



HAL
open science

The Dynamics of Epistemic Attitudes in Resource-Bounded Agents

Philippe Balbiani, David Fernández-Duque, Emiliano Lorini

► **To cite this version:**

Philippe Balbiani, David Fernández-Duque, Emiliano Lorini. The Dynamics of Epistemic Attitudes in Resource-Bounded Agents. *Studia Logica*, 2019, Special Issue: 40 years of FDE, 107, pp.457-488. 10.1007/s11225-018-9798-4 . hal-02390930

HAL Id: hal-02390930

<https://hal.science/hal-02390930v1>

Submitted on 14 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.

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. 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 $\varphi \rightarrow \psi$ 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.*, $\varphi \rightarrow \psi$ 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.

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.*, [1, 22, 24, 36]) starting from the seminal work of Fagin and 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 $DES4_n$ 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

71 certain agent i knows something, with tense-like operators of type $\langle F_i \rangle$, that
 72 allows to represent the fact that a certain agent i will *possibly* get to know
 73 something.¹ Duc's system shares with our logic DLEK the idea that (i) an
 74 agent gets to know (or believe) something by performing inferences, and (ii)
 75 making inferences takes time. Nonetheless, while in our logic DLEK infer-
 76 ential operations are represented both at the syntactic level, via dynamic
 77 operators in the DEL style, and a semantic level, as model update opera-
 78 tions, in Duc's system and its formal semantics given by [2] they are not.
 79 The formal semantics for DES4_n given by Ågotnes and Alechina is a Kripke-
 80 style semantics in which every operator $\langle F_i \rangle$ is associated with an abstract
 81 accessibility relation R_i over possible worlds, where wR_iv means that world
 82 w is related with world v . This semantics does not say anything about the
 83 kind of inferential operation (or sequence of inferential operations) that is
 84 responsible for the transition from world w to world v . More generally, the
 85 system DES4_n does not support reasoning about the consequences of a spe-
 86 cific kind of inferential operation such as application of *modus ponens* or
 87 closure under conjunction on an agent's knowledge or beliefs. On the con-
 88 trary, this is something our logic DLEK can express.

89 Another related work is [4], which presents a semantics for a variant
 90 of Alternating-time Temporal Logic (ATL) extended by explicit knowledge
 91 operators. This logic supports reasoning about, among other things, the
 92 capability of an agent of gaining some knowledge by applying a certain rea-
 93 soning rule such as reasoning with *modus ponens* or reasoning monotonically.
 94 This logic does not take into account the distinction between explicit knowl-
 95 edge (or belief) and background knowledge which is fundamental for our
 96 analysis. But, the most important difference between this system and ours
 97 is their belonging to two different families of logics. Ågotnes and Walicki's
 98 logic is a logic for strategic reasoning with a semantics based on the concept
 99 of concurrent game structure (CGS). It has modal operators of strategic
 100 capability that represent what an agent or coalition of agents can achieve
 101 by playing a certain strategy. On the contrary, our logic DLEK belongs to
 102 the family of dynamic epistemic logics whose main constituents are dynamic
 103 modal operators that allow to represent the consequences of a specific com-
 104 municative action or mental operation. Apart from some rare exceptions
 105 (see, e.g., [3, 32]), the connection between the two families of logics has been
 106 rarely explored. Moreover, it is not the object of the present work.

¹The 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.

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

140 **Plan of the paper.** The paper is organized as follows. In Section 2 we present
 141 the conceptual foundation of our logic DLEK, namely the general view about
 142 dynamics of beliefs in resource-bounded agents which underlies our formal
 143 theory. The general idea is that new beliefs can be formed either by per-
 144 ception or by inferring them from existing beliefs in working memory and
 145 by retrieving information from background knowledge in long-term memory.



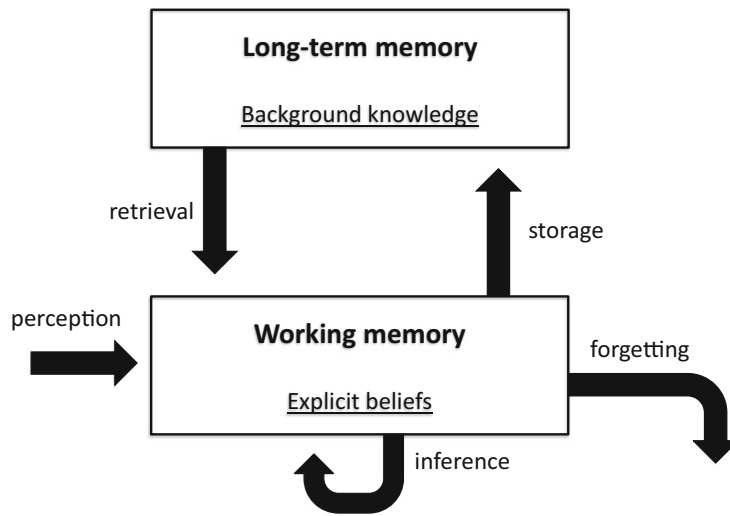


Figure 1. Cognitive view of the relationship between background knowledge and explicit beliefs

146 Section 3 presents the syntax and the semantics of DLEK. We will show that,
 147 in DLEK, perceptive and inferential steps are modelled as special kinds of
 148 model-update operations. Section 4 is devoted to presenting an axiomatic
 149 system and showing its completeness. Section 5 presents complexity results
 150 for the satisfiability problem of the static fragment of DLEK, called LEK, as
 151 well as a decidability result for the satisfiability problem of DLEK. Given
 152 that the semantics of the static fragment of DLEK are non-standard, in Sec-
 153 tion 6 we show how they can be reduced to purely relational semantics,
 154 albeit with additional modalities.

155 2. Conceptual Framework

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

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

163 The information available in LTM, generally called *background knowl-*
 164 *edge*, includes both knowledge of specific events in the past and conceptual
 165 or causal knowledge representing the agent's unproblematic interpretation

166 of reality.² For example, an agent may have background conceptual knowl-
 167 edge about how restaurants are organized or background causal knowledge
 168 about the relation between smoke and fire. In particular, she may know that
 169 restaurants have waiters, chairs and tables or that if smoke comes out from
 170 a window of a certain house, then there is fire inside the house.

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

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

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

²In 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.

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

200 3. Logical Framework

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

209 3.1. Syntax

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

214 The language of DLEK, denoted by \mathcal{L}_{DLEK} , is defined by the following
215 grammar in Backus–Naur form:

$$216 \quad \begin{aligned} \alpha & ::= \top(\varphi, \psi) \mid \cap(\varphi, \psi) \mid +\varphi \mid -\varphi \\ \varphi, \psi & ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathbf{B}_i\varphi \mid \mathbf{K}_i\varphi \mid [(G:\alpha)_H]\varphi \end{aligned}$$

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

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

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

$$224 \quad \varphi, \psi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathbf{B}_i\varphi \mid \mathbf{K}_i\varphi.$$

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

226 The formula $\mathbf{B}_i\varphi$ is read “the agent i explicitly believes that φ is true” or,
227 more shortly, “the agent i believes that φ is true”. As explained in Section 2,
228 explicit beliefs are accessible in working memory and are the basic elements
229 of the agents' reasoning process. Following the logic of local reasoning by
230 Fagin and Halpern [15], we wish to give a purely semantical model of explicit

231 belief. As we will show in Section 4, an effect of this approach is that agents
 232 cannot distinguish between logically equivalent formulas, i.e., if two facts
 233 φ and ψ are logically equivalent and an agent explicitly believes that φ is
 234 true, then she believes that ψ is true as well. There are other approaches,
 235 such as justification logics [5], that do not have this feature. However, the
 236 interpretation of these logics relies highly on syntax, a situation we wish to
 237 avoid in the present paper.

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

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

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

271 We distinguish four types of mental operations α which allow us to cap-
 272 ture some of the dynamic properties of explicit beliefs and background
 273 knowledge informally described in Section 2 above: $+\varphi$, $-\varphi$, $\vdash(\varphi,\psi)$ and
 274 $\cap(\varphi,\psi)$. The operations $+\varphi$ and $-\varphi$ correspond, respectively, to the men-
 275 tal operations of forming an explicit belief via perception and forgetting
 276 an explicit belief represented in Figure 1. $\vdash(\varphi,\psi)$ and $\cap(\varphi,\psi)$ characterize
 277 two basic operations of forming explicit beliefs via inference. Specifically,
 278 $\vdash(\varphi,\psi)$ is the mental operation which consists in inferring ψ from φ in case
 279 φ is believed and, according to an agent's background knowledge, ψ is a
 280 logical consequence of φ . In other words, by performing this mental opera-
 281 tion, an agent tries to retrieve from her background knowledge in long-term
 282 memory the information that φ implies ψ and, if she succeeds, she starts
 283 to believe ψ .⁴ $\cap(\varphi,\psi)$ is the mental operation which consists in closing the
 284 explicit belief that φ and the explicit belief that ψ under conjunction. In
 285 other words, $\cap(\varphi,\psi)$ characterizes the mental operation of deducing $\varphi \wedge \psi$
 286 from the explicit belief that φ and the explicit belief that ψ .

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

292 3.2. Semantics

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

296 DEFINITION 1. (LEK/LEK⁺ model) We define a LEK model to be a tuple
 297 $M = (W, N, R_1, \dots, R_n, V)$ where:

- 298 (a) W is a set of worlds or situations;
- 299 (b) for every $i \in \text{Agt}$, $R_i \subseteq W \times W$ is a serial, transitive and Euclidean⁵
 300 relation on W ;
- 301 (c) $N : \text{Agt} \times W \longrightarrow 2^{2^W}$ is a neighborhood function such that for all
 302 $i \in \text{Agt}$, $w, v \in W$ and $X \subseteq W$:

⁴Note 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.

⁵Recall 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 .

- 303 (C1) if $X \in N(i, w)$ then $X \subseteq R_i(w)$,
 304 (C2) if wR_iv then $N(i, w) = N(i, v)$;
 305 (d) $V : W \longrightarrow 2^{Atm}$ is a valuation function.

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

307 For every $i \in Agt$ and $w \in W$, $R_i(w) = \{v \in W : wR_iv\}$ identifies the set
 308 of situations that agent i considers possible at world w . In cognitive terms,
 309 $R_i(w)$ can be conceived as the set of all situations that agent i can retrieve
 310 from her long-term memory and reason about them. More generally, $R_i(w)$
 311 is called agent i 's epistemic state at w . In LEK^+ -models, the reason why R_i
 312 is an equivalence relation is that it is used to model a form of omniscient
 313 background knowledge instead of omniscient background belief. The latter
 314 could be modelled, as in LEK-models, by replacing the equivalence relations
 315 R_i by serial, transitive and Euclidean relations commonly used in doxastic
 316 logic to model a notion of belief.

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

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

326 Truth conditions of DLEK formulas are inductively defined as follows.

327 DEFINITION 2. For a LEK model $M = (W, N, R_1, \dots, R_n, V)$, a world $w \in$
 328 W , a formula $\varphi \in \mathcal{L}_{LEK}$, and an action α , we define the truth relation
 329 $M, w \models \varphi$ and a new LEK model $M^{(G:\alpha)_H}$ by simultaneous recursion on α
 330 and φ as follows. Below, we write

$$331 \quad \|\varphi\|_{i,w}^M = \{v \in W : wR_iv \text{ and } M, v \models \varphi\}$$

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

$$333 \quad M, w \models p \iff p \in V(w)$$

$$334 \quad M, w \models \neg\varphi \iff M, w \not\models \varphi$$

$$335 \quad M, w \models \varphi \wedge \psi \iff M, w \models \varphi \text{ and } M, w \models \psi$$

$$336 \quad M, w \models B_i\varphi \iff \|\varphi\|_{i,w}^M \in N(i, w)$$

$$337 \quad M, w \models K_i\varphi \iff M, v \models \varphi \text{ for all } v \in R_i(w)$$

$$338 \quad M, w \models [(G:\alpha)_H]\varphi \iff M^{(G:\alpha)_H}, w_0 \models \varphi$$

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

341 First, for $i \in \text{Agt}$ and $w \in W$, set

$$342 \quad N^{+\psi}(i, w) = N(i, w) \cup \{ \|\psi\|_{i,w}^M \}$$

$$343 \quad N^{-\psi}(i, w) = N(i, w) \setminus \{ \|\psi\|_{i,w}^M \}$$

$$344 \quad N^{+(\psi,\chi)}(i, w) = \begin{cases} N(i, w) \cup \{ \|\chi\|_{i,w}^M \} & \text{if } M, w \models \mathbf{B}_i\psi \wedge \mathbf{K}_i(\psi \rightarrow \chi) \\ N(i, w) & \text{otherwise} \end{cases}$$

$$345 \quad N^{\cap(\psi,\chi)}(i, w) = \begin{cases} N(i, w) \cup \{ \|\psi \wedge \chi\|_{i,w}^M \} & \text{if } M, w \models \mathbf{B}_i\psi \wedge \mathbf{B}_i\chi \\ N(i, w) & \text{otherwise} \end{cases}$$

346 Then we define

$$347 \quad W^{(G:\alpha)_H} = W \times \{0, 1\}$$

$$348 \quad N^{(G:\alpha)_H}(i, w_0) = \{X \times \{0\} : X \in N^\alpha(i, w)\} \text{ if } i \in G$$

$$349 \quad N^{(G:\alpha)_H}(i, w_1) = \{X \times \{1\} : X \in N(i, w)\} \text{ if } i \in G$$

$$350 \quad N^{(G:\alpha)_H}(i, w_0) = \{X \times \{0\} : X \in N(i, w)\} \text{ if } i \in H \setminus G$$

$$351 \quad N^{(G:\alpha)_H}(i, w_1) = \{X \times \{1\} : X \in N(i, w)\} \text{ if } i \in H \setminus G$$

$$352 \quad N^{(G:\alpha)_H}(i, w_0) = \{X \times \{1\} : X \in N(i, w)\} \text{ if } i \notin H$$

$$353 \quad N^{(G:\alpha)_H}(i, w_1) = \{X \times \{1\} : X \in N(i, w)\} \text{ if } i \notin H$$

$$354 \quad R_i^{(G:\alpha)_H} = \{(w_0, v_0) : wR_iv\} \cup \{(w_1, v_1) : wR_iv\} \text{ if } i \in H$$

$$355 \quad R_i^{(G:\alpha)_H} = \{(w_0, v_1) : wR_iv\} \cup \{(w_1, v_1) : wR_iv\} \text{ if } i \notin H$$

$$356 \quad V^{(G:\alpha)_H}(w_x) = V(w) \text{ with } x \in \{0, 1\}$$

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

359 The fact that the mental operation α is performed by the agents in G
 360 and this is only common knowledge for the agents in H is captured by the
 361 duplication of the original model M . The 0-part of the original model is
 362 the part in which agents in G update their explicit beliefs via the mental
 363 operation α , while the 1-part is the part in which nothing happens. All
 364 agents inside H are aware that those in G have changed their beliefs via the

370 mental operation α , as they have access to the 0-part. On the contrary, the
 371 agents outside of H are not aware that those in G have changed their beliefs
 372 via the mental operation α , as they only have access to the 1-part thereby
 373 assuming that nothing happens.

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

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

384 According to the previous truth conditions, an agent i explicitly believes
 385 φ at world w if and only if, at world w , agent i has the fact corresponding
 386 to the formula φ (i.e., $\|\varphi\|_{i,w}^M$) included in her explicit belief set. Note that
 387 this semantic interpretation is different from the usual one for neighbor-
 388 hood formulas, as we restrict the extension of φ to the agent's epistemically
 389 accessible worlds. This is justified by the fact that an agent's explicit beliefs
 390 are relative to her background knowledge. The latter guarantees that if an
 391 agent explicitly believes that φ is true (i.e., $\text{B}_i\varphi$) while having background
 392 knowledge that φ is false (i.e., $\text{K}_i\neg\varphi$) then, she explicitly entertains a con-
 393 tradiction (i.e., $\text{B}_i\perp$). This seems a reasonable property of explicit beliefs.
 394 Moreover, an agent has background knowledge that φ is true if and only if
 395 φ is true in all situations that are included in the agent's epistemic state.

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

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

412 3.3. Some Validities

413 The following four validities capture the basic properties of the four mental
 414 operations $\vdash(\varphi, \psi)$, $\cap(\varphi, \psi)$, $+\varphi$ and $-\varphi$ semantically defined above. Let
 415 $\varphi, \psi \in \text{Prop}$ and $i \in G \subseteq H \subseteq \text{Agt}$. Then, the following are valid over the
 416 class of LEK -models:

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

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

427 The reason why we need to impose that φ and ψ are propositional
 428 formulas is that there are DLEK -formulas such as the Moore-like formula
 429 $p \wedge \neg \mathbf{B}_ip$ for which the previous four principles do not hold. For instance,
 430 $[(\{i\}: +(p \wedge \neg \mathbf{B}_ip))_{\{i\}}] \mathbf{B}_i(p \wedge \neg \mathbf{B}_ip)$ is not valid.

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

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

- 441 1. $\models_{\text{DLEK}} (\mathbf{B}_i\varphi \wedge \mathbf{B}_i(\varphi \rightarrow \psi)) \rightarrow [\cap(\varphi, \varphi \rightarrow \psi)][\vdash(\varphi \wedge \psi, \psi)] \mathbf{B}_i\psi$
- 442 2. $\models_{\text{DLEK}} \mathbf{B}_i\varphi \rightarrow [\vdash(\varphi, \varphi \rightarrow (\varphi \vee \psi))] \mathbf{B}_i(\varphi \vee \psi)$
- 443 3. $\models_{\text{DLEK}} \mathbf{B}_i\psi \rightarrow [\vdash(\psi, \psi \rightarrow (\varphi \vee \psi))] \mathbf{B}_i(\varphi \vee \psi)$

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

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

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

456 3.4. Example

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

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

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

470 Thus, let us suppose that $G = H = \text{Agt} = \{A, B\}$ and $\text{Atm} = \{r, g, h, b\}$.
 471 We represent the initial situation by a minimal LEK model satisfying
 472 the hypothesis H1 and H2 and which only excludes the impossible situ-
 473 ations in which the electric light is both red and green and the person
 474 needs help and takes a rest at the same time.⁶ This model is the tuple
 475 $\text{MR} = (W, N, R_A, R_B, V)$ (where MR stands for ‘model of the robots’) such
 476 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 .

- 477 • $W = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9\}$;
- 478 • $N(i, w) = \emptyset$ for all $i \in \text{Agt}$ and for all $w \in W$;
- 479 • the quotient set of W by R_A is $\{\{w_1, w_2, w_3, w_4, w_5\}, \{w_6, w_7\}, \{w_8, w_9\}\}$;
- 480 • the quotient set of W by R_B is $\{\{w_1, w_2, w_3, w_4, w_5, w_6, w_7\}, \{w_8, w_9\}\}$;
- 481 • $V(w_1) = \{r, h\}$, $V(w_2) = \{h\}$, $V(w_3) = \{g, b\}$, $V(w_4) = \{b\}$, $V(w_5) = \emptyset$,
- 482 $V(w_6) = \{g\}$, $V(w_7) = \{g, h\}$, $V(w_8) = \{r\}$, and $V(w_9) = \{r, b\}$.

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

487 The first thing to observe is that in the actual situation the hypothesis
 488 H1 and H2 are both satisfied. Indeed, as for H1, we have: $\text{MR}, w_1 \models \text{K}_A(r \rightarrow$
 489 $h) \wedge \text{K}_A(g \rightarrow b)$. As for H2, we have: $\text{MR}, w_1 \models \text{K}_B(r \rightarrow h) \wedge \neg \text{K}_B(g \rightarrow b)$.

490 Let us suppose that the person switches on the red light in order to
 491 signal to the two robots that she needs help. This event is represented by
 492 the mental operation $+r$ which leads from model MR to the updated model
 493 $\text{MR}^{+r} = (W, N^{+r}, R_A, R_B, V)$ such that:

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

495 Note that the only difference between between A 's and B 's neighbour-
 496 hoods at MR^{+r} is on worlds w_6 and w_7 .

497 It is easy to check that in the new situation, after the mental operation
 498 $+r$ has been executed, the two robots explicitly believe that the light is red.
 499 That is: $\text{MR}^{+r}, w_1 \models \text{B}_{\{A,B\}}r$, where $\text{B}_{\{A,B\}}\varphi$ is a abbreviation of $\text{B}_A\varphi \wedge \text{B}_B\varphi$.
 500 However, the mental operation is not sufficient to guarantee that the robots
 501 believe that the person needs helps. Indeed, we have: $\text{MR}^{+r}, w_1 \models \neg \text{B}_A h \wedge$
 502 $\neg \text{B}_B h$.

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

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

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: $(\mathbf{MR}^{+r})^{\vdash(r,h)}, w_1 \models \mathbf{B}_{\{A,B\}}h$. To sum up, we have that the following holds:

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

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:

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

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

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

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

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

$$545 \quad M, w \models [?_{i,j}\alpha]\varphi \iff M^{?_{i,j}\alpha}, w_0 \models \varphi$$

546 $M^{?_{i,j}\alpha} = (W^{?_{i,j}\alpha}, N^{?_{i,j}\alpha}, R_1^{?_{i,j}\alpha}, \dots, R_n^{?_{i,j}\alpha}, V^{?_{i,j}\alpha})$ is the updated model
 547 such that for all $k \in \text{Agt}$ and for all $w \in W$:

$$548 \quad W^{?_{i,j}\alpha} = W \times \{0, 1\}$$

$$549 \quad N^{?_{i,j}\alpha}(j, w_0) = \{X \times \{0\} : X \in N^\alpha(j, w)\}$$

$$550 \quad N^{?_{i,j}\alpha}(j, w_1) = \{X \times \{1\} : X \in N(j, w)\}$$

$$551 \quad N^{?_{i,j}\alpha}(i, w_0) = \{X \times \{0, 1\} : X \in N(i, w)\}$$

$$552 \quad N^{?_{i,j}\alpha}(i, w_1) = \{X \times \{0, 1\} : X \in N(i, w)\}$$

$$553 \quad N^{?_{i,j}\alpha}(k, w_0) = \{X \times \{1\} : X \in N(k, w)\} \text{ if } k \neq i \text{ and } k \neq j$$

$$554 \quad N^{?_{i,j}\alpha}(k, w_1) = \{X \times \{1\} : X \in N(k, w)\} \text{ if } k \neq i \text{ and } k \neq j$$

$$555 \quad R_i^{?_{i,j}\alpha} = \{(w_x, v_y) : x, y \in \{0, 1\} \text{ and } wR_iv\}$$

$$556 \quad R_j^{?_{i,j}\alpha} = \{(w_0, v_0) : wR_jv\} \cup \{(w_1, v_1) : wR_jv\}$$

$$557 \quad R_k^{?_{i,j}\alpha} = \{(w_0, v_1) : wR_kv\} \cup \{(w_1, v_1) : wR_kv\} \text{ if } k \neq i \text{ and } k \neq j$$

$$558 \quad V^{?_{i,j}\alpha}(w_x) = V(w) \text{ with } x \in \{0, 1\}$$

560 The following proposition guarantees that the semantics of the operator
 561 $[?_{i,j}\alpha]$ with respect to the class of LEK models is well defined:

Footnote 7 continued

that j 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.

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

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

571 4. Axiomatization

572 Let us now present sound and complete axiomatizations for the logics LEK,
 573 LEK⁺ and their dynamic extensions DLEK, DLEK⁺. For a logic λ , we write
 574 $\lambda \vdash \varphi$ to denote the fact that φ is a theorem of λ , i.e., that φ belongs to the
 575 least set containing all axioms of λ and closed under the rules of λ .

576 DEFINITION 3. We define LEK⁺ to be the extension of classical propositional
 577 logic given by the following rules and axioms:

$$\begin{array}{ll}
 578 & (K_i\varphi \wedge K_i(\varphi \rightarrow \psi)) \rightarrow K_i\psi & (\mathbf{K}_{K_i}) \\
 579 & K_i\varphi \rightarrow \varphi & (\mathbf{T}_{K_i}) \\
 580 & \neg K_i(\varphi \wedge \neg\varphi) & (\mathbf{D}_{K_i}) \\
 581 & K_i\varphi \rightarrow K_i K_i\varphi & (\mathbf{4}_{K_i}) \\
 582 & \neg K_i\varphi \rightarrow K_i \neg K_i\varphi & (\mathbf{5}_{K_i}) \\
 583 & (B_i\varphi \wedge K_i(\varphi \leftrightarrow \psi)) \rightarrow B_i\psi & (\mathbf{Mix1}_{K_i, B_i}) \\
 584 & B_i\varphi \rightarrow K_i B_i\varphi & (\mathbf{Mix2}_{K_i, B_i}) \\
 585 & \frac{\varphi}{K_i\varphi} & (\mathbf{Nec}_{K_i}) \\
 586 & &
 \end{array}$$

587 The logic LEK is the logic obtained by removing (\mathbf{T}_{K_i}) .

588 It is straightforward to check that all axioms are valid and all rules pre-
 589 serve validity, from which we obtain:

590 LEMMA 1. *The logic LEK⁽⁺⁾ is sound for the class of LEK⁽⁺⁾ models.*

591 Observe that (\mathbf{D}_{K_i}) follows from (\mathbf{T}_{K_i}) , but the former is needed in
 592 LEK. Most of the axioms are familiar from modal and epistemic logic,
 with the possible exceptions of $(\mathbf{Mix1}_{K_i, B_i})$ and $(\mathbf{Mix2}_{K_i, B_i})$. The latter

593 simply corresponds to **(C2)**, which states that the neighborhood relation
 594 does not vary within $R_i(w)$. Regarding **(Mix1 $_{K_i, B_i}$)**, observe that if $\varphi \leftrightarrow \psi$
 595 were valid, then so would $B_i\varphi \rightarrow B_i\psi$, as then $\|\varphi\|_{i,w}^M = \|\psi\|_{i,w}^M$ regard-
 596 less of M or w . However, as $\|\varphi\|_{i,w}^M, \|\psi\|_{i,w}^M \subseteq R_i(w)$, it suffices that
 597 $\|\varphi\|_{i,w}^M \cap R_i(w) = \|\psi\|_{i,w}^M \cap R_i(w)$, i.e., that $K_i(\varphi \leftrightarrow \psi)$ is true. Note moreover
 598 that $K_i(\varphi \rightarrow \psi)$ is *not* sufficient for $B_i\varphi \rightarrow B_i\psi$ to hold, as the neighborhood
 599 relation is not assumed monotone.

600 The axiomatics of the logic $DLEK^{(+)}$ includes all principles of the logic
 601 $LEK^{(+)}$ plus a set of reduction axioms and the rule of replacement of equiv-
 602 alents.

603 **DEFINITION 4.** We define $DLEK^{(+)}$ to be the extension of $LEK^{(+)}$ generated
 604 by the following axioms and rule of inference, where $G \subseteq H \subseteq Agt$, $i \in G$,
 605 $k \in H \setminus G$ and $j \notin H$:

$$606 \quad [(G:\alpha)_H]p \leftrightarrow p \quad (\mathbf{Red1})$$

$$607 \quad [(G:\alpha)_H]\neg\varphi \leftrightarrow \neg[(G:\alpha)_H]\varphi \quad (\mathbf{Red2})$$

$$608 \quad [(G:\alpha)_H](\varphi \wedge \psi) \leftrightarrow ([(G:\alpha)_H]\varphi \wedge [(G:\alpha)_H]\psi) \quad (\mathbf{Red3})$$

$$609 \quad [(G:\alpha)_H]K_i\varphi \leftrightarrow K_i[(G:\alpha)_H]\varphi \quad (\mathbf{Red4})$$

$$610 \quad [(G: + \varphi)_H]B_i\psi \leftrightarrow (B_i[(G: + \varphi)_H]\psi \vee K_i([(G: + \varphi)_H]\psi \leftrightarrow \varphi)) \quad (\mathbf{Red5})$$

$$611 \quad [(G: - \varphi)_H]B_i\psi \leftrightarrow (B_i[(G: - \varphi)_H]\psi \wedge \neg K_i([(G: - \varphi)_H]\psi \leftrightarrow \varphi)) \quad (\mathbf{Red6})$$

$$612 \quad [(G: \vdash(\varphi, \psi))_H]B_i\chi \leftrightarrow \left(B_i[(G: \vdash(\varphi, \psi))_H]\chi \vee ((B_i\varphi \wedge K_i(\varphi \rightarrow \psi)) \wedge \right. \quad (\mathbf{Red7})$$

$$613 \quad \left. K_i([(G: \vdash(\varphi, \psi))_H]\chi \leftrightarrow \psi)) \right)$$

$$614 \quad [(G: \cap(\varphi, \psi))_H]B_i\chi \leftrightarrow \left(B_i[(G: \cap(\varphi, \psi))_H]\chi \vee ((B_i\varphi \wedge B_i\psi) \wedge \right. \quad (\mathbf{Red8})$$

$$615 \quad \left. K_i([(G: \cap(\varphi, \psi))_H]\chi \leftrightarrow (\varphi \wedge \psi)) \right)$$

$$616 \quad [(G:\alpha)_H]K_k\varphi \leftrightarrow K_k[(G:\alpha)_H]\varphi \quad (\mathbf{Red9})$$

$$617 \quad [(G:\alpha)_H]B_k\varphi \leftrightarrow B_k[(G:\alpha)_H]\varphi \quad (\mathbf{Red10})$$

$$618 \quad [(G:\alpha)_H]K_j\varphi \leftrightarrow K_j\varphi \quad (\mathbf{Red11})$$

$$619 \quad [(G:\alpha)_H]B_j\varphi \leftrightarrow B_j\varphi \quad (\mathbf{Red12})$$

621 and the following rule of inference:

$$622 \quad \frac{\psi_1 \leftrightarrow \psi_2}{\varphi \leftrightarrow \varphi[\psi_1/\psi_2]} \quad (\mathbf{RRE})$$

624 The first four axioms are standard from dynamic epistemic logic. Mean-
 625 while, for $i \in G$ the intuition behind all of the **(Red5)**–**(Red8)** axioms is

626 similar; consider, for example, (Red5). The left-hand side states that, after
 627 adding φ as a neighborhood, the agent will believe ψ , i.e. the extension of
 628 ψ within $R_i(w)$ will be a neighborhood in the new model. This can hap-
 629 pen for precisely two reasons: either the extension of ψ in the new model,
 630 i.e. $[(G: + \varphi)_H]\psi$, already *was* a neighborhood, in which case $B_i[(G: + \varphi)_H]\psi$
 631 holds; or, this extension is precisely the neighborhood that was added,
 632 i.e. $K_i([(G: + \varphi)_H]\psi \leftrightarrow \varphi)$ holds.

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

636 For (Red11) and (Red12) the intuition is that, for agents not in H , the
 637 set of possible worlds is *exactly the same* as it was before those in G realized
 638 the private mental operation. With this in mind, it is straightforward to
 639 check that all axioms are valid and all rules preserve validity in the class of
 640 $\text{LEK}^{(+)}$ -models, from which the following is an immediate consequence:

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

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

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

- 649 • W^λ is the set of all maximal λ -consistent subsets of \mathcal{L}_{LEK} .
- 650 • $wR_i^\lambda v$ if and only if, for all formulas φ , $K_i\varphi \in w$ implies that $\varphi \in v$.
- 651 • In order to define N^λ , for $w \in W$ and $\varphi \in \mathcal{L}_{\text{LEK}}$, first define $A_\varphi(i, w) =$
 652 $\{v \in R_i^\lambda(w) : \varphi \in v\}$. Then, define N^λ by letting $N^\lambda(i, w) = \{A_\varphi(i, w) :$
 653 $B_i\varphi \in w\}$.
- 654 • Finally, we define the valuation V^λ by $w \in V^\lambda(p)$ if and only if $p \in w$.

655 The following is standard and we omit the proof:

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

658. $K_i\varphi \in w$ if and only if, for every v such that $wR_i^\lambda v$, $\varphi \in v$, and
659. if $wR_i^\lambda v$ and $B_i\varphi \in w$, then $B_i\varphi \in v$.

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

662 LEMMA 4. For every $w \in W^\lambda$ and $B_i\varphi, B_i\psi \in \mathcal{L}_{LEK}$, if $B_i\varphi \in w$ but $B_i\psi \notin w$,
 663 it follows that there is $v \in R_i^\lambda(w)$ such that either $\varphi \in v$ but $\neg\psi \in v$, or
 664 $\neg\varphi \in v$ but $\psi \in v$.

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

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

670 LEMMA 5. For every $\varphi \in \mathcal{L}_{LEK}$ and every $w \in W^\lambda$, $\varphi \in w$ if and only if
 671 $M^\lambda, w \models \varphi$.

672 PROOF. The proof proceeds by a standard induction on the construction of
 673 φ . All cases are routine except $\varphi = B_i\psi$.

674 First assume that $B_i\psi \in w$. Then, $A_\psi(i, w) \in N^\lambda(i, w)$. But, $A_\psi(i, w) =$
 675 $\{v \in R_i^\lambda(w) : \psi \in v\} \stackrel{\text{IH}}{=} \|\psi\| \cap R_i^\lambda(w)$. Thus, $M^\lambda, w \models B_i\psi$.

676 Now, suppose $B_i\psi \notin w$, so that $\neg B_i\psi \in w$. We must check that $\|\psi\| \cap$
 677 $R_i^\lambda(w) \notin N(i, w)$. Choose an arbitrary set $A \in N(i, w)$; by definition, $A =$
 678 $A_\theta(i, w)$ for some θ with $B_i\theta \in w$. By Lemma 4, there is some $v \in R_i^\lambda(w)$ such
 679 that $\psi, \neg\theta \in v$ or $\neg\psi, \theta \in v$; in the first case, this shows using the induction
 680 hypothesis that $v \in (\|\psi\| \cap R_i^\lambda(w)) \setminus A_\theta(i, w)$, while in the second we obtain
 681 $v \in A_\theta(i, w) \setminus (\|\psi\| \cap R_i^\lambda(w))$. In either case we obtain $A_\theta(i, w) \neq \|\psi\| \cap R_i^\lambda(w)$,
 682 and since $A = A_\theta(i, w)$ was an arbitrary element of $N^\lambda(i, w)$, we conclude
 683 that $\|\psi\| \cap R_i^\lambda(w) \notin N^\lambda(i, w)$ and thus $M^\lambda, w \not\models B_i\psi$. ■

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

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

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

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

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

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

696 As a corollary, we get the following:

697 THEOREM 2. *DLEK is strongly complete for the class of LEK models and*
698 *mono-agent DLEK⁽⁺⁾ is strongly complete for the class of LEK⁽⁺⁾ models.*

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

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

707 5. Complexity

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

714 5.1. Mono-Agent Case

715 Assume $\text{card}(\text{Agt}) = 1$. Let φ be a satisfiable formula. Let $M = (W, N, V)$
716 be a model and $w \in W$ be such that $M, w \models \varphi$; note that in the mono-
717 agent case we can assume that R is the total relation, so we omit it. Let
718 $\mathbf{K}\psi_1, \dots, \mathbf{K}\psi_m$ and $\mathbf{B}\chi_1, \dots, \mathbf{B}\chi_n$ be lists of all subformulas of φ of the form
719 $\mathbf{K}\psi$ and $\mathbf{B}\chi$. Let $\hat{K} = \{i : 1 \leq i \leq m \ \& \ M, w \not\models \mathbf{K}\psi_i\}$, $\hat{B}^+ = \{j : 1 \leq j \leq$
720 $n \ \& \ M, w \models \mathbf{B}\chi_j\}$ and $\hat{B}^- = \{k : 1 \leq k \leq n \ \& \ M, w \not\models \mathbf{B}\chi_k\}$. For all $i \in \hat{K}$,
721 let $v_i \in W$ be such that $M, v_i \not\models \psi_i$. Such v_i exists because $M, w \not\models \mathbf{K}\psi_i$. For
722 all $j \in \hat{B}^+$, let $A_j \in N$ be such that $A_j = \{v \in W : M, v \models \chi_j\}$. Such A_j
723 exists because $M, w \models \mathbf{B}\chi_j$. For all $j \in \hat{B}^+$ and for all $k \in \hat{B}^-$, let $u_{j,k} \in W$
724 be such that $M, u_{j,k} \not\models \chi_j \leftrightarrow \chi_k$. Such $u_{j,k}$ exists because $M, w \models \mathbf{B}\chi_j$ and
725 $M, w \not\models \mathbf{B}\chi_k$. Let $M' = (W', N', V')$ be the model defined by:

- 726 • $W' = \{w\} \cup \{v_i : i \in \hat{K}\} \cup \{u_{j,k} : j \in \hat{B}^+ \ \& \ k \in \hat{B}^-\}$,

- 727 • $N' = \{A_j \cap W' : j \in \hat{B}^+\},$
- 728 • for all $p \in \text{Atm}, V'(p) = V(p) \cap W'.$

729 Obviously, $\text{card}(W')$ and $\text{card}(N')$ are polynomial in the size of φ . Let $\Phi =$
 730 $SF(\varphi)$, the closure under single negations of the set of all of φ 's subformulas.

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

733 PROOF. By induction on ψ . We only consider the cases $\psi = B\chi$.

734 Suppose $M, s \models B\chi$ and $M', s \not\models B\chi$. Let $j \in \hat{B}^+$ be such that $B\chi = B\chi_j$.
 735 Hence, $A_j = \{t \in W : M, t \models \chi\}$. By induction hypothesis, $A_j \cap W' = \{t \in$
 736 $W' : M', t \models \chi\}$. Thus, $M', s \models B\chi$: a contradiction.

737 Suppose $M', s \models B\chi$ and $M, s \not\models B\chi$. Let $j \in \hat{B}^+$ be such that $A_j \cap$
 738 $W' = \{t \in W' : M', t \models \chi\}$. By induction hypothesis, $A_j \cap W' = \{t \in$
 739 $W' : M, t \models \chi\}$. Let $k \in \hat{B}^-$ be such that $B\chi = B\chi_k$. Remember that
 740 $A_j = \{t \in W : M, t \models \chi_j\}$. Moreover, $u_{j,k} \in W'$ is such that $M, u_{j,k} \models \chi_j$
 741 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,
 742 $u_{j,k} \in A_j \cap W'$. Hence, $M, u_{j,k} \models \chi$: a contradiction. In the latter case,
 743 $u_{j,k} \in A_j \cap W'$. Thus, $M, u_{j,k} \models \chi_j$: a contradiction. ■

744 THEOREM 3. *If $\text{card}(\text{Agt}) = 1$ then satisfiability problem of LEK^+ is NP-*
 745 *complete.*

746 PROOF. Membership in NP follows from Lemma 7. NP-hardness follows
 747 from the NP-hardness of classical propositional logic. ■

748 5.2. Multi-Agent Case

749 Our study of the computability in the multi-agent case will be based on the
 750 modal tableaux approach developed by Halpern and Moses [23]. Assume
 751 $\text{card}(\text{Agt}) \geq 2$.

752 A set \mathcal{T} of formulas is said to be fully expanded if for all formulas φ in \mathcal{T}
 753 and for all formulas ψ in $SF(\varphi)$, either $\psi \in \mathcal{T}$, or $\neg\psi \in \mathcal{T}$. A propositional
 754 tableau is a set \mathcal{T} of formulas such that: (i) for all formulas φ , if $\neg\neg\varphi \in \mathcal{T}$
 755 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
 756 $\neg\psi \in \mathcal{T}$; (iii) for all formulas φ, ψ , if $\varphi \vee \psi \in \mathcal{T}$ then either $\varphi \in \mathcal{T}$, or
 757 $\psi \in \mathcal{T}$. A propositional tableau \mathcal{T} is said to be blatantly consistent iff for
 758 all formulas φ , either $\varphi \notin \mathcal{T}$, or $\neg\varphi \notin \mathcal{T}$.

759 A modal tableau is a structure of the form $\mathcal{T} = (W, R, L)$ where W is
 760 a nonempty set of states (with typical members denoted w, v , etc), R is a
 761 function associating a binary relation R_i on W to each $i \in \text{Agt}$ and L is a

762 function assigning to each $w \in W$ a blatantly consistent and fully expanded
 763 propositional tableau $L(w)$ such that for all $w \in W$: (i) for all formulas φ ,
 764 if $\neg K_i \varphi \in L(w)$ then there exists $v \in R_i(w)$ such that $\neg \varphi \in L(v)$; (ii) for
 765 all formulas φ , if $B_i \varphi \in L(w)$ then for all formulas ψ , if $\neg B_i \psi \in L(w)$
 766 then there exists $v \in R_i(w)$ such that either $\varphi \in L(v)$ and $\neg \psi \in L(v)$, or
 767 $\neg \varphi \in L(v)$ and $\psi \in L(v)$; (iii) for all formulas φ , if $K_i \varphi \in L(w)$ then for all
 768 $v \in (R_i \cup R_i^{-1})^*(w)$, $\varphi \in L(v)$ and $K_i \varphi \in L(v)$; (iv) for all formulas φ , if
 769 $B_i \varphi \in L(w)$ then for all $v \in (R_i \cup R_i^{-1})^*(w)$, $B_i \varphi \in L(v)$. For all formulas φ ,
 770 let $L^{-1}(\varphi) = \{w : w \in W \ \& \ \varphi \in L(w)\}$. We shall say that a modal tableau
 771 $\mathcal{T} = (W, R, L)$ is a modal tableau for a formula φ if $L^{-1}(\varphi) \neq \emptyset$.

772 Given a model $M = (W, R, N, V)$, let $\mathcal{T}' = (W', R', L')$ be defined by:
 773 $W' = W$, $R'_i = R_i$ for each $i \in \text{Agt}$, L' is the function assigning to each
 774 $w \in W'$ the propositional tableau $L'(w) = \{\varphi : M, w \models \varphi\}$. The proof that
 775 \mathcal{T}' is a modal tableau is easy.

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

778 Given a modal tableau $\mathcal{T} = (W, R, L)$, let $M' = (W', R', N', V')$ be
 779 defined by: $W' = W$, $R'_i = (R_i \cup R_i^{-1})^*$ for each $i \in \text{Agt}$, $N'_i(w) = \{(R_i \cup$
 780 $R_i^{-1})^*(w) \cap L^{-1}(\varphi) : B_i \varphi \in L(w)\}$ for each $w \in W$ and for each $i \in \text{Agt}$,
 781 V' is the function assigning to each $p \in \text{Atm}$ the subset $V'(p) = \{w : w \in$
 782 $W \ \& \ p \in L(w)\}$ of W' . The proof that \mathcal{T}' is a model is easy. Moreover, one
 783 can prove by induction on φ that for all $w \in W$, $\varphi \in L(w)$ iff $M', w \models \varphi$.

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

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

- 792 • For all given formulas, the above algorithm terminates and runs in poly-
 793 nomial space,
- 794 • for all given formulas φ , the algorithm returns “there is a modal tableau
 795 for φ ” iff there is a modal tableau for φ .

796 THEOREM 4. *If $\text{card}(\text{Agt}) \geq 2$ then satisfiability problem of LEK is*
 797 *PSPACE-complete.*

798 PROOF. Membership in PSPACE follows from the above discussion. Mean-
 799 while, PSPACE-hardness follows from the PSPACE-hardness of multi-agent
 800 epistemic logic. ■

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

806 COROLLARY 1. *The satisfiability problems of mono-agent $DLEK^{(+)}$ and mul-*
 807 *ti-agent DLEK are decidable.*

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

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

813 6. Translation of \mathcal{L}_{LEK} into an Ordinary Modal Language: The 814 Mono-Agent Case

815 We will show how one can reduce the semantics of LEK^+ to a purely rela-
 816 tional semantics, albeit with additional modalities. For simplicity, we will
 817 consider only the mono-agent case, and we will write K, B instead of K_1, B_1 .
 818 Simple changes in the arguments presented below would allow the reader to
 819 adapt our line of reasoning to LEK . Recall that in the mono-agent case, we
 820 may assume that in any LEK^+ -model (W, N, R, V) we have that R is the
 821 total relation, so we may omit it. Moreover, since $N(w)$ does not depend
 822 on $w \in W$ (given that it is assumed R -invariant), we may identify N with
 823 the set of neighbourhoods $N(w_0)$, where $w_0 \in W$ is arbitrary.

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

- 830 • To interpret $K\varphi$ at $w \in W$, we include an equivalence relation R_K linking
 831 w to all ‘real’ worlds. It does not matter how R_K behaves on neighbour-
 832 hoods, so on N we may just let R_K be the identity.

- 833 • To interpret $\mathbf{B}\varphi$ at $w \in W$, we include:
- 834 – a relation R_L linking each $w \in W$ to each $U \in N(w)$. Note in this
- 835 case that every point is related to every neighborhood, so we may
- 836 simply define $R_L = W \times N$, and
- 837 – a relation R_\square linking each $U \in N$ to its elements (i.e., $UR_\square w$ if and
- 838 only if $w \in U$).

839 This setup would suffice to interpret monotone neighborhood semantics [17].

840 However, for our *strict* interpretation, we need some additional structure.

841 Namely, to check if U is a witness for $\mathbf{B}\varphi$, we must check that every element

842 of U satisfies φ , but also that every element *not* in U satisfies $\neg\varphi$. We achieve

843 this by an additional accessibility relation R_\blacksquare , where $UR_\blacksquare w$ holds whenever

844 $w \notin U$; a similar trick has been used in [18, 20] to deal with the ‘inaccessible

845 worlds’ logic of Humberstone [26].

846 Let us now formalize these ideas. We begin by defining a new modal

847 language, \mathcal{L}_O , given by:

$$848 \quad \varphi, \psi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi \mid L\varphi \mid \square\varphi \mid \blacksquare\varphi$$

849 We will use the following abbreviations: $\hat{K}\varphi$ for $\neg K\neg\varphi$, $\hat{L}\varphi$ for $\neg L\neg\varphi$,

850 $\diamond\varphi$ for $\neg\square\neg\varphi$ and $\blacklozenge\varphi$ for $\neg\blacksquare\neg\varphi$. Let $\tau : \mathcal{L}_{\text{LEK}} \longrightarrow \mathcal{L}_O$ be the translation

851 function defined by

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

852 An \mathcal{L}_O -model is a structure of the form $(W, R_K, R_L, R_\square, R_\blacksquare, V)$ where

853 W is a nonempty set of possible worlds, R_K , R_L , R_\square and R_\blacksquare are binary

854 relations on W and $V : W \longrightarrow 2^{Atm}$ is a valuation function. The truth-

855 conditions are inductively defined as in ordinary modal logic. We will say

856 that the model $(W, R_K, R_L, R_\square, R_\blacksquare, V)$ is prestandard iff

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

857 If, furthermore, $R_\square \cap R_\blacksquare = \emptyset$ then we will say that the prestandard

858 model $(W, R_K, R_L, R_\square, R_\blacksquare, V)$ is standard. We will say that $\varphi \in \mathcal{L}_O$ is

859 prestandard-valid (respectively, standard-valid), in symbols $\models_{pr} \varphi$ (respec-
860 tively, $\models_{st} \varphi$), iff φ is true at every possible world of every prestandard
861 (respectively, standard) \mathcal{L}_O -model.

862 LEMMA 8. Let $M = (W, R_K, R_L, R_{\square}, R_{\blacksquare}, V)$ be a standard \mathcal{L}_O -model. Let
863 $s_0 \in W$. Let $M' = (W', N', V')$ be the LEK^+ -model such that

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

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

868 PROOF. By induction on φ . ■

869 LEMMA 9. Let $M = (W, N, V)$ be a LEK^+ -model such that, for any $w \in W$,
870 $N(w) \cap W = \emptyset$; this condition is easily obtained if we do not allow the
871 elements of W to be subsets of W . Let $M' = (W', R'_K, R'_L, R'_{\square}, R'_{\blacksquare}, V')$ be
872 the standard \mathcal{L}_O -model such that

- 873 • $W' = W \cup N$,
- 874 • $R'_K = (W \times W) \cup Id_N$,
- 875 • $R'_L = W \times N$,
- 876 • $R'_{\square} = \{(A, s) : A \in N \ \& \ s \in W \ \& \ s \in A\}$,
- 877 • $R'_{\blacksquare} = \{(A, s) : A \in N \ \& \ s \in W \ \& \ s \notin A\}$,
- 878 • $V' : p \mapsto V'(p) = V(p)$.

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

880 PROOF. By induction on φ . ■

881 PROPOSITION 5. Let $\varphi \in \mathcal{L}_{LEK}$. The following conditions are equivalent:

- 882 1. $\models_{LEK^+} \varphi$; 2. $\models_{st} \tau(\varphi)$.

883 PROOF. 1. \Rightarrow 2.) By Lemma 8.

884 2. \Rightarrow 1.) By Lemma 9. ■

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

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

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

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

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

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

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

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

903 PROOF. Left to the reader. ■

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

906 PROOF. (1. \Rightarrow 2.) By Proposition 6. (2. \Rightarrow 3.) By Lemma 10 and the
 907 Bounded Morphism Lemma in ordinary modal logic. (3. \Rightarrow 1.) By Proposi-
 908 tion 7. ■

909 7. Conclusion

910 Let's take stock. In the paper we have introduced DLEK, a logical theory
 911 of belief dynamics for resource-bounded agents inspired by existing psy-
 912 chological theories of human memory. We have provided decidability and

913 complexity results for DLEK as well as for its static fragment LEK. In addi-
914 tion, we have introduced the logics DLEK^+ , LEK^+ , which are suitable for
915 the mono-agent setting.

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

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

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

946 **Acknowledgements.** David Fernández-Duque's work was partially sup-
947 ported by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-
948 0002-02.

949 **References**

- 950 [1] ÅGOTNES, T., and N. ALECHINA, A logic for reasoning about knowledge of unawareness, *Journal of Logic, Language and Information* 23(2):197–217, 2014.
- 951 [2] ÅGOTNES, T., and N. ALECHINA, Semantics for dynamic syntactic epistemic logics, in *Proceedings of the Tenth International Conference on Knowledge Representation and Reasoning (KR 2006)*. AAAI Press, 2006, pp. 411–419.
- 952 [3] ÅGOTNES, T., P. BALBIANI, H. VAN DITMARSCH, and P. SEBAN, Group announcement logic, *Journal of Applied Logic* 8(1):62–81, 2014.
- 953 [4] ÅGOTNES, T., and M. WALICKI, Syntactic knowledge: A logic of reasoning, communication and cooperation, in *Proceedings of the Second European Workshop on Multi-Agent Systems (EUMAS 2004)*, 2004.
- 954 [5] ARTËMOV, S. N., The logic of justification, *Review of Symbolic Logic* 1(4):477–513, 2008.
- 955 [6] ALECHINA, N., B. LOGAN, and M. WHITSEY, A complete and decidable logic for resource-bounded agents, in *Proceedings of AAMAS 2014*. IEEE Computer Society, 2004, pp. 606–613.
- 956 [7] BADDELEY, A. D., and G. HITCH, Working memory, in G. H. Bower, editor, *The Psychology of Learning and Motivation: Advances in Research and Theory*. Academic Press, 1974, pp. 47–89.
- 957 [8] BALTAG, A., and L. S. MOSS, Logics for epistemic programs, *Synthese*, 139(2):165–224, 2004.
- 958 [9] BALTAG, A., and S. SMETS, A qualitative theory of dynamic interactive belief revision, in *Proceedings of LOFT 7*. Amsterdam University Press, 2008, pp. 13–60.
- 959 [10] COWAN, N., 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*. Cambridge University Press, 1999, pp. 62–101.
- 960 [11] DUC, H. N., Reasoning about rational, but not logically omniscient, agents, *Journal of Logic and Computation*, 7(5):633–648, 1997.
- 961 [12] ELGOT-DRAPKIN, J., 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*. Morgan Kaufmann, 1987, pp. 349–357.
- 962 [13] ELGOT-DRAPKIN, J., 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.
- 963 [14] ELGOT-DRAPKIN, J., S. KRAUS, M. MILLER, M. NIRKHE, and D. PERLIS, Active logics: A unified formal approach to episodic reasoning, *Technical Report CS-TR-4072*, 1999.
- 964 [15] FAGIN, R., and J. Y. HALPERN, Belief, awareness, and limited reasoning, *Artificial Intelligence*, 34:39–76, 1988.
- 965 [16] FAGIN, R., J. Y. HALPERN, Y. MOSES, and M. Y. VARDI, *Reasoning About Knowledge*. MIT Press, 1995.
- 966 [17] GASQUET, O., and A. HERZIG, Translation-based deduction methods for modal logics, in *Proceedings of the 5th International Conference on Information Processing and*

- 992 *Management of Uncertainty in Knowledge-Based Systems (IPMU'94)*. Springer, 1995,
993 pp. 399–408.
- 994 [18] GASQUET, O., and A. HERZIG, Translating inaccessible worlds logic into bimodal logic,
995 in *Proceedings of the European Conference on Symbolic and Quantitative Approaches*
996 *to Reasoning and Uncertainty (ECSQARU '93)*. Springer, 1993, pp. 145–150.
- 997 [19] GOLDMAN, A., Theory of mind, in E. Margolis, R. Samuels, and S. Stich, editors,
998 *Oxford Handbook of Philosophy and Cognitive Science*, Oxford University Press, 2012.
- 999 [20] GORANKO, V., Completeness and incompleteness in the bimodal base $\mathcal{L}(R, -R)$, in
1000 *Mathematical Logic*. Springer, 1990, pp. 311–326.
- 1001 [21] GRANT, J., S. KRAUS, and D. PERLIS, A logic for characterizing multiple bounded
1002 agents *Autonomous Agents and Multi-Agent Systems*, 3(4):351–387, 2000.
- 1003 [22] GROSSI, D., and F. R. VELÁZQUEZ-QUESADA, Twelve angry men: A study on the
1004 fine-grain of announcements, in *Proceedings of LORI 2009, Volume 5834 of LNCS*.
1005 Springer, 2009, pp. 147–160.
- 1006 [23] HALPERN, J., and Y. MOSES, A guide to completeness and complexity for modal
1007 logics of knowledge and belief *Artificial Intelligence*, 54:319–379, 1992.
- 1008 [24] HALPERN, J. Y., and L. C. RÊGO, Reasoning about knowledge of unawareness, *Games*
1009 *and Economic Behavior* 67(2):503–525, 2009.
- 1010 [25] HINTIKKA, J., Impossible possible worlds vindicated, *Journal of Philosophical Logic*
1011 4:475–484, 1975.
- 1012 [26] HUMBERSTONE, I. L., Inaccessible worlds, *Notre Dame Journal of Formal Logic*
1013 24(3):346–352, 1983.
- 1014 [27] KAHNEMAN, D., and A. TVERSKY., Variants of uncertainty, *Cognition* 11:143–157,
1015 1982.
- 1016 [28] KONOLIGE, K., What awareness isn't: A sentential view of implicit and explicit belief,
1017 in *Proceedings of the 1st Conference on Theoretical Aspects of Reasoning About Knowl-*
1018 *edge (TARK'86)*. Morgan Kaufmann, 1986, pp. 241–250.
- 1019 [29] JAGO, M., Epistemic Logic for Rule-Based Agents, *Journal of Logic, Language and*
1020 *Information* 18(1):131–158, 2009.
- 1021 [30] LAIRD, J. E., *The Soar Cognitive Architecture*. MIT Press, 2012.
- 1022 [31] LEVESQUE, H., A logic of implicit and explicit belief, in *Proceedings of AAAI-84*.
1023 AAAI Press, 1984, pp. 198–202.
- 1024 [32] DE LIMA, T., Alternating-time temporal dynamic epistemic logic, *Journal of Logic*
1025 *and Computation* 24(6):1145–1178, 2014.
- 1026 [33] LOVETT, M. C., L. M. REDER, and C. LEBIERE, Modeling working memory in a
1027 unified architecture: An ACT-R perspective, in A. Miyake and P. Shah, editors, *Mod-*
1028 *els of Working Memory: Mechanisms of Active Maintenance and Executive Control*.
1029 Cambridge University Press, 1999, pp. 135–182.
- 1030 [34] LUTZ, C., Complexity and succinctness of public announcement logic, in *Proceedings*
1031 *of AAMAS 2006*. ACM Press, 2006, pp. 137–144.
- 1032 [35] VAN BENTHEM, J., and E. PACUIT, Dynamic Logics of Evidence-Based Beliefs, *Journal*
1033 *of Logic, Language and Information* 99(1-3):61–92, 2011.
- 1034 [36] VAN DITMARSCH, H., and T. FRENCH, Semantics for knowledge and change of aware-
1035 ness, *Journal of Logic, Language and Information* 23(2):169–195, 2014.

- 1036 [37] VAN DITMARSCH, H., W. VAN DER HOEK, and B. KOOI, *Dynamic Epistemic Logic*.
1037 Springer, 2008.
- 1038 [38] VELÁZQUEZ-QUESADA, F. R., Dynamic Epistemic Logic for Implicit and Explicit
1039 Beliefs, *Journal of Logic, Language and Information* 23(2):107–140, 2014.
- 1040 [39] VELÁZQUEZ-QUESADA, F. R., Explicit and implicit knowledge in neighbourhood mod-
1041 els, in *Proceedings of LORI 2013, Volume 8196 of LNCS*. Springer, 2013, pp. 239–252.
- 1042 [40] YOUNG, R. M., and R. L. LEWIS, The Soar cognitive architecture and human working
1043 memory, in A. Miyake and P. Shah, editors, *Models of Working Memory: Mechanisms*
1044 *of Active Maintenance and Executive Control*. Cambridge University Press, 1999, pp.
1045 135–182.

1046 P. BALBIANI, E. LORINI
1047 IRIT
1048 Toulouse University
1049 Toulouse
1050 France
1051 `lorini@irit.fr`

1052
1053 P. BALBIANI
1054 `balbiani@irit.fr`

1055 P. BALBIANI, D. FERNÁNDEZ-DUQUE
1056 Department of Mathematics
1057 Ghent University
1058 Ghent
1059 Belgium
1060 `David.FernandezDuque@UGent.be`