4}_n$ by Duc [11] share a similar view with us as well. In particular, Velázquez-Quesada shares with us the idea of modeling inference steps by means of dynamic operators in the style of dynamic epistemic logic (DEL) [37]. Nonetheless, our conceptual framework is different from his (see Section 2). He does not distinguish the concept of explicit belief and the concept of background knowledge, which is the fundamental distinction of our logic

DLEK. Furthermore, he does not provide any axiomatization or decidability result for his logic of inference steps, as he only provides a semantics.

Duc's system DES4_n combines epistemic operators for truthful and non-omniscient knowledge of type K_i, that allows to represent the fact that a certain agent i knows something, with tense-like operators of type $\langle F_i \rangle$, that allows to represent the fact that a certain agent i will *possibly* get to know something.[1] Duc's system shares with our logic DLEK the idea that (i) an agent gets to know (or believe) something by performing inferences, and (ii) making inferences takes time. Nonetheless, while in our logic DLEK inferential operations are represented both at the syntactic level, via dynamic operators in the DEL style, and a semantic level, as model update operations, in Duc's system and its formal semantics given by [2] they are not. The formal semantics for DES4_n given by Ågotnes & Alechina is a Kripke-style semantics in which every operator $\langle F_i \rangle$ is associated with an abstract accessibility relation R_i over possible worlds, where wR_iv means that world w is related with world v. This semantics does not say anything about the kind of inferential operation (or sequence of inferential operations) that is responsible for the transition from world w to world v. More generally, the system DES4_n does not support reasoning about the consequences of a specific kind of inferential operation such as application of *modus ponens* or closure under conjunction on an agent's knowledge or beliefs. On the contrary, this is something our logic DLEK can express.

Another related work is [4], which presents a semantics for a variant of Alternating-time Temporal Logic (ATL) extended by explicit knowledge operators. This logic supports reasoning about, among other things, the capability of an agent of gaining some knowledge by applying a certain reasoning rule such as reasoning with *modus ponens* or reasoning monotonically. This logic does not take into account the distinction between explicit knowledge (or belief) and background knowledge which is fundamental for our analysis. But, the most important difference between this system and ours is their belonging to two different families of logics. Ågotnes & Walicki's logic is a logic for strategic reasoning with a semantics based on the concept of concurrent game structure (CGS). It has modal operators of strategic capability that represent what an agent or coalition of agents can achieve by playing a certain strategy. On the contrary, our logic DLEK belongs to the family of dynamic epistemic logics whose main constituents are dynamic modal operators that allow to represent the consequences of a specific com-

[1] The dual of $\langle F_i \rangle$, denoted by $[F_i]$, allows to represent the fact that a certain agent i will *necessarily* get to know something.

municative action or mental operation. Apart from some rare exceptions (see, e.g., [32, 3]), the connection between the two families of logics has been rarely explored. Moreover, it is not the object of the present work.

Our constructive approach to explicit beliefs also distinguishes DLEK from existing logics of time-bounded reasoning which represent reasoning as a process that requires time (see, e.g., [6, 21]). Active logics [12, 14] are representative of the research in this area. They account for the formation of new beliefs due to the time-consuming application of inference rules to the beliefs that are under the focus of attention. Specifically, the basic semantics of active logics, as presented in [14], includes three components: (i) an agent's belief set, identifying all facts that an agent explicitly believes, (ii) an observation function, identifying all facts that an agent observes at a given time point, and (iii) an inference function, specifying what an agent should believe at the next time point on the basis of the application of the inference rules she possesses on her belief set, given her actual observations. Nonetheless, there are important differences between active logics and our logic DLEK. First of all, active logics do not belong to the family of modal logics, while DLEK does. The latter has an advantage. As we will show in the paper, we can use existing techniques from modal logic in order to prove results about mathematical and computational properties of our logic. This includes the canonical-model argument for proving completeness of its axiomatics, the filtration argument for proving decidability of its satisfiability problem, and modal tableaux for stating complexity of this problem. Second, while active logics provide models of reasoning based on long-term memory and short-term memory (or working memory) (see, e.g., [13]), they do not distinguish between the notion of *explicit belief* and the notion of *background knowledge*, conceived as different kinds of epistemic attitudes. As we will show in Section 2, these two notions are the basic building blocks of our theory of belief dynamics in resource-bounded agents. Third, by exploiting the rich model-update semantics of dynamic epistemic logic (DEL) [37], our logic DLEK accounts for a variety of mental operations (or processes) that have not been explored in the active logic literature. This includes, for example, *forgetting* that a certain fact is true, *ascribing* to someone the execution of certain inference step, *being uncertain* that someone has performed a certain inference step. The latter mental operations correspond to basic operations of mindreading in the sense of Theory of Mind (ToM) [19].

Plan of the paper The paper is organized as follows. In Section 2 we present the conceptual foundation of our logic DLEK, namely the general view about dynamics of beliefs in resource-bounded agents which underlies our formal theory. The general idea is that new beliefs can be formed either by perception or by inferring them from existing beliefs in working memory and by retrieving information from background knowledge in long-term memory. Section 3 presents the syntax and the semantics of DLEK. We will show that, in DLEK, perceptive and inferential steps are modelled as special kinds of model-update operations. Section 4 is devoted to presenting an axiomatic system and showing its completeness. Section 5 presents complexity results for the satisfiability problem of the static fragment of DLEK, called LEK, as well as a decidability result for the satisfiability problem of DLEK. Given that the semantics of the static fragment of DLEK are non-standard, in Section 6 we show how they can be reduced to purely relational semantics, albeit with additional modalities.

2. Conceptual framework

The cognitive architecture underlying the logic DLEK (Dynamic Logic of Explicit Beliefs and Knowledge) is represented in Figure 1. It clarifies the processing of information in human agents and human-like artificial agents.

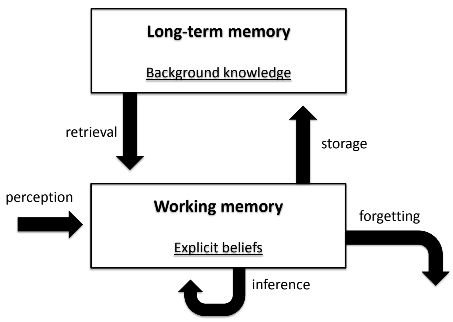

Figure 1. Cognitive view of the relationship between background knowledge and explicit beliefs

In accordance with existing psychological theories and computational models of memory and attention [7, 10, 40, 33], we assume that an agent has two kinds of information in her mind, those available in long-term memory (LTM) and those directly accessible in working memory (WM).

The information available in LTM, generally called *background knowledge*, includes both knowledge of specific events in the past and conceptual or causal knowledge representing the agent's unproblematic interpretation of reality.[2] For example, an agent may have background conceptual knowledge about how restaurants are organized or background causal knowledge about the relation between smoke and fire. In particular, she may know that restaurants have waiters, chairs and tables or that if smoke comes out from a window of a certain house, then there is fire inside the house.

WM retains information in an accessible state suitable for carrying out any kind of reasoning or decision task. In particular, following [27], we assume that the information available in an agent's working memory includes all *explicit beliefs* of the agent that occupy her consciousness and draw on her limited capacity of attention.[3] Some explicit beliefs are formed via *perception*. Formation of explicit beliefs via perception just consists of adding a new belief to the set of beliefs that are under the focus of the agent's attention. For example, an agent may look outside the window, see that it is raining outside, and thereby start believing that it is raining outside.

An agent can also use her explicit beliefs as premises of an *inference* which leads to the formation of a new belief. In some cases, formation of explicit beliefs via inference requires the *retrieval* of information from long-term memory. For example, suppose that an agent sees that smoke comes out from the window of a certain house and, as a result, she starts to explicitly believe this. The agent retrieves from her background knowledge stored in her long-term memory the information that if smoke comes out from a window of a certain house, then it means that there is fire inside the house. The agent can use this information together with the belief that smoke comes out from the window available in her working memory, to infer that there is fire inside the house and to form the corresponding belief.

Information can also be lost from working memory through *forgetting*: an agent may explicitly believe something but not believe it anymore at a later point. Information can also be removed from working memory and stored in long-term memory to make it available for a later moment. *Storage* of information in long-term memory might be necessary, given the limited capabilities of working memory. In the paper, we do not discuss this latter operation. We leave its formalization for future work.

In the next section we present the syntax and the semantics of the logic DLEK which makes precise all concepts informally discussed in this section.

[2]In the Soar architecture [30] these two kinds of background are called, respectively, episodic memory and semantic memory.

[3]Some psychologists (e.g., [10]) distinguish focus of attention from working memory, as they assume that there might be information activated in working memory which is not under the focus of the agent's attention. For simplicity, we here assume that focus of attention and working memory are coextensive.

3. Logical framework

DLEK is a logic which consists of a static component and a dynamic one. The static component, called LEK, is a logic of explicit beliefs and background knowledge. The dynamic component extends the static one with dynamic operators capturing the consequences of the agents' mental operations on their explicit beliefs. Since private mental operations may lead other agents to not contemplate the real world as a possibility, we will work with KD45 models, although we also consider logics DLEK^+, LEK^+, where a truthful S5 notion of background knowledge is used instead.

3.1. Syntax

Assume a countable set of atomic propositions $Atm = \{p, q, \ldots\}$ and a finite set of agents $Agt = \{1, \ldots, n\}$. By *Prop* we denote the set of all propositional formulas, i.e. the set of all Boolean formulas built out of the set of atomic propositions Atm.

The language of DLEK, denoted by $\mathcal{L}_{\mathsf{DLEK}}$, is defined by the following grammar in Backus-Naur form:

$$\begin{array}{lll} \alpha & ::= & \vdash(\varphi,\psi) \mid \cap(\varphi,\psi) \mid +\varphi \mid -\varphi \\ \varphi, \psi & ::= & p \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathsf{B}_i\varphi \mid \mathsf{K}_i\varphi \mid [(G{:}\alpha)_H]\varphi \end{array}$$

where p ranges over Atm and G, H range over 2^{Agt} with $G \subseteq H$.

The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from p, $\neg$ and $\wedge$ in the standard way. For every formula φ, let $SF(\varphi)$ be the closure under single negations of the set of all of φ's subformulas.

The language of LEK, the fragment of DLEK without dynamic operators, is denoted by $\mathcal{L}_{\mathsf{LEK}}$ and defined by the following grammar in Backus-Naur Form:

$$\varphi, \psi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathsf{B}_i\varphi \mid \mathsf{K}_i\varphi.$$

In what follows, we explain the meaning of the operators of our logic.

The formula $\mathsf{B}_i\varphi$ is read "the agent i explicitly believes that φ is true" or, more shortly, "the agent i believes that φ is true". As explained in Section 2, explicit beliefs are accessible in working memory and are the basic elements of the agents' reasoning process. Following the logic of local reasoning by Fagin & Halpern [15], we wish to give a purely semantical model of explicit belief. As we will show in Section 4, an effect of this approach is that agents cannot distinguish between logically equivalent formulas, i.e., if two facts φ and ψ are logically equivalent and an agent explicitly believes that φ is true, then she believes that ψ is true as well. There are other approaches, such as justification logics [5], that do not have this feature. However, the interpretation of these logics relies highly on syntax, a situation we wish to avoid in the present paper.

The modal operator K_i captures the notion of background knowledge discussed in Section 2. It represents the information that agent *i can* use to infer new explicit beliefs. Some pieces of background knowledge are either conceptual or causal and represent the agent's unproblematic interpretation of reality.

Unlike explicit beliefs, background knowledge is assumed to satisfy 'omniscience' principles like closure under conjunction and known implication, closure under logical consequence and introspection. Moreover, background knowledge is assumed to be veridical. Specifically, as we will show below, K_i is nothing but the well-known S5 (or, more generally, KD45) operator for knowledge widely used in computer science [16]. The fact that background knowledge is veridical and closed under logical consequence is justified in two terms. First of all, we assume it to be veridical since knowledge, differently from belief, is commonly assumed to be truthful. Secondly, we assume it to be closed under logical consequence since we conceive it as a kind of deductively closed *knowledge base*. Specifically, we assume background knowledge to include all facts that the agent has stored in her long-term memory (LTM), after having processed them in her working memory (WM), as well as all logical consequences of these facts.

The formula $[(G{:}\alpha)_H]\psi$ should be read "ψ holds after the mental operation (or mental action) α is performed by all agents in G, and the agents in H have common knowledge about this fact". Of course G and H may be singletons, but we also wish to model situations where a group of agents G makes a collective mental operation (say, after a discussion or public announcement) and the agents in a larger group H of which G is a subset have common knowledge about this fact. For example, suppose Ann and Bob are sitting on the sofa in the living-room while Charles gets into it. Ann and Bob see Charles getting into the living-room (i.e., $\{Ann, Bob\}$ is the set G). Charles is so noisy that everybody knows that Ann and Bob have seen him getting into the living-room. More generally, Ann, Bob and Charles have common knowledge that Ann and Bob have seen Charles getting into the living-room (i.e., $\{Ann, Bob, Charles\}$ is the set H). We may sometimes write $[\alpha]$ instead of $[(G{:}\alpha)_H]$ when G and H are clear from context.

We distinguish four types of mental operations α which allow us to capture some of the dynamic properties of explicit beliefs and background knowledge informally described in Section 2 above: $+\varphi$, $-\varphi$, $\vdash(\varphi,\psi)$ and $\cap(\varphi,\psi)$.

The operations $+\varphi$ and $-\varphi$ correspond, respectively, to the mental operations of forming an explicit belief via perception and forgetting an explicit belief represented in Figure 1. $\vdash(\varphi,\psi)$ and $\cap(\varphi,\psi)$ characterize two basic operations of forming explicit beliefs via inference. Specifically, $\vdash(\varphi,\psi)$ is the mental operation which consists in inferring ψ from φ in case φ is believed and, according to an agent's background knowledge, ψ is a logical consequence of φ. In other words, by performing this mental operation, an agent tries to retrieve from her background knowledge in long-term memory the information that φ implies ψ and, if she succeeds, she starts to believe ψ.[4] $\cap(\varphi,\psi)$ is the mental operation which consists in closing the explicit belief that φ and the explicit belief that ψ under conjunction. In other words, $\cap(\varphi,\psi)$ characterizes the mental operation of deducing $\varphi \wedge \psi$ from the explicit belief that φ and the explicit belief that ψ.

In this paper we assume that, differently from explicit beliefs, background knowledge is irrevocable in the sense of being stable over time [9]. In the conclusion, we will offer some insights on how to make background knowledge dynamic by including in our logic the operation of storing information in long-term memory, represented in Figure 1.

3.2. Semantics

The main notion in semantics is given by the following definition of LEK (LEK^+) model which provides the basic components for the interpretation of the static logics:

Definition 1 (LEK/LEK^+ model). *We define a LEK model to be a tuple $M = (W, N, R_1, \ldots, R_n, V)$ where:*

(a) W is a set of worlds or situations;

(b) for every $i \in Agt$, $R_i \subseteq W \times W$ is a serial, transitive and Euclidean[5] *relation on W;*

(c) $N : Agt \times W \longrightarrow 2^{2^W}$ is a neighborhood function such that for all $i \in Agt$, $w, v \in W$ and $X \subseteq W$:

(C1) *if $X \in N(i, w)$ then $X \subseteq R_i(w)$,*

(C2) *if wR_iv then $N(i, w) = N(i, v)$;*

---

[4]Note that retrieving the information "φ implies ψ" from long-term memory here just means that the agent uses it for drawing her inference, without 'loading' it in her explicit beliefs.

[5]Recall that R is *serial* if, for every w, there is a y such that wRy, and *Euclidean* if, whenever wRu and wRv, it follows that uRv.

(d) $V : W \longrightarrow 2^{Atm}$ *is a valuation function.*

A LEK^+ model *is a* LEK *model such that each* R_i *is reflexive.*

For every $i \in Agt$ and $w \in W$, $R_i(w) = \{v \in W : wR_iv\}$ identifies the set of situations that agent i considers possible at world w. In cognitive terms, $R_i(w)$ can be conceived as the set of all situations that agent i *can* retrieve from her long-term memory and reason about them. More generally, $R_i(w)$ is called agent i's epistemic state at w. In LEK^+-models, the reason why R_i is an equivalence relation is that it is used to model a form of omniscient background knowledge instead of omniscient background belief. The latter could be modelled, as in LEK-models, by replacing the equivalence relations R_i by serial, transitive and Euclidean relations commonly used in doxastic logic to model a notion of belief.

For every $i \in Agt$ and every $w \in W$, $N(i, w)$ defines the set of all facts that agent i explicitly believes at world w, a fact being identified with a set of worlds. More precisely, if $A \in N(i, w)$ then, at world w, agent i has the fact A under the focus of her attention and believes it. $N(i, w)$ is called agent i's explicit belief set at world w.

Constraint (**C1**) just means that an agent can have explicit in her mind only facts which are compatible with her current epistemic state. According to Constraint (**C2**), if world v is compatible with agent i's epistemic state at world w, then agent i should have the same explicit beliefs at w and v.

Truth conditions of DLEK formulas are inductively defined as follows.

Definition 2. *For a* LEK *model* $M = (W, N, R_1, \ldots, R_n, V)$*, a world* $w \in W$*, a formula* $\varphi \in \mathcal{L}_{\mathsf{LEK}}$*, and an action* α*, we define the truth relation* $M, w \models \varphi$ *and a new* LEK *model* $M^{(G:\alpha)_H}$ *by simultaneous recursion on* α *and* φ *as follows. Below, we write*

$$\|\varphi\|^M_{i,w} = \{v \in W : wR_iv \text{ and } M, v \models \varphi\}$$

whenever $M, v \models \varphi$ *is well-defined. Then, we set:*

$$\begin{aligned}
M, w \models p &\iff p \in V(w)\\
M, w \models \neg\varphi &\iff M, w \not\models \varphi\\
M, w \models \varphi \wedge \psi &\iff M, w \models \varphi \text{ and } M, w \models \psi\\
M, w \models \mathsf{B}_i\varphi &\iff ||\varphi||^M_{i,w} \in N(i, w)\\
M, w \models \mathsf{K}_i\varphi &\iff M, v \models \varphi \text{ for all } v \in R_i(w)\\
M, w \models [(G{:}\alpha)_H]\varphi &\iff M^{(G:\alpha)_H}, w_0 \models \varphi
\end{aligned}$$

where $M^{(G:\alpha)_H} = (W^{(G:\alpha)_H}, N^{(G:\alpha)_H}, R_1^{(G:\alpha)_H}, \ldots, R_n^{(G:\alpha)_H}, V^{(G:\alpha)_H})$ *is defined as follows.*

First, for $i \in Agt$ *and* $w \in W$*, set*

$$N^{+\psi}(i, w) = N(i, w) \cup \{||\psi||_{i,w}^M\}$$

$$N^{-\psi}(i, w) = N(i, w) \setminus \{||\psi||_{i,w}^M\}$$

$$N^{\vdash(\psi,\chi)}(i, w) = \begin{cases} N(i, w) \cup \{||\chi||_{i,w}^M\} & \text{if } M, w \models \mathsf{B}_i\psi \wedge \mathsf{K}_i(\psi \rightarrow \chi) \\ N(i, w) & \text{otherwise} \end{cases}$$

$$N^{\cap(\psi,\chi)}(i, w) = \begin{cases} N(i, w) \cup \{||\psi \wedge \chi||_{i,w}^M\} & \text{if } M, w \models \mathsf{B}_i\psi \wedge \mathsf{B}_i\chi \\ N(i, w) & \text{otherwise} \end{cases}$$

Then we define

$$\begin{aligned}
W^{(G:\alpha)_H} &= W \times \{0, 1\} \\
N^{(G:\alpha)_H}(i, w_0) &= \{X \times \{0\} : X \in N^{\alpha}(i, w)\} \text{ if } i \in G \\
N^{(G:\alpha)_H}(i, w_1) &= \{X \times \{1\} : X \in N(i, w)\} \text{ if } i \in G \\
N^{(G:\alpha)_H}(i, w_0) &= \{X \times \{0\} : X \in N(i, w)\} \text{ if } i \in H \setminus G \\
N^{(G:\alpha)_H}(i, w_1) &= \{X \times \{1\} : X \in N(i, w)\} \text{ if } i \in H \setminus G \\
N^{(G:\alpha)_H}(i, w_0) &= \{X \times \{1\} : X \in N(i, w)\} \text{ if } i \notin H \\
N^{(G:\alpha)_H}(i, w_1) &= \{X \times \{1\} : X \in N(i, w)\} \text{ if } i \notin H \\
R_i^{(G:\alpha)_H} &= \{(w_0, v_0) : wR_iv\} \cup \{(w_1, v_1) : wR_iv\} \text{ if } i \in H \\
R_i^{(G:\alpha)_H} &= \{(w_0, v_1) : wR_iv\} \cup \{(w_1, v_1) : wR_iv\} \text{ if } \text{ if } i \notin H \\
V^{(G:\alpha)_H}(w_x) &= V(w) \text{ with } x \in \{0, 1\}
\end{aligned}$$

where, for notational convenience, elements of $W^{(G:\alpha)_H}$ *are denoted by* w_0*,* w_1 *instead of* $(w, 0)$*,* $(w, 1)$*.*

The fact that the mental operation α is performed by the agents in G and this is only common knowledge for the agents in H is captured by the duplication of the original model M. The 0-part of the original model is the part in which agents in G update their explicit beliefs via the mental operation α, while the 1-part is the part in which nothing happens. All agents inside H are aware that those in G have changed their beliefs via the mental operation α, as they have access to the 0-part. On the contrary, the agents outside of H are not aware that those in G have changed their beliefs via the mental operation α, as they only have access to the 1-part thereby assuming that nothing happens.

Note that such an operation will cause the accessibility relation for agents not in H to no longer be reflexive, *even if they originally were.* Nevertheless, these operations *are* well-defined over the class of LEK models. Remark that for all dynamic operators α, for all $i \in Agt$ and for all $w \in W$, the neighborhood function $N^{\alpha}(i, w)$ still satisfies the constraints (**C1**) and (**C2**). Note that thanks to Constraint (**C1**), we can assume, in a mono-agent LEK model $M = (W, N, R_1, V)$, that R_1 is the universal relation on W.

PROPOSITION 1. *If M is a LEK model then $M^{(G:\alpha)_H}$ is also a LEK model. If moreover $H = Agt$ and M is a LEK^+ model, then $M^{(G:\alpha)_H}$ is also a LEK^+ model.*

According to the previous truth conditions, an agent i explicitly believes φ at world w if and only if, at world w, agent i has the fact corresponding to the formula φ (i.e., $||\varphi||^M_{i,w}$) included in her explicit belief set. Note that this semantic interpretation is different from the usual one for neighborhood formulas, as we restrict the extension of φ to the agent's epistemically accessible worlds. This is justified by the fact that an agent's explicit beliefs are relative to her background knowledge. The latter guarantees that if an agent explicitly believes that φ is true (i.e., $\mathsf{B}_i\varphi$) while having background knowledge that φ is false (i.e., $\mathsf{K}_i\neg\varphi$) then, she explicitly entertains a contradiction (i.e., $\mathsf{B}_i\bot$). This seems a reasonable property of explicit beliefs. Moreover, an agent has background knowledge that φ is true if and only if φ is true in all situations that are included in the agent's epistemic state.

Mental operations of the form α are formalized as model update operations that expand or contract the agents' explicit belief sets. In particular, the mental operation $+\psi$ consists of perceiving ψ and adding it to the explicit belief set, while the mental operation $-\psi$ consists of forgetting ψ and removing it from the explicit belief set. The mental operation $\vdash(\psi,\chi)$ consists of adding the explicit belief χ to an agent's explicit belief set if the agent believes ψ and has background knowledge that ψ implies χ. The mental operation $\cap(\psi,\chi)$ consists of adding the explicit belief $\psi \wedge \chi$ to an agent's explicit belief set if the agent explicitly believes both ψ and χ. Note that the preconditions for the mental operations $\vdash(\psi,\chi)$ and $\cap(\psi,\chi)$ do not appear in the semantic interpretation of the modality (as in, e.g., public announcements [37]); they are rather 'embedded' in the definition of the operation.

We write $\models_{\mathsf{DLEK}} \varphi$ ($\models_{\mathsf{DLEK}^+} \varphi$) to denote that φ is valid, i.e., φ is true at every world w of every LEK(LEK^+)-model M. In the next section we show some interesting validities of the logics DLEK and DLEK^+.

3.3. Some validities

The following four validities capture the basic properties of the four mental operations $\vdash(\varphi,\psi)$, $\cap(\varphi,\psi)$, $+\varphi$ and $-\varphi$ semantically defined above. Let $\varphi, \psi \in \mathit{Prop}$ and $i \in G \subseteq H \subseteq \mathit{Agt}$. Then, the following are valid over the class of LEK-models:

1. $(\mathsf{K}_i(\varphi \rightarrow \psi) \wedge \mathsf{B}_i\varphi) \rightarrow [(G{:} \vdash(\varphi,\psi))_H]\mathsf{B}_i\psi$
2. $(\mathsf{B}_i\varphi \wedge \mathsf{B}_i\psi) \rightarrow [(G{:} \cap (\varphi,\psi))_H]\mathsf{B}_i(\varphi \wedge \psi)$
3. $[(G{:} +\varphi)_H]\mathsf{B}_i\varphi$
4. $[(G{:} -\varphi)_H]\neg\mathsf{B}_i\varphi$

For instance, according to the first validity, if φ and ψ are propositional formulas, agent i explicitly believes φ and has background knowledge that φ implies ψ then, as a consequence of the mental operation $\vdash(\varphi,\psi)$, she will start to believe ψ. According to the third validity, if φ is a propositional formula then, as a consequence of perceiving that φ is true, agent i starts to explicitly believe that φ is true.

The reason why we need to impose that φ and ψ are propositional formulas is that there are DLEK-formulas such as the Moore-like formula $p \wedge \neg\mathsf{B}_i p$ for which the previous four principles do not hold. For instance, $[(\{i\}{:} + (p \wedge \neg\mathsf{B}_i p))_{\{i\}}]\mathsf{B}_i(p \wedge \neg\mathsf{B}_i p)$ is not valid.

It is worth noting that in the logics DLEK and DLEK^+ we can 'simulate' in a dynamic way the rule of necessitation. Indeed, for $\lambda \in \{\mathsf{DLEK}^+, \mathsf{DLEK}\}$, if $\models_\lambda \varphi$ and $i \in G$, then $\models_\lambda [(G{:} + \top)_H]\mathsf{B}_i\varphi$. This is a consequence of the semantic interpretation of the explicit belief operator (Definition 2) in which the extension of φ is restricted to the agent's epistemically accessible worlds.

We have only included a conjunction introduction operation for explicit beliefs for the sake of technical simplicity, but in the case of propositional formulas, other inference rules can be simulated using these operations, via the following validities, for $\varphi, \psi \in \mathit{Prop}$ and $i \in G \subseteq H \subseteq \mathit{Agt}$. Recall that for notational convenience we write α instead of $(G{:}\alpha)_H$.

1. $\models_{\mathsf{DLEK}} \quad (\mathsf{B}_i\varphi \wedge \mathsf{B}_i(\varphi \rightarrow \psi)) \rightarrow [\cap(\varphi,\varphi \rightarrow \psi)][\vdash(\varphi \wedge \psi,\psi)]\mathsf{B}_i\psi$
2. $\models_{\mathsf{DLEK}} \quad \mathsf{B}_i\varphi \rightarrow [\vdash(\varphi,\varphi \rightarrow (\varphi \vee \psi))]\mathsf{B}_i(\varphi \vee \psi)$
3. $\models_{\mathsf{DLEK}} \quad \mathsf{B}_i\psi \rightarrow [\vdash(\psi,\psi \rightarrow (\varphi \vee \psi))]\mathsf{B}_i(\varphi \vee \psi)$

4. $\models_{\mathsf{DLEK}}$ $\mathsf{B}_i(\varphi \wedge \psi) \rightarrow [\vdash(\varphi \wedge \psi, \varphi \wedge \psi \rightarrow \varphi))]\mathsf{B}_i\varphi$
5. $\models_{\mathsf{DLEK}}$ $\mathsf{B}_i(\varphi \wedge \psi) \rightarrow [\vdash(\varphi \wedge \psi, \varphi \wedge \psi \rightarrow \psi))]\mathsf{B}_i\psi$

Thus we can define modus ponens, disjunction introduction, and conjunction elimination using the mental operations we have presented. The general idea here is that in order to apply modus ponens, disjunction introduction, or conjunction elimination, an agent has to retrieve some information from her background knowledge. For example, in the case of conjunction elimination, she has to retrieve the information that "$\varphi \wedge \psi$ implies φ". Of course, an alternative approach where all of these operations are 'hard-wired' into the logic and do not require retrieval of information from background knowledge may be preferable if e.g we wish to minimize the number of reasoning steps that the agents need to use.

3.4. Example

In this section we are going to illustrate our logic DLEK with the help of a concrete example. The scenario goes as follows. There are two resource-bounded robotic assistants, say robot A (Anne) and robot B (Bob), who have to take care of a person. The person communicates with the robots via a coloured electric light which can be either red or green. The communication code is the following one: (i) if the electric light is red (atom r) then it means that the person needs help (atom h), and (ii) if the electric is green (atom g) then it means that the person is having a rest and wants not to be bothered (atom b). We assume that:

H1. robot A has full knowledge about the communication code as she knows that r implies h and that g implies b, and

H2. robot B has only partial knowledge about the communication code as he knows that r implies h but he does not know that g implies b.

Thus, let us suppose that $G = H = Agt = \{A, B\}$ and $Atm = \{r, g, h, b\}$. We represent the initial situation by a minimal LEK model satisfying the hypothesis H1 and H2 and which only excludes the impossible situations in which the electric light is both red and green and the person needs help and takes a rest at the same time.[6] This model is the tuple $\mathtt{MR} = (W, N, R_A, R_B, V)$ (where MR stands for 'model of the robots') such that:

[6]Note that the model satisfies the additional hypothesis that robot A and robot B have common knowledge that they share a part of the communication code, namely that each of them knows that r implies h.

- $W = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9\}$;
- $N(i, w) = \emptyset$ for all $i \in Agt$ and for all $w \in W$;
- the quotient set of W by R_A is $\{\{w_1, w_2, w_3, w_4, w_5\}, \{w_6, w_7\}, \{w_8, w_9\}\}$;
- the quotient set of W by R_B is $\{\{w_1, w_2, w_3, w_4, w_5, w_6, w_7\}, \{w_8, w_9\}\}$;
- $V(w_1) = \{r, h\}$, $V(w_2) = \{h\}$, $V(w_3) = \{g, b\}$, $V(w_4) = \{b\}$, $V(w_5) = \emptyset$, $V(w_6) = \{g\}$, $V(w_7) = \{g, h\}$, $V(w_8) = \{r\}$, and $V(w_9) = \{r, b\}$.

The fact that $N(i, w) = \emptyset$ for all $i \in Agt$ and for all $w \in W$ just means that in the initial situation the robots have no explicit belief in their short-term memories. Let us assume that w_1 is the actual situation in which the electric light is red and the person needs help.

The first thing to observe is that in the actual situation the hypothesis H1 and H2 are both satisfied. Indeed, as for H1, we have: $\mathtt{MR}, w_1 \models \mathsf{K}_A(r \rightarrow h) \wedge \mathsf{K}_A(g \rightarrow b)$. As for H2, we have: $\mathtt{MR}, w_1 \models \mathsf{K}_B(r \rightarrow h) \wedge \neg\mathsf{K}_B(g \rightarrow b)$.

Let us suppose that the person switches on the red light in order to signal to the two robots that she needs help. This event is represented by the mental operation $+r$ which leads from model $\mathtt{MR}$ to the updated model $\mathtt{MR}^{+r} = (W, N^{+r}, R_A, R_B, V)$ such that:

$N^{+r}(A, w_1) = \{\{w_1\}\}$, $N^{+r}(A, w_2) = \{\{w_1\}\}$, $N^{+r}(A, w_3) = \{\{w_1\}\}$,
$N^{+r}(A, w_4) = \{\{w_1\}\}$, $N^{+r}(A, w_5) = \{\{w_1\}\}$, $N^{+r}(A, w_6) = \emptyset$,
$N^{+r}(A, w_7) = \emptyset$, $N^{+r}(A, w_8) = \{\{w_8, w_9\}\}$, $N^{+r}(A, w_9) = \{\{w_8, w_9\}\}$,
$N^{+r}(B, w_1) = \{\{w_1\}\}$, $N^{+r}(B, w_2) = \{\{w_1\}\}$, $N^{+r}(B, w_3) = \{\{w_1\}\}$,
$N^{+r}(B, w_4) = \{\{w_1\}\}$, $N^{+r}(B, w_5) = \{\{w_1\}\}$, $N^{+r}(B, w_6) = \{\{w_1\}\}$,
$N^{+r}(B, w_7) = \{\{w_1\}\}$, $N^{+r}(B, w_8) = \{\{w_8, w_9\}\}$, $N^{+r}(B, w_9) = \{\{w_8, w_9\}\}$.

Note that the only difference between between A's and B's neighbourhoods at $\mathtt{MR}^{+r}$ is on worlds w_6 and w_7.

It is easy to check that in the new situation, after the mental operation $+r$ has been executed, the two robots explicitly believe that the light is red. That is: $\mathtt{MR}^{+r}, w_1 \models \mathsf{B}_{\{A,B\}}r$, where $\mathsf{B}_{\{A,B\}}\varphi$ is a abbreviation of $\mathsf{B}_A\varphi \wedge \mathsf{B}_B\varphi$. However, the mental operation is not sufficient to guarantee that the robots believe that the person needs helps. Indeed, we have: $\mathtt{MR}^{+r}, w_1 \models \neg\mathsf{B}_A h \wedge \neg\mathsf{B}_B h$.

It is by trying to infer that the person needs help from the fact that she switched on the red light, represented by $\vdash(r,h)$, that the robots can form this explicit belief. The mental operation $\vdash(r,h)$ leads from model $\mathtt{MR}^{+r}$ to the updated model $(\mathtt{MR}^{+r})^{\vdash(r,h)} = (W, (N^{+r})^{\vdash(r,h)}, R_A, R_B, V)$ such that:

$(N^{+r})^{\vdash(r,h)}(A, w_1) = \{\{w_1\}, \{w_1, w_2\}\}$, $(N^{+r})^{\vdash(r,h)}(A, w_2) = \{\{w_1\}, \{w_1, w_2\}\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_3) = \{\{w_1\}, \{w_1, w_2\}\}$, $(N^{+r})^{\vdash(r,h)}(A, w_4) = \{\{w_1\}, \{w_1, w_2\}\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_5) = \{\{w_1\}, \{w_1, w_2\}\}$, $(N^{+r})^{\vdash(r,h)}(A, w_6) = \emptyset$,
$(N^{+r})^{\vdash(r,h)}(A, w_7) = \emptyset$, $(N^{+r})^{\vdash(r,h)}(A, w_8) = \{\{w_8, w_9\}\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_9) = \{\{w_8, w_9\}\}$, $(N^{+r})^{\vdash(r,h)}(B, w_1) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_2) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$, $(N^{+r})^{\vdash(r,h)}(B, w_3) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_4) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$, $(N^{+r})^{\vdash(r,h)}(B, w_5) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_6) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$, $(N^{+r})^{\vdash(r,h)}(B, w_7) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_8) = \{\{w_8, w_9\}\}$, $(N^{+r})^{\vdash(r,h)}(B, w_9) = \{\{w_8, w_9\}\}$.

It is easy to check that in the new situation, after the mental operation $\vdash(r,h)$ has been executed, the two robots explicitly believe that the person needs help. That is: $(\mathtt{MR}^{+r})^{\vdash(r,h)}, w_1 \models \mathsf{B}_{\{A,B\}}h$. To sum up, we have that the following holds:

$$\mathtt{MR}, w_1 \models [+r]\mathsf{B}_{\{A,B\}}r \wedge [+r](\neg\mathsf{B}_A h \wedge \neg\mathsf{B}_B h) \wedge [+r][\vdash(r,h)]\mathsf{B}_{\{A,B\}}h.$$

It is just routine to check that the mental operations $+g$ and $\vdash(g,b)$ are also sufficient for robot A to start to believe b explicitly after performing them, but they are not sufficient for robot B since he does not have background knowledge that g implies b. In formal terms, we have:

$$\mathtt{MR}, w_1 \models [+g][\vdash(g,b)]\mathsf{B}_A b, \qquad \mathtt{MR}, w_1 \models [+g][\vdash(g,b)]\neg\mathsf{B}_B b.$$

3.5. Discussion: uncertainty about mental operations of others

In the previous semantics of events $(G{:}\alpha)_H$, we assumed that if agents of a group G perform a mental operation, then every agent in H *knows* this and agents in H have common knowledge about this fact. In certain situations, it would be useful to relax this assumption by representing the fact that an agent i may *be uncertain* that another agent j has performed a certain mental operation. In particular, we would like to represent the following situation:

> Agent j performs the mental operation α, while agent i envisages that j could have possibly performed it and that j has performed no mental operation at all.

Our discussion could be generalized to the situation in which agent i envisages a *finite* set Σ of mental operations that j could have possibly performed. For simplicity of exposition, we only consider the case in which Σ is a singleton.[7]

[7] Notice that it would not make sense to consider the situation in which, for every mental operation α, agent i considers it possible both that agent j has performed α and that j

In order to represent the previous situation in a formal way, we discuss here the semantics of a new type of events and corresponding dynamic operators of the form $[?_{i,j}\alpha]$ with $i \neq j$ where $?_{i,j}\alpha$ represents the event of agent i questioning whether agent j has performed the mental operation α.

When the event $?_{i,j}\alpha$ takes place, at the same time, (i) agent i considers it possible that agent j has performed the mental operation α and (ii) agent i considers it possible that agent j has performed no mental operation at all. In other words, agent i is uncertain whether the mental operation α has been performed by agent j.

Truth conditions of formulas $[?_{i,j}\alpha]\varphi$ are defined as follows:

$$M, w \models [?_{i,j}\alpha]\varphi \iff M^{?_{i,j}\alpha}, w_0 \models \varphi$$

$M^{?_{i,j}\alpha} = (W^{?_{i,j}\alpha}, N^{?_{i,j}\alpha}, R_1^{?_{i,j}\alpha}, \ldots, R_n^{?_{i,j}\alpha}, V^{?_{i,j}\alpha})$ is the updated model such that for all $k \in Agt$ and for all $w \in W$:

$$\begin{aligned}
W^{?_{i,j}\alpha} &= W \times \{0,1\} \\
N^{?_{i,j}\alpha}(j, w_0) &= \{X \times \{0\} : X \in N^{\alpha}(j,w)\} \\
N^{?_{i,j}\alpha}(j, w_1) &= \{X \times \{1\} : X \in N(j,w)\} \\
N^{?_{i,j}\alpha}(i, w_0) &= \{X \times \{0,1\} : X \in N(i,w)\} \\
N^{?_{i,j}\alpha}(i, w_1) &= \{X \times \{0,1\} : X \in N(i,w)\} \\
N^{?_{i,j}\alpha}(k, w_0) &= \{X \times \{1\} : X \in N(k,w)\} \text{ if } k \neq i \text{ and } k \neq j \\
N^{?_{i,j}\alpha}(k, w_1) &= \{X \times \{1\} : X \in N(k,w)\} \text{ if } k \neq i \text{ and } k \neq j \\
R_i^{?_{i,j}\alpha} &= \{(w_x, v_y) : x, y \in \{0,1\} \text{ and } wR_iv\} \\
R_j^{?_{i,j}\alpha} &= \{(w_0, v_0) : wR_jv\} \cup \{(w_1, v_1) : wR_jv\} \\
R_k^{?_{i,j}\alpha} &= \{(w_0, v_1) : wR_kv\} \cup \{(w_1, v_1) : wR_kv\} \text{ if } k \neq i \text{ and } k \neq j \\
V^{?_{i,j}\alpha}(w_x) &= V(w) \text{ with } x \in \{0,1\}
\end{aligned}$$

The following proposition guarantees that the semantics of the operator $[?_{i,j}\alpha]$ with respect to the class of LEK models is well defined:

PROPOSITION 2. *If M is a LEK model then $M^{?_{i,j}\alpha}$ is also a LEK model.*

---

has not performed α, as this is incompatible with the concept of resource-bounded agent. Indeed, there exists an infinite number of possible mental operations α. Furthermore, by definition, a resource-bounded agent is able to represent in her mind and to imagine an finite number of events that might occur at a given time but is unable to imagine an infinite number of events.

Notice that, in a way similar to the semantics of the events $(G{:}\alpha)_H$, in the definition of the updated model $M^{?_{i,j}\alpha}$, there are two copies of the original model. In the copy 0, agent j performs the mental operation α whereas in the copy 1 agent j does nothing. Agent i cannot distinguish copy 0 from copy 1. This is the reason why i is uncertain about the performance of the mental operation α by agent j. All agents different from i and j are not aware of the fact that j has performed the mental operation α, as they only envisage the copy 1 of the original model.

4. Axiomatization

Let us now present sound and complete axiomatizations for the logics LEK, LEK^+ and their dynamic extensions DLEK, DLEK^+. For a logic λ, we write $\lambda \vdash \varphi$ to denote the fact that φ is a theorem of λ, i.e., that φ belongs to the least set containing all axioms of λ and closed under the rules of λ.

Definition 3. *We define LEK^+ to be the extension of classical propositional logic given by the following rules and axioms:*

$(\mathbf{K}_{\mathsf{K}_i})$	$(\mathsf{K}_i\varphi \land \mathsf{K}_i(\varphi \to \psi)) \to \mathsf{K}_i\psi$
$(\mathbf{T}_{\mathsf{K}_i})$	$\mathsf{K}_i\varphi \to \varphi$
$(\mathbf{D}_{\mathsf{K}_i})$	$\neg\mathsf{K}_i(\varphi \land \neg\varphi)$
$(\mathbf{4}_{\mathsf{K}_i})$	$\mathsf{K}_i\varphi \to \mathsf{K}_i\mathsf{K}_i\varphi$
$(\mathbf{5}_{\mathsf{K}_i})$	$\neg\mathsf{K}_i\varphi \to \mathsf{K}_i\neg\mathsf{K}_i\varphi$
$(\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i})$	$(\mathsf{B}_i\varphi \land \mathsf{K}_i(\varphi \leftrightarrow \psi)) \to \mathsf{B}_i\psi$
$(\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{B}_i})$	$\mathsf{B}_i\varphi \to \mathsf{K}_i\mathsf{B}_i\varphi$
$(\mathbf{Nec}_{\mathsf{K}_i})$	$\dfrac{\varphi}{\mathsf{K}_i\varphi}$

The logic LEK is the logic obtained by removing $(\mathbf{T}_{\mathsf{K}_i})$.

It is straightforward to check that all axioms are valid and all rules preserve validity, from which we obtain:

Lemma 1. *The logic $\mathsf{LEK}^{(+)}$ is sound for the class of $\mathsf{LEK}^{(+)}$ models.*

Observe that $(\mathbf{D}_{\mathsf{K}_i})$ follows from $(\mathbf{T}_{\mathsf{K}_i})$, but the former is needed in LEK. Most of the axioms are familiar from modal and epistemic logic, with the possible exceptions of $(\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i})$ and $(\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{B}_i})$. The latter simply corresponds to $(\mathbf{C2})$, which states that the neighborhood relation does not vary within $R_i(w)$. Regarding $(\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i})$, observe that if

$\varphi \leftrightarrow \psi$ were valid, then so would $\mathsf{B}_i\varphi \rightarrow \mathsf{B}_i\psi$, as then $||\varphi||^M_{i,w} = ||\psi||^M_{i,w}$ regardless of M or w. However, as $||\varphi||^M_{i,w}, ||\psi||^M_{i,w} \subseteq R_i(w)$, it suffices that $||\varphi||^M_{i,w} \cap R_i(w) = ||\psi||^M_{i,w} \cap R_i(w)$, i.e., that $\mathsf{K}_i(\varphi \leftrightarrow \psi)$ is true. Note moreover that $\mathsf{K}_i(\varphi \rightarrow \psi)$ is *not* sufficient for $\mathsf{B}_i\varphi \rightarrow \mathsf{B}_i\psi$ to hold, as the neighborhood relation is not assumed monotone.

The axiomatics of the logic $\mathsf{DLEK}^{(+)}$ includes all principles of the logic $\mathsf{LEK}^{(+)}$ *plus* a set of reduction axioms and the rule of replacement of equivalents.

Definition 4. *We define $\mathsf{DLEK}^{(+)}$ to be the extension of $\mathsf{LEK}^{(+)}$ generated by the following axioms and rule of inference, where $G \subseteq H \subseteq Agt$, $i \in G$, $k \in H \setminus G$ and $j \notin H$:*

(**Red1**) $[(G{:}\alpha)_H]p \leftrightarrow p$

(**Red2**) $[(G{:}\alpha)_H]\neg\varphi \leftrightarrow \neg[(G{:}\alpha)_H]\varphi$

(**Red3**) $[(G{:}\alpha)_H](\varphi \wedge \psi) \leftrightarrow ([(G{:}\alpha)_H]\varphi \wedge [(G{:}\alpha)_H]\psi)$

(**Red4**) $[(G{:}\alpha)_H]\mathsf{K}_i\varphi \leftrightarrow \mathsf{K}_i[(G{:}\alpha)_H]\varphi$

(**Red5**) $[(G{:}+\varphi)_H]\mathsf{B}_i\psi \leftrightarrow (\mathsf{B}_i[(G{:}+\varphi)_H]\psi \vee \mathsf{K}_i([(G{:}+\varphi)_H]\psi \leftrightarrow \varphi))$

(**Red6**) $[(G{:}-\varphi)_H]\mathsf{B}_i\psi \leftrightarrow (\mathsf{B}_i[(G{:}-\varphi)_H]\psi \wedge \neg\mathsf{K}_i([(G{:}-\varphi)_H]\psi \leftrightarrow \varphi))$

(**Red7**) $$[(G{:}\vdash(\varphi,\psi))_H]\mathsf{B}_i\chi \leftrightarrow \Big(\mathsf{B}_i[(G{:}\vdash(\varphi,\psi))_H]\chi \vee \big((\mathsf{B}_i\varphi \wedge \mathsf{K}_i(\varphi \rightarrow \psi)) \wedge \mathsf{K}_i[(G{:}\vdash(\varphi,\psi))_H]\chi \leftrightarrow \psi)\big)\Big)$$

(**Red8**) $$[(G{:}\cap(\varphi,\psi))_H]\mathsf{B}_i\chi \leftrightarrow \Big(\mathsf{B}_i[(G{:}\cap(\varphi,\psi))_H]\chi \vee \big((\mathsf{B}_i\varphi \wedge \mathsf{B}_i\psi) \wedge \mathsf{K}_i([(G{:}\cap(\varphi,\psi))_H]\chi \leftrightarrow (\varphi \wedge \psi))\big)\Big)$$

(**Red9**) $[(G{:}\alpha)_H]\mathsf{K}_k\varphi \leftrightarrow \mathsf{K}_k[(G{:}\alpha)_H]\varphi$

(**Red10**) $[(G{:}\alpha)_H]\mathsf{B}_k\varphi \leftrightarrow \mathsf{B}_k[(G{:}\alpha)_H]\varphi$

(**Red11**) $[(G{:}\alpha)_H]\mathsf{K}_j\varphi \leftrightarrow \mathsf{K}_j\varphi$

(**Red12**) $[(G{:}\alpha)_H]\mathsf{B}_j\varphi \leftrightarrow \mathsf{B}_j\varphi$

and the following rule of inference:

(**RRE**) $$\frac{\psi_1 \leftrightarrow \psi_2}{\varphi \leftrightarrow \varphi[\psi_1/\psi_2]}$$

The first four axioms are standard from dynamic epistemic logic. Meanwhile, for $i \in G$ the intuition behind all of the (**Red5**))-((**Red8**) axioms is similar; consider, for example, (**Red5**). The left-hand side states that, after adding φ as a neighborhood, the agent will believe ψ, i.e. the extension of ψ within $R_i(w)$ will be a neighborhood in the new model. This can happen for precisely two reasons: either the extension of ψ in the new model, i.e. $[(G\colon + \varphi)_H]\psi$, already *was* a neighborhood, in which case $\mathsf{B}_i[(G\colon + \varphi)_H]\psi$ holds; or, this extension is precisely the neighborhood that was added, i.e. $\mathsf{K}_i([(G\colon + \varphi)_H]\psi \leftrightarrow \varphi)$ holds.

For (**Red9**) and (**Red10**) the intuition is that agents in $H \setminus G$ have not performed the mental operation α but are aware of the fact that mental operation α has been performed by the agents in G.

For (**Red11**) and (**Red12**) the intuition is that, for agents not in H, the set of possible worlds is *exactly the same* as it was before those in G realized the private mental operation. With this in mind, it is straightforward to check that all axioms are valid and all rules preserve validity in the class of $\mathsf{LEK}^{(+)}$-models, from which the following is an immediate consequence:

LEMMA 2. *The logic DLEK is sound for the class of LEK-models, and the logic DLEK^+ is sound for the class of single-agent LEK^+ models.*

Our goal now is to prove that LEK and LEK^+ are strongly complete for their intended semantics. We will achieve this by a fairly standard canonical-model argument, although the neighborhood structure will require some care.

DEFINITION 5 (canonical λ-model). *For $\lambda \in \{\mathsf{LEK}, \mathsf{LEK}^+\}$, we define the canonical λ-model $M^\lambda = (W^\lambda, N^\lambda, R_1^\lambda \ldots, R_n^\lambda, V^\lambda)$, where:*

- *W^λ is the set of all maximal λ-consistent subsets of $\mathcal{L}_{\mathsf{LEK}}$.*
- *$wR_i^\lambda v$ if and only if, for all formulas φ, $\mathsf{K}_i\varphi \in w$ implies that $\varphi \in v$.*
- *In order to define N^λ, for $w \in W$ and $\varphi \in \mathcal{L}_{\mathsf{LEK}}$, first define $A_\varphi(i, w) = \{v \in R_i^\lambda(w) : \varphi \in v\}$. Then, define N^λ by letting $N^\lambda(i, w) = \{A_\varphi(i, w) : \mathsf{B}_i\varphi \in w\}$.*
- *Finally, we define the valuation V^λ by $w \in V^\lambda(p)$ if and only if $p \in w$.*

The following is standard and we omit the proof:

LEMMA 3. *The structure M^λ defined above is a λ-model. Moreover, if $w \in W^\lambda$ and $\varphi \in \mathcal{L}_{\mathsf{LEK}}$, then*

1. *$\mathsf{K}_i\varphi \in w$ if and only if, for every v such that $wR_i^\lambda v$, $\varphi \in v$, and*
2. *if $wR_i^\lambda v$ and $\mathsf{B}_i\varphi \in w$, then $\mathsf{B}_i\varphi \in v$.*

We also need to prove that M^λ has a somewhat less familiar property. This will be used later in the truth lemma, for the case of B_i.

LEMMA 4. *For every $w \in W^\lambda$ and $\mathsf{B}_i\varphi, \mathsf{B}_i\psi \in \mathcal{L}_{\mathsf{LEK}}$, if $\mathsf{B}_i\varphi \in w$ but $\mathsf{B}_i\psi \notin w$, it follows that there is $v \in R_i^\lambda(w)$ such that either $\varphi \in v$ but $\neg\psi \in v$, or $\neg\varphi \in v$ but $\psi \in v$.*

PROOF. Let $w \in W^\lambda$ and φ, ψ be such that $\mathsf{B}_i\varphi \in w$, $\mathsf{B}_i\psi \notin w$. Towards a contradiction, assume that for every $v \in R_i^\lambda(w)$, either $\varphi, \psi \in v$ or $\neg\varphi, \neg\psi \in v$; then, it follows from Lemma 3 that $\mathsf{K}_i(\varphi \leftrightarrow \psi) \in w$, so that by Axiom ($\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i}$), $\mathsf{B}_i\psi \in w$, contrary to our assumption. ■

With this, we may state and prove our version of the Truth Lemma:

LEMMA 5. *For every $\varphi \in \mathcal{L}_{\mathsf{LEK}}$ and every $w \in W^\lambda$, $\varphi \in w$ if and only if $M^\lambda, w \models \varphi$.*

PROOF. The proof proceeds by a standard induction on the construction of φ. All cases are routine except $\varphi = \mathsf{B}_i\psi$.

First assume that $\mathsf{B}_i\psi \in w$. Then, $A_\psi(i, w) \in N^\lambda(i, w)$. But, $A_\psi(i, w) = \{v \in R_i^\lambda(w) : \psi \in v\} \overset{\text{IH}}{=} \|\psi\| \cap R_i^\lambda(w)$. Thus, $M^\lambda, w \models \mathsf{B}_i\psi$.

Now, suppose $\mathsf{B}_i\psi \notin w$, so that $\neg\mathsf{B}_i\psi \in w$. We must check that $\|\psi\| \cap R_i^\lambda(w) \notin N(i, w)$. Choose an arbitrary set $A \in N(i, w)$; by definition, $A = A_\theta(i, w)$ for some θ with $\mathsf{B}_i\theta \in w$. By Lemma 4, there is some $v \in R_i^\lambda(w)$ such that $\psi, \neg\theta \in v$ or $\neg\psi, \theta \in v$; in the first case, this shows using the induction hypothesis that $v \in (\|\psi\| \cap R_i^\lambda(w)) \setminus A_\theta(i, w)$, while in the second we obtain $v \in A_\theta(i, w) \setminus (\|\psi\| \cap R_i^\lambda(w))$. In either case we obtain $A_\theta(i, w) \neq \|\psi\| \cap R_i^\lambda(w)$, and since $A = A_\theta(i, w)$ was an arbitrary element of $N^\lambda(i, w)$, we conclude that $\|\psi\| \cap R_i^\lambda(w) \notin N^\lambda(i, w)$ and thus $M^\lambda, w \not\models \mathsf{B}_i\psi$. ■

We are now ready to prove that the static logics are strongly complete.

THEOREM 1. *For $\lambda \in \{\mathsf{LEK}, \mathsf{LEK}^+\}$, λ is strongly complete for the class of λ-models.*

PROOF. Any consistent set of formulas Φ may be extended to a maximal consistent set of formulas $w_* \in W^\lambda$, and $M^\lambda, w_* \models \Phi$ by Lemma 5. ■

The strong completeness of the dynamic logics follows from this result, in view of the fact that the reduction axioms may be used to find, for any formula, a provably equivalent formula in the static fragment.

LEMMA 6. *If φ is any formula of $\mathcal{L}_{\mathsf{DLEK}}$, there is a formula $\widetilde{\varphi}$ in $\mathcal{L}_{\mathsf{LEK}}$ such that $\mathsf{DLEK} \vdash \varphi \leftrightarrow \widetilde{\varphi}$.*

Proof. This follows by a routine induction on φ using the reduction axioms and the rule of replacement of equivalents (**RRE**) from Definition 4. ■

As a corollary, we get the following:

Theorem 2. *DLEK is strongly complete for the class of LEK models and mono-agent $\mathsf{DLEK}^{(+)}$ is strongly complete for the class of $\mathsf{LEK}^{(+)}$ models.*

Proof. If Γ is a $\mathsf{DLEK}^{(+)}$-consistent set of $\mathcal{L}_{\mathsf{DLEK}}$ formulas, then $\widetilde{\Gamma} = \{\widetilde{\varphi} : \varphi \in \Gamma\}$ is a $\mathsf{LEK}^{(+)}$-consistent set of $\mathcal{L}_{\mathsf{LEK}}$ formulas (since $\mathsf{DLEK}^{(+)}$ is an extension of $\mathsf{LEK}^{(+)}$), and hence by Theorem 1, there is a $\mathsf{LEK}^{(+)}$-model M with a world w such that $M, w \models \widetilde{\Gamma}$. But, since $\mathsf{DLEK}^{(+)}$ is sound and for each $\varphi \in \Gamma$, $\mathsf{DLEK}^{(+)} \vdash \varphi \leftrightarrow \widetilde{\varphi}$, it follows that $M, w \models \Gamma$. ■

Thus our logics are strongly complete, but the construction we have given will in general produce infinite models. In the next section, we will consider the complexity of the satisfiability problem.

5. Complexity

We study the computability of the satisfiability problem of LEK^{+}: given a formula, determine whether it is satisfiable. Simple changes in the arguments presented below would allow the reader to adapt our line of reasoning to LEK. We will also provide a decidability result of the satisfiability problem of DLEK. We first consider the single-agent case and move then to the multi-agent case. Below, $card(A)$ denotes the cardinality of the set A.

5.1. Mono-agent case

Assume $card(Agt) = 1$. Let φ be a satisfiable formula. Let $M = (W, N, V)$ be a model and $w \in W$ be such that $M, w \models \varphi$; note that in the mono-agent case we can assume that R is the total relation, so we omit it. Let $\mathsf{K}\psi_1, \dots, \mathsf{K}\psi_m$ and $\mathsf{B}\chi_1, \dots, \mathsf{B}\chi_n$ be lists of all subformulas of φ of the form $\mathsf{K}\psi$ and $\mathsf{B}\chi$. Let $\hat{\mathsf{K}} = \{i : 1 \leq i \leq m \ \&\ M, w \not\models \mathsf{K}\psi_i\}$, $\hat{\mathsf{B}}^{+} = \{j : 1 \leq j \leq n \ \&\ M, w \models \mathsf{B}\chi_j\}$ and $\hat{\mathsf{B}}^{-} = \{k : 1 \leq k \leq n \ \&\ M, w \not\models \mathsf{B}\chi_k\}$. For all $i \in \hat{\mathsf{K}}$, let $v_i \in W$ be such that $M, v_i \not\models \psi_i$. Such v_i exists because $M, w \not\models \mathsf{K}\psi_i$. For all $j \in \hat{\mathsf{B}}^{+}$, let $A_j \in N$ be such that $A_j = \{v \in W : M, v \models \chi_j\}$. Such A_j exists because $M, w \models \mathsf{B}\chi_j$. For all $j \in \hat{\mathsf{B}}^{+}$ and for all $k \in \hat{\mathsf{B}}^{-}$, let $u_{j,k} \in W$ be such that $M, u_{j,k} \not\models \chi_j \leftrightarrow \chi_k$. Such $u_{j,k}$ exists because $M, w \models \mathsf{B}\chi_j$ and $M, w \not\models \mathsf{B}\chi_k$. Let $M' = (W', N', V')$ be the model defined by:

- $W' = \{w\} \cup \{v_i : i \in \hat{\mathsf{K}}\} \cup \{u_{j,k} : j \in \hat{\mathsf{B}}^+ \ \& \ k \in \hat{\mathsf{B}}^-\}$,
- $N' = \{A_j \cap W' : j \in \hat{\mathsf{B}}^+\}$,
- for all $p \in Atm$, $V'(p) = V(p) \cap W'$.

Obviously, $card(W')$ and $card(N')$ are polynomial in the size of φ. Let $\Phi = SF(\varphi)$, the closure under single negations of the set of all of φ's subformulas.

Lemma 7. *Let ψ be a formula. If $\psi \in \Phi$ then for all $s \in W'$, $M, s \models \psi$ iff $M', s \models \psi$.*

Proof. By induction on ψ. We only consider the cases $\psi = \mathsf{B}\chi$.
Suppose $M, s \models \mathsf{B}\chi$ and $M', s \not\models \mathsf{B}\chi$. Let $j \in \hat{\mathsf{B}}^+$ be such that $\mathsf{B}\chi = \mathsf{B}\chi_j$. Hence, $A_j = \{t \in W : M, t \models \chi\}$. By induction hypothesis, $A_j \cap W' = \{t \in W' : M', t \models \chi\}$. Thus, $M', s \models \mathsf{B}\chi$: a contradiction.
Suppose $M', s \models \mathsf{B}\chi$ and $M, s \not\models \mathsf{B}\chi$. Let $j \in \hat{\mathsf{B}}^+$ be such that $A_j \cap W' = \{t \in W' : M', t \models \chi\}$. By induction hypothesis, $A_j \cap W' = \{t \in W' : M, t \models \chi\}$. Let $k \in \hat{\mathsf{B}}^-$ be such that $\mathsf{B}\chi = \mathsf{B}\chi_k$. Remember that $A_j = \{t \in W : M, t \models \chi_j\}$. Moreover, $u_{j,k} \in W'$ is such that $M, u_{j,k} \models \chi_j$ and $M, u_{j,k} \not\models \chi_k$, or $M, u_{j,k} \not\models \chi_j$ and $M, u_{j,k} \models \chi_k$. In the former case, $u_{j,k} \in A_j \cap W'$. Hence, $M, u_{j,k} \models \chi$: a contradiction. In the latter case, $u_{j,k} \in A_j \cap W'$. Thus, $M, u_{j,k} \models \chi_j$: a contradiction. ∎

Theorem 3. *If $card(Agt) = 1$ then satisfiability problem of LEK^+ is NP-complete.*

Proof. Membership in NP follows from Lemma 7. NP-hardness follows from the NP-hardness of classical propositional logic. ∎

5.2. Multi-agent case

Our study of the computability in the multi-agent case will be based on the modal tableaux approach developed by Halpern and Moses [23]. Assume $card(Agt) \geq 2$.
A set $\mathcal{T}$ of formulas is said to be fully expanded if for all formulas φ in $\mathcal{T}$ and for all formulas ψ in $SF(\varphi)$, either $\psi \in \mathcal{T}$, or $\neg\psi \in \mathcal{T}$. A propositional tableau is a set $\mathcal{T}$ of formulas such that: (i) for all formulas φ, if $\neg\neg\varphi \in \mathcal{T}$ then $\varphi \in \mathcal{T}$; (ii) for all formulas φ, ψ, if $\neg(\varphi \vee \psi) \in \mathcal{T}$ then $\neg\varphi \in \mathcal{T}$ and $\neg\psi \in \mathcal{T}$; (iii) for all formulas φ, ψ, if $\varphi \vee \psi \in \mathcal{T}$ then either $\varphi \in \mathcal{T}$, or $\psi \in \mathcal{T}$. A propositional tableau $\mathcal{T}$ is said to be blatantly consistent iff for all formulas φ, either $\varphi \notin \mathcal{T}$, or $\neg\varphi \notin \mathcal{T}$.

A modal tableau is a structure of the form $\mathcal{T} = (W, R, L)$ where W is a nonempty set of states (with typical members denoted w, v, etc), R is a function associating a binary relation R_i on W to each $i \in Agt$ and L is a function assigning to each $w \in W$ a blatantly consistent and fully expanded propositional tableau $L(w)$ such that for all $w \in W$: (i) for all formulas φ, if $\neg \mathsf{K}_i \varphi \in L(w)$ then there exists $v \in R_i(w)$ such that $\neg \varphi \in L(v)$; (ii) for all formulas φ, if $\mathsf{B}_i \varphi \in L(w)$ then for all formulas ψ, if $\neg \mathsf{B}_i \psi \in L(w)$ then there exists $v \in R_i(w)$ such that either $\varphi \in L(v)$ and $\neg \psi \in L(v)$, or $\neg \varphi \in L(v)$ and $\psi \in L(v)$; (iii) for all formulas φ, if $\mathsf{K}_i \varphi \in L(w)$ then for all $v \in (R_i \cup R_i^{-1})^\star(w)$, $\varphi \in L(v)$ and $\mathsf{K}_i \varphi \in L(v)$; (iv) for all formulas φ, if $\mathsf{B}_i \varphi \in L(w)$ then for all $v \in (R_i \cup R_i^{-1})^\star(w)$, $\mathsf{B}_i \varphi \in L(v)$. For all formulas φ, let $L^{-1}(\varphi) = \{w : w \in W \ \& \ \varphi \in L(w)\}$. We shall say that a modal tableau $\mathcal{T} = (W, R, L)$ is a modal tableau for a formula φ if $L^{-1}(\varphi) \neq \emptyset$.

Given a model $M = (W, R, N, V)$, let $\mathcal{T}' = (W', R', L')$ be defined by: $W' = W$, $R'_i = R_i$ for each $i \in Agt$, L' is the function assigning to each $w \in W'$ the propositional tableau $L'(w) = \{\varphi : M, w \models \varphi\}$. The proof that $\mathcal{T}'$ is a modal tableau is easy.

Proposition 3. *Let φ be a formula. If φ is satisfiable then there exists a modal tableau for φ.*

Given a modal tableau $\mathcal{T} = (W, R, L)$, let $M' = (W', R', N', V')$ be defined by: $W' = W$, $R'_i = (R_i \cup R_i^{-1})^\star$ for each $i \in Agt$, $N'_i(w) = \{(R_i \cup R_i^{-1})^\star(w) \cap L^{-1}(\varphi) : \ \mathsf{B}_i \varphi \in L(w)\}$ for each $w \in W$ and for each $i \in Agt$, V' is the function assigning to each $p \in Atm$ the subset $V'(p) = \{w : w \in W \ \& \ p \in L(w)\}$ of W'. The proof that $\mathcal{T}'$ is a model is easy. Moreover, one can prove by induction on φ that for all $w \in W$, $\varphi \in L(s)$ iff $M', s \models \varphi$.

Proposition 4. *Let φ be a formula. If there exists a modal tableau for φ then φ is satisfiable.*

By Propositions 3 and 4, satisfiability is reducible to the following decision problem (MT): given a formula, determine whether there exists a modal tableau for it. Based on the tools and techniques developed in [23] for ordinary epistemic logics, one can design an algorithm that tries to construct a modal tableau for a given formula. The main properties of such algorithm are:

- For all given formulas, the above algorithm terminates and runs in polynomial space,
- for all given formulas φ, the algorithm returns "there is a modal tableau for φ" iff there is a modal tableau for φ.

Theorem 4. *If $card(Agt) \geq 2$ then satisfiability problem of* LEK *is PSPACE-complete.*

Proof. Membership in PSPACE follows from the above discussion. Meanwhile, PSPACE-hardness follows from the PSPACE-hardness of multi-agent epistemic logic. ■

The reduction axioms and the rule of replacement of equivalents in Definition 4 may be used to give a decision procedure for the satisfiability of DLEK; however, due to exponential blow-up in the size of formulas, this algorithm would no longer remain in PSPACE without modification. Thus we will state only the following:

Corollary 1. *The satisfiability problems of mono-agent* $\mathsf{DLEK}^{(+)}$ *and multi-agent* DLEK *are decidable.*

Proof. Immediate from the decidability of $\mathsf{LEK}^{(+)}$ and the fact that, given a formula φ of $\mathcal{L}_{\mathsf{DLEK}}$, the formula $\widetilde{\varphi}$ is clearly computable from φ. ■

However, we do not believe that such a procedure would be optimal, and indeed conjecture that DLEK is in PSPACE. We leave the computation of its precise complexity for future work.

6. Translation of $\mathcal{L}_{\mathsf{LEK}}$ into an ordinary modal language: the mono-agent case

We will show how one can reduce the semantics of LEK^+ to a purely relational semantics, albeit with additional modalities. For simplicity, we will consider only the mono-agent case, and we will write K, B instead of $\mathsf{K}_1, \mathsf{B}_1$. Simple changes in the arguments presented below would allow the reader to adapt our line of reasoning to LEK. Recall that in the mono-agent case, we may assume that in any LEK^+-model (W, N, R, V) we have that R is the total relation, so we may omit it. Moreover, since $N(w)$ does not depend on $w \in W$ (given that it is assumed R-invariant), we may identify N with the set of neighbourhoods $N(w_0)$, where $w_0 \in W$ is arbitrary.

The intuition behind our reduction is as follows. Suppose that a mono-agent model $M = (W, N, V)$ is given. Then, one can think of M as being a substructure of a bigger model M', whose domain is $W \cup N$; that is, we may view the neigbhorhoods themselves as new 'worlds'. In order for M' to contain enough information to interpret the language of LEK^+, we need to include the following relations:

- To interpret $\mathsf{K}\varphi$ at $w \in W$, we include an equivalence relation R_K linking w to all 'real' worlds. It does not matter how R_K behaves on neighbourhoods, so on N we may just let R_K be the identity.
- To interpret $\mathsf{B}\varphi$ at $w \in W$, we include:
  - a relation R_L linking each $w \in W$ to each $U \in N(w)$. Note in this case that every point is related to every neighborhood, so we may simply define $R_L = W \times N$, and
  - a relation $R_\Box$ linking each $U \in N$ to its elements (i.e., $UR_\Box w$ if and only if $w \in U$).

This setup would suffice to interpret monotone neighborhood semantics [17]. However, for our *strict* interpretation, we need some additional structure. Namely, to check if U is a witness for $\mathsf{B}\varphi$, we must check that every element of U satisfies φ, but also that every element *not* in U satisfies $\neg\varphi$. We achieve this by an additional accessibility relation $R_\blacksquare$, where $UR_\blacksquare w$ holds whenever $w \notin U$; a similar trick has been used in [18, 20] to deal with the 'inaccessible worlds' logic of Humberstone [26].

Let us now formalize these ideas. We begin by defining a new modal language, $\mathcal{L}_O$, given by:

$$\varphi, \psi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi \mid L\varphi \mid \Box\varphi \mid \blacksquare\varphi$$

We will use the following abbreviations: $\hat{K}\varphi$ for $\neg K\neg\varphi$, $\hat{L}\varphi$ for $\neg L\neg\varphi$, $\Diamond\varphi$ for $\neg\Box\neg\varphi$ and $\blacklozenge\varphi$ for $\neg\blacksquare\neg\varphi$. Let $\tau : \mathcal{L}_{\mathsf{LEK}} \longrightarrow \mathcal{L}_O$ be the translation function defined by

- $\tau(p) = p$,
- $\tau(\neg\varphi) = \neg\tau(\varphi)$,
- $\tau(\varphi \wedge \psi) = \tau(\varphi) \wedge \tau(\psi)$,
- $\tau(\mathsf{K}\varphi) = K\tau(\varphi)$,
- $\tau(\mathsf{B}\varphi) = \hat{L}(\Box\tau(\varphi) \wedge \blacksquare\neg\tau(\varphi))$.

An $\mathcal{L}_O$-model is a structure of the form $(W, R_K, R_L, R_\Box, R_\blacksquare, V)$ where W is a nonempty set of possible worlds, R_K, R_L, $R_\Box$ and $R_\blacksquare$ are binary relations on W and $V : W \longrightarrow 2^{Atm}$ is a valuation function. The truth-conditions are inductively defined as in ordinary modal logic. We will say that the model $(W, R_K, R_L, R_\Box, R_\blacksquare, V)$ is prestandard iff

- R_K is reflexive, transitive and Euclidean,
- $R_K \circ R_L \subseteq R_L$,
- $R_L \subseteq R_\Box^{-1} \cup R_\blacksquare^{-1}$,
- $R_L \circ R_\Box \subseteq R_K$,
- $R_L \circ R_\blacksquare \subseteq R_K$.

If, furthermore, $R_\Box \cap R_\blacksquare = \emptyset$ then we will say that the prestandard model $(W, R_K, R_L, R_\Box, R_\blacksquare, V)$ is standard. We will say that $\varphi \in \mathcal{L}_O$ is prestandard-valid (respectively, standard-valid), in symbols $\models_{pr} \varphi$ (respectively, $\models_{st} \varphi$), iff φ is true at every possible world of every prestandard (respectively, standard) $\mathcal{L}_O$-model.

LEMMA 8. *Let $M = (W, R_K, R_L, R_\Box, R_\blacksquare, V)$ be a standard $\mathcal{L}_O$-model. Let $s_0 \in W$. Let $M' = (W', N', V')$ be the* LEK^+*-model such that*

- $W' = R_K(s_0)$,
- $N' = \{R_\Box(A) : \ A \in W \ \& \ s_0 R_L A\}$,
- $V' : \ p \mapsto V'(p) = V(p) \cap R_K(s_0)$.

Let $\varphi \in \mathcal{L}_{\mathsf{LEK}}$. For all $s \in W'$, $M, s \models \tau(\varphi)$ iff $M', s \models \varphi$.

PROOF. By induction on φ. ■

LEMMA 9. *Let $M = (W, N, V)$ be a* LEK^+*-model such that, for any $w \in W$, $N(w) \cap W = \varnothing$; this condition is easily obtained if we do not allow the elements of W to be subsets of W. Let $M' = (W', R'_K, R'_L, R'_\Box, R'_\blacksquare, V')$ be the standard $\mathcal{L}_O$-model such that*

- $W' = W \cup N$,
- $R'_K = (W \times W) \cup Id_N$,
- $R'_L = W \times N$,
- $R'_\Box = \{(A, s) : \ A \in N \ \& \ s \in W \ \& \ s \in A\}$,
- $R'_\blacksquare = \{(A, s) : \ A \in N \ \& \ s \in W \ \& \ s \notin A\}$,
- $V' : \ p \mapsto V'(p) = V(p)$.

Let $\varphi \in \mathcal{L}_{\mathsf{LEK}}$. For all $s \in W$, $M, s \models \varphi$ iff $M', s \models \tau(\varphi)$.

PROOF. By induction on φ. ■

PROPOSITION 5. *Let $\varphi \in \mathcal{L}_{\mathsf{LEK}}$. The following conditions are equivalent: 1. $\models_{\mathsf{LEK}^+} \varphi$; 2. $\models_{st} \tau(\varphi)$.*

PROOF. 1. ⇒ 2.) By Lemma 8.
2. ⇒ 1.) By Lemma 9. ■

For those interested by axiomatizing the set of all standard-valid $\mathcal{L}_O$-formulas, let O be the least normal modal logic in $\mathcal{L}_O$ containing the following axioms:

- $K\varphi \to \varphi$
- $K\varphi \to KK\varphi$
- $\hat{K}\varphi \to K\hat{K}\varphi$,
- $L\varphi \to KL\varphi$,
- $\varphi \to L(\Diamond\varphi \vee \blacklozenge\varphi)$,
- $K\varphi \to L\Box\varphi$,
- $K\varphi \to L\blacksquare\varphi$.

PROPOSITION 6. *Let $\varphi \in \mathcal{L}_O$. If φ is O-derivable then $\models_{st} \varphi$.*

Moreover, the above axioms are Sahlqvist formulas and correspond to the elementary properties characterizing prestandard $\mathcal{L}_O$-models. As a result, by Sahlqvist Theorems in ordinary modal logic,

PROPOSITION 7. *Let $\varphi \in \mathcal{L}_O$. If $\models_{pr} \varphi$ then φ is O-derivable.*

LEMMA 10. *Let $M = (W, R_K, R_L, R_\Box, R_\blacksquare, V)$ be a prestandard $\mathcal{L}_O$-model. Let $M' = (W', R'_K, R'_L, R'_\Box, R'_\blacksquare, V')$ be the standard $\mathcal{L}_O$-model such that*

- $W' = W \times \{0, 1\}$,
- $(s, \alpha)R'_K(t, \beta)$ *iff* sR_Kt,
- $(s, \alpha)R'_L(t, \beta)$ *iff* sR_Lt,
- $(s, \alpha)R'_\Box(t, \beta)$ *iff* $sR_\Box t$ & $(sR_\blacksquare t \Rightarrow \beta = 0)$,
- $(s, \alpha)R'_\blacksquare(t, \beta)$ *iff* $sR_\blacksquare t$ & $(sR_\Box t \Rightarrow \beta = 1)$,
- $V' : \ p \mapsto V'(p) = V(p) \times \{0, 1\}$.

Let $f : W' \longrightarrow W$ be such that $f(s, \alpha) = s$. f is a surjective bounded morphism from M' to M.

PROOF. Left to the reader. ■

PROPOSITION 8. *Let $\varphi \in \mathcal{L}_O$. The following conditions are equivalent: 1. φ is O-derivable; 2. $\models_{st} \varphi$; 3. $\models_{pr} \varphi$.*

PROOF. (1. $\Rightarrow$ 2.) By Proposition 6. (2. $\Rightarrow$ 3.) By Lemma 10 and the Bounded Morphism Lemma in ordinary modal logic. (3. $\Rightarrow$ 1.) By Proposition 7. ■

7. Conclusion

Let's take stock. In the paper we have introduced DLEK, a logical theory of belief dynamics for resource-bounded agents inspired by existing psychological theories of human memory. We have provided decidability and complexity results for DLEK as well as for its static fragment LEK. In addition, we have introduced the logics DLEK^+, LEK^+, which are suitable for the mono-agent setting.

Directions of future research are manifold. On the conceptual level, we plan to complete the conceptual framework described in Section 2 by extending the family of mental operations with operations of *storage* of information in long-term memory. We believe that these kinds of operations can be modelled as special kinds of epistemic actions in the DEL sense. Specifically, a storage operation modifies an agent's background knowledge by restricting the epistemic relation R_i to worlds in which the information φ to be stored is true, under the condition that this information is already available in the agent's working memory and explicitly believed by the agent.

In the paper, we have adopted a *normative* view of the notion of background knowledge, as the latter is assumed to satisfy 'omniscience' principles such as closure under logical consequence. In future work, we plan to provide a *descriptive* analysis of background knowledge by enriching our logical framework with the notion of *degree of activation* used in the context of computational models of human cognition such as ACT-R [33]. The idea is that every piece of information in an agent's background knowledge is associated with a certain degree of activation: the higher the degree of activation, the higher the availability of the information in the agent's long-term memory. Furthermore, an agent can retrieve from her long-term memory only the background knowledge whose degree of activation is equal to or higher than a certain threshold θ. From this perspective, it is not necessarily the case that, after having stored a formula φ in her background knowledge, an agent can retrieve all logical consequences of φ and her previous background knowledge. Indeed, some of these logical consequences may be inactive, i.e., their degrees of activation may be below the threshold θ.

On the technical level, as emphasized in Section 5, we plan to obtain a result about complexity of the satisfiability problem of DLEK. In particular, we plan to prove that this problem is PSPACE-complete by appropriately adapting to our framework the technique proposed by Lutz [34] for studying complexity of the satisfiability problem of public announcement logic (PAL).

Acknowledgements. David Fernández-Duque's work was partially supported by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-0002-02.

References

[1] T. Ågotnes and N. Alechina. A logic for reasoning about knowledge of unawareness. *Journal of Logic, Language and Information*, 23(2):197–217, 2014.

[2] T. Ågotnes and N. Alechina. Semantics for dynamic syntactic epistemic logics. In *Proceedings of the Tenth International Conference on Knowledge Representation and Reasoning (KR 2006)*, pages 411–419. AAAI Press, 2006.

[3] T. Ågotnes, P. Balbiani, H. van Ditmarsch, and P. Seban. Group announcement logic. *Journal of Applied Logic*, 8(1):62–81, 2014.

[4] T. Ågotnes and M. Walicki. Syntactic knowledge: A logic of reasoning, communication and cooperation. *Proceedings of the Second European Workshop on Multi-Agent Systems (EUMAS 2004)*, 2004.

[5] S. N. Artëmov. The logic of justification. *Rewiev of Symbolic Logic* , 1(4):477–513, 2008.

[6] N. Alechina, B. Logan, and M. Whitsey. A complete and decidable logic for resource-bounded agents. In *Proceedings of AAMAS 2014*, pages 606–613. IEEE Computer Society, 2004.

[7] A. D. Baddeley and G. Hitch. Working memory. In G. H. Bower, editor, *The psychology of learning and motivation: Advances in research and theory*, pages 47–89. Academic Press, 1974.

[8] A. Baltag and L. S. Moss. Logics for Epistemic Programs. *Synthese*, 139(2):165–224, 2004.

[9] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In *Proceedings of LOFT 7*, pages 13–60. Amsterdam University Press, 2008.

[10] N Cowan. An embedded process model of working memory. In A. Miyake and P. Shah, editors, *Models of working memory: Mechanisms of Active Maintenance and Executive Control*, pages 62–101. Cambridge University Press, 1999.

[11] H. N. Duc. Reasoning about rational, but not logically omniscient, agents. *Journal of Logic and Computation*, 7(5):633–648, 1997.

[12] J. Elgot-Drapkin, M. Miller, and D. Perlis. Life on a desert island: Ongoing work in real-time reasoning. In *Proceedings of the 1987 workshop on The Frame Problem*, pages 349–357. Morgan Kaufmann, 1987.

[13] J. Elgot-Drapkin, M. Miller, and D. Perlis. Memory, reason and time: the step-logic approach. In R. Cummins and J. L. Pollock, editors, *Philosophy and AI: Essays at the Interface*, MIT Press, 1991.

[14] J. Elgot-Drapkin, S. Kraus, M. Miller, M. Nirkhe, and D. Perlis. Active logics: A unified formal approach to episodic reasoning. *Technical Report CS-TR-4072*, 1999.

[15] R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76, 1988.

[16] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge.* MIT Press, 1995.

[17] O. Gasquet and A. Herzig. Translation-based deduction methods for modal logics. In *Proceedings of the 5th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'94)*, pages 399–408. Springer, 1995.

[18] O. Gasquet and A. Herzig. Translating inaccessible worlds logic into bimodal logic. In *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU '93)*, pages 145–150. Springer, 1993.

[19] A. Goldman. Theory of Mind. In E. Margolis, R. Samuels, and S. Stich, editors, *Oxford Handbook of Philosophy and Cognitive Science*, Oxford University Press, 2012.

[20] V. Goranko. Completeness and Incompleteness in the Bimodal Base $\mathcal{L}(R, -R)$. In *Mathematical Logic*, pages 311–326. Springer, 1990.

[21] J. Grant, S. Kraus, and D. Perlis. A logic for characterizing multiple bounded agents. *Autonomous Agents and Multi-Agent Systems*, 3(4):351–387, 2000.

[22] D. Grossi and F. R. Velázquez-Quesada. Twelve Angry Men: A study on the fine-grain of announcements. In *Proceedings of LORI 2009*, volume 5834 of *LNCS*, pages 147–160. Springer, 2009.

[23] J. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[24] J. Y. Halpern and L. C. Rêgo. Reasoning about knowledge of unawareness. *Games and Economic Behavior*, 67(2):503–525, 2009.

[25] J. Hintikka. Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4:475–484, 1975.

[26] I. L. Humberstone. Inaccessible worlds. *Notre Dame Journal of Formal Logic*, 24(3):346–352, 1983.

[27] D. Kahneman and A. Tversky. Variants of uncertainty. *Cognition*, 11:143–157, 1982.

[28] K. Konolige. What Awareness Isn't: A Sentential View of Implicit and Explicit Belief. In *Proceedings of the 1st Conference on Theoretical Aspects of Reasoning about Knowledge (TARK'86)*, pages 241–250. Morgan Kaufmann, 1986.

[29] M. Jago. Epistemic Logic for Rule-Based Agents. *Journal of Logic, Language and Information*, 18(1):131–158, 2009.

[30] J. E. Laird. *The Soar Cognitive Architecture*. MIT Press, 2012.

[31] H. Levesque. A logic of implicit and explicit belief. In *Proceedings of AAAI-84*, pages 198–202. AAAI Press, 1984.

[32] T. de Lima. Alternating-time temporal dynamic epistemic logic. *Journal of Logic and Computation*, 24(6):1145–1178, 2014.

[33] M. C. Lovett, L. M. Reder, and C. Lebiere. Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake and P. Shah, editors, *Models of working memory: Mechanisms of Active Maintenance and Executive Control*, pages 135–182. Cambridge University Press, 1999.

[34] C. Lutz. Complexity and Succinctness of Public Announcement Logic. In *Proceedings of AAMAS 2006*, pages 137–144. ACM Press, 2006.

[35] J. van Benthem and E. Pacuit. Dynamic Logics of Evidence-Based Beliefs. *Journal of Logic, Language and Information*, 99(1-3):61–92, 2011.

[36] H. van Ditmarsch and T. French. Semantics for knowledge and change of awareness. *Journal of Logic, Language and Information*, 23(2):169–195, 2014.

[37] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, 2008.

[38] F. R. Velázquez-Quesada. Dynamic Epistemic Logic for Implicit and Explicit Beliefs. *Journal of Logic, Language and Information*, 23(2):107–140, 2014.

[39] F. R. Velázquez-Quesada. Explicit and implicit knowledge in neighbourhood models. In *Proceedings of LORI 2013*, volume 8196 of *LNCS*, pages 239–252. Springer, 2013.

[40] R. M. Young and R. L. Lewis. The Soar cognitive architecture and human working memory. In A. Miyake and P. Shah, editors, *Models of working memory: Mechanisms of Active Maintenance and Executive Control*, pages 135–182. Cambridge University Press, 1999.

Philippe Balbiani
Emiliano Lorini
IRIT
Toulouse University
Toulouse, France
{balbiani,lorini}@irit.fr

David Fernández-Duque
Department of Mathematics
Ghent University
Ghent, Belgium
David.FernandezDuque@UGent.be