Rend. Istit. Mat. Univ. Trieste
Vol. 53 (2021), Art. No. 25, pp. 1–25 (electronic)
DOI: 10.13137/2464-8728/33311

Epistemic logics for modeling group dynamics of cooperative agents, and aspects of Theory of Mind

Stefania Costantini

Eugenio has been for me a mentor and a friend. I wish to dedicate to him some verses by the Italian poet Camillo Sbarbaro, dedicated to his father, that properly and synthetically express my feelings for Eugenio (though he is not my father): "Padre, se anche tu non fossi il mio padre, se anche fossi a me un estraneo, per te stesso egualmente t'amerei."

Abstract. *Logic has been proved useful to model various aspects of the reasoning process of agents and Multi-Agent Systems (MAS). In this paper, we report about a line of work carried on in cooperation with Andrea Formisano (former Eugenio's Ph.D. student) and Valentina Pitoni, to explore some social aspects of such systems. The aim is to formally model (aspects of) the group dynamics of cooperative agents. We have proposed a particular logical framework (the Logic of "Inferable" L-DINF), where a group of cooperative agents can jointly perform actions. I.e., at least one agent of the group can perform the action, either with the approval of the group or on behalf of the group. We have been able to take into consideration actions'* cost*, and the preferences that each agent may have for what concerns performing each action. Our focus is on: (i) explainability, i.e., the syntax of our logic is especially devised to make it possible to transpose a proof into a natural language explanation, in the perspective of trustworthy Artificial Intelligence (AI); (ii) the capability to construct and execute joint plans within a group of agents; (iii) the formalization of aspects of the Theory of Mind, which is an important social-cognitive skill that involves the ability to attribute mental states, including emotions, desires, beliefs, and knowledge both one's own and those of others, and to reason about the practical consequences of such mental states; this capability is very relevant when agents have to interact with humans, and in particular in robotic applications; (iv) connection between theory and practice, so as to make our logic actually usable by systems' designers. In this paper, we summarize our past work and propose some discussions, possible extensions and considerations.*

Keywords: Epistemic logic, agents and multi-agent systems, theory of mind.
MS Classification 2020: 03-02, 03B42, 03B45, 03B70.

1. Introduction

The metaphor adopted in Artificial Intelligence (AI) to model societies whose members are to some extent cooperative towards each other is that of agents and Multi-Agent Systems (MAS). Agents, in fact, can be either cooperative or competitive. In the cooperative case, the agents pursue common goals, can share to some extent their private knowledge, and can expect benevolent intentions from other agents. Agents in a competitive MAS setting have instead non-aligned goals, and individual agents seek only to maximize their own gains. In our work we have focused on the former case[1]. To achieve better results via cooperation, agents belonging to a MAS must be able to reason about what a group of agents can do, because it is often the case that a group can fulfil objectives that are out of reach for the single agent: each participating agent, in fact, may not in general be able to solve a whole problem or reach an overall goal by itself, rather, often it can only cope with small subproblems/subgoals, for which it has the required competences. The overall result/goal is, in general, accomplished by means of cooperation with other agents. In the course of the cooperation, an agent may have to bid for solving some aspect of the problem or perform some action instead of some other one, or to negotiate with other agents for the distribution of tasks. Several agent-oriented programming languages and systems exist, many of them based upon computational logic (cf., e.g., [3, 4, 19] for recent surveys on such languages), and thus endowed (at least in principle) with a logical semantics.

Many kinds of logical frameworks can be found in the literature, which try to emulate cognitive aspects of human beings, also from the cooperative point of view. In our past work (joint work with Andrea Formisano and Valentina Pitoni) [10, 11, 13], we defined the new Logic of "Inferable", called *L-DINF*, as an extension of an existing logic by Lorini & Balbiani [2], which considers an agent in the context of some cooperative group(s). We introduced conditions for the cooperative executability of physical actions taking into account feasibility, costs and budget, and also preferences of single agents concerning their willingness to execute actions that they are allowed to do.

A relevant feature of our approach is that the conditions concerning whether an agent (and thus its group) is allowed to execute some action, and to which extent it is willing to perform it, are not specified in the logical theory defining an agent: rather, we envisage separate modules from which the agent's logical theory "inputs" the results. Such modules might be specified in some other logic or also, pragmatically, via pieces of code whenever, e.g., feasibility of actions should be verified according to agents' environmental conditions.

---

[1] In practice however, agents in a system can show a wide range of behaviors that may either be cooperative or competitive, depending on their present circumstances.

The rationale of this approach can be exposed as follows. On the one hand, logic is a good tool to express the semantics underlying (aspects of) agent-oriented programming languages, as it allows properties of the behaviour of an agent or a group of agents to be expressed and proved. To this aim however, it is important to keep the complexity of the logic low enough to be practically manageable. Modularity is an important property to ensure, as it allows programmers to better organize the definition of the application at hand, and allows an agent-systems' definition to be more flexible and customizable. As notable examples, in [14] it is shown how an agent behaviour can significantly change by leaving its 'main' definition unchanged, while modifying only its communication modalities, i.e., which kind of messages, and from/to whom, the agent is available to manage. In [22], it is shown that a different sequencing and duration of agent's activities determines a very different 'external' behaviour, again over the same main program. Moreover, modularity can be an advantage for explainability, in the sense of making the explanation itself modular.

So, our approach aims to combine the rigour of logic and the flexibility of modularity. We allow one to define in a separate way which actions are allowed for each agent to perform at each stage, and with which degree of preference. A programmer will then be able to define suitable pieces of code specifying where, when, and why each action is indeed allowed, and, possibly, which is the 'rationale' of a certain degree of preference of an agent in performing an action. Thus, modular changes to the conditions for actions to be enabled and to the reasons for an agent's preference to perform or not an action, may affect in a relevant way the behaviour of both an agent and the group(s) to which it belongs.

In the original formulation of *L-DINF*, we considered the notion of executability of agents' inferential actions (also called *mental* actions). In our approach, when an agent belongs to a group, if that agent is not able to perform an intended action which in principle it should be able to perform, it may be supported by its group. The reason for not being able can be that an action may require resource consumption (and hence, involve a *cost*). So, in order to execute an action the agent must possess the necessary *budget*, or borrow it from the group. We then extended the logic by introducing further possibilities of solidarity between the members of a cooperative group of agents, in particular to support each other in performing actions in place of some other agent who is not enabled or not wishing to do that itself. In this extension, the reason of not being able to perform an action can be that the agent is not allowed to perform that action in the present state; moreover, the agent might be allowed and still not willing to execute that action.

'Our' agents are *aware* of themselves, of the group they belong to, and possibly of other groups. Since we assume that agents belonging to a group are cooperative with respect to action execution, an action can be executed

by the group if at least one agent therein is able and allowed and willing to execute it, and the group can bear (in some way) the cost. In case more agents can perform an action, the one who is best willing will be selected, based on a concept of preference.

In [10], we have thoroughly discussed the relationship of logic *L-DINF* with related work, emphasizing that this logic draws inspiration from the concepts of Theory of Mind [20] and of Social Intelligence [21].

We are also indebted to [18], concerning the point of view that an agent reaches a certain belief state by performing inferences, and that making inferences takes time. We tackled the issue of time in previous work, discussed in [9, 12, 23]. Differently from these works however, in *L-DINF* inferential actions are represented both at the syntactic level, via dynamic operators in the DEL style, and at a semantic level as neighborhood-update operations. Also, *L-DINF* enables an agent to reason on executability of inferential actions. We try to introduce (even though the formalization is not complete yet) the concept that an action may take a certain number of steps in order to be enabled or suitable for execution.

One relevant aim of this work is to take into account the relationship between the semantic and the practical aspects of agents' specification and engineering, which is often neglected, thus leading to an undesirable (in our view) detachment between theory and practice. Therefore, we provide action-related reserved syntax, specifying explicitly what an agent can do, does, and has done, or to which degree it is willing to perform the feasible actions. For some of these expressions we assume a "semantic attachment" to the external environment in which an agent will be situated, i.e., some kind of sensor/actuator device which actually performs actions, which is opaque at the logical level but in our view still needs representation (we were inspired by the discussion, that goes way back to a long time ago, proposed in [25]). This approach is aimed at making the formalization more complete and comprehensible for developers, and intends to improve explainability of an agent's operation, by being able to translate logical proofs into natural language expressions that are intelligible to human users, also thanks to the explicit standard representations of action-related aspects.

A long-term goal is to formalize in our logic aspects of the "Theory of Mind" (ToM), which is an important social-cognitive skill that involves the ability to attribute mental states, including emotions, desires, beliefs, and knowledge, to oneself and to others, and to reason about the practical consequences of such mental states. Theory of Mind (ToM), developed originally by Philosophers and Psychologists, is starting to be applied to robotics, and some suitable logics are being developed [16]. In fact, with the arrival of "service robots" devised to support users in their everyday tasks (e.g., in eHealth applications, robots may support on the one hand patients, by reminding them to take their

medicines and by providing advice and reassurance, and on the other hand doctors, by constantly monitoring the user's vital parameters, and by creating alerts whenever necessary). In order to render these robots acceptable and even appreciated by users, they will have to be programmed so as to mimic basic social skills and behave in a socially acceptable manner, which means that their behaviour is to some extent predictable by the user, as it conforms to social standards. Theory of Mind is linked to affective computing (which is a set of techniques able to elicit a human's emotional states from physical signs), to enable the system to respond intelligently to human emotional feedback, and to enhance ToM activities by providing it with perceptions related to the user's emotional signs.

The paper is organized as follows. Section 2 introduces syntax and semantics of *L-DINF*, together with an axiomatization of the proposed logical system. In Section 3 we present an example of application of the new logic. Canonical models, and the proof of strong completeness of the logic, are discussed in Section 4. In Section 5 we introduce interesting possible future developments: namely, we discuss the possibility of formalizing the fact that a goal is meant to be reached (or has been reached) within a certain number of steps, and we outline how to extend our logic so as to model significant aspects of the Theory of Mind. Finally, in Section 6 we conclude.

2. Logical framework

L-DINF is a logic which consists of a static component and a dynamic one. The static component, called *L-INF*, is a logic of explicit beliefs and background knowledge. The dynamic component, called *L-DINF*, extends the static one with dynamic operators capturing the consequences of the agents' inferential actions on their explicit beliefs as well as a dynamic operator capturing what an agent can conclude by performing some inferential action in her repertoire.

2.1. Syntax of *L-DINF*

Let $Atm = \{p, q, \ldots\}$ be a countable set of atomic propositions. 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. A subset Atm_A of the atomic propositions represent the physical actions that an agent can perform, including "active sensing" actions (e.g., "let's check whether it rains", "let's measure the temperature"). Below, let $\phi_A \in Atm_A$. Let $d, d_{max} \in \mathbb{N}$ where $0 \leq d \leq d_{max}$. Moreover, let Agt be a set of agents. The language of *L-DINF*, denoted by

$\mathcal{L}_{L\text{-}DINF}$, is defined by the following grammar in Backus-Naur form:

$$\begin{array}{lll} \varphi, \psi & ::= & p \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathbf{B}_i\,\varphi \mid \mathbf{K}_i\,\varphi \mid \\ & & \mathit{do}_i(\phi_A) \mid \mathit{do}_i^P(\phi_A) \mid \mathit{can\_do}_i(\phi_A) \mid \mathit{pref\_do}_i(\phi_A, d) \mid \\ & & \mathit{do}_G(\phi_A) \mid \mathit{do}_G^P(\phi_A) \mid \mathit{can\_do}_G(\phi_A) \mid \mathit{pref\_do}_G(i, \phi_A) \mid \\ & & \mathit{intend}_i(\phi_A) \mid \mathit{intend}_G(\phi_A) \mid \mathit{exec}_i(\alpha) \mid \mathit{exec}_G(\alpha) \mid [G : \alpha]\,\varphi \\ \alpha & ::= & \vdash(\varphi,\psi) \mid \cap(\varphi,\psi) \mid \downarrow(\varphi, \psi) \end{array}$$

where p ranges over Atm and $i \in \mathit{Agt}$. (Other Boolean operators are defined from $\neg$ and $\wedge$ in the standard manner.)

The language of *inferential actions* (or "mental actions") of type α is denoted by $\mathcal{L}_{\mathsf{ACT}}$. Plainly, the static part *L-INF* of *L-DINF*, includes only those formulas not having sub-formulas of type α, namely, no inferential operation is admitted.

Notice the expression $\mathit{intend}_i(\phi_A)$, where it is required that $\phi_A \in \mathit{Atm}_A$. This expression indicates the intention of agent i to perform action ϕ_A in the sense of the BDI agent model [24]. This intention can be part of an agent's knowledge base from the beginning, or it can be derived later. In this paper, we do not cope with the formalization of BDI, for which the reader may refer, e.g., to [17]. So, we will treat intentions rather informally, assuming also that $\mathit{intend}_G(\phi_A)$ holds whenever all agents in group G intend to perform action ϕ_A.

The expressions $\mathit{can\_do}_i(\phi_A)$ and $\mathit{pref\_do}_i(\phi_A, d)$ (where it is required that $\phi_A \in \mathit{Atm}_A$) are closely related to $\mathit{do}_i(\phi_A)$. In fact, $\mathit{can\_do}_i(\phi_A)$ is to be seen as an enabling condition, indicating that agent i is enabled to execute action ϕ_A, while instead $\mathit{pref\_do}_i(\phi_A, d)$ indicates the level d of preference/willingness of agent i to perform that action. $\mathit{pref\_do}_G(i, \phi_A)$ indicates that agent i exhibits the *maximum level* of preference on performing action ϕ_A within all group members. Notice that, if a group of agents intends to perform an action ϕ_A, this will entail that the entire group intends to do ϕ_A, that will be enabled to be actually executed only if at least one agent $i \in G$ can do it, i.e., it can derive $\mathit{can\_do}_i(\phi_A)$.

$\mathit{do}_i(\phi_A)$, where again it is required that $\phi_A \in \mathit{Atm}_A$, indicates *actual execution* of action ϕ_A by agent i, automatically recorded by the new belief $\mathit{do}_i^P(\phi_A)$ (postfix "P" standing for "past" action). In fact, do and do^P (and similarly do_G and do_G^P) are not axiomatized, as they are realized by what has been called in [25] a *semantic attachment*, i.e., a procedure which connects an agent with its external environment in a way that is unknown at the logical level. The axiomatization concerns only the relationship between doing and being enabled to do.

Unlike explicit beliefs, an agent's background knowledge is assumed to satisfy *omniscience* principles, such as closure under conjunction and known implication, and closure under logical consequence, and introspection. More specifically, $\mathbf{K}_i$ is nothing but the well-known S5 modal operator often used

to model/represent knowledge. The assumption that background knowledge is closed under logical consequence is justified because we conceive it as a kind of stable reliable *knowledge base*. The background knowledge, in our view, includes facts (formulas) known by the agent from the beginning, plus facts the agent decided to store in her long-term memory (by means of some decision mechanism not treated here, after having processed them in her working memory), as well their logical consequences. To our present aims however, we assume background knowledge to be irrevocable, in the sense of being stable over time.

A formula of the form $[G : \alpha]\,\varphi$, with $G \in 2^{Agt}$, and where α must be an inferential action, states that "φ holds after action α has been performed by at least one of the agents in G, and all agents in G have common knowledge about this fact".

REMARK 2.1. If an inferential action is performed by an agent $i \in G$, the other agents belonging to the same group G have full visibility of this action and, therefore, as we suppose agents to be cooperative, it is as if they had performed the action themselves.

Borrowing from [1], we distinguish four types of mental actions α which allow us to capture some of the dynamic properties of explicit beliefs and background knowledge: $\vdash(\varphi,\psi)$, $\cap(\varphi,\psi)$ and $\downarrow(\varphi,\psi)$. These actions characterize the basic operations of forming explicit beliefs via inference:

- $\downarrow(\varphi,\psi)$ is the inferential action which consists in inferring ψ from φ in case φ is believed and, according to agent's background knowledge, ψ is a logical consequence of φ. In other words, by performing this inferential action, 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 believing ψ;

- $\cap(\varphi,\psi)$ is the inferential action which closes the explicit belief φ and the explicit belief ψ under conjunction. In other words, $\cap(\varphi,\psi)$ characterizes the inferential action of deducing $\varphi \wedge \psi$ from the explicit belief φ and the explicit belief ψ;

- $\dashv(\varphi,\psi)$ is the inferential action that performs a simple form of "belief revision", i.e., removes ψ from the working memory in case φ is believed and, according to agent's background knowledge, $\neg\psi$ is logical consequence of φ. Both ψ and φ are required to be ground atoms.

- $\vdash(\varphi,\psi)$ is the inferential action which adds ψ to the working memory in case φ is believed and, according to agent's working memory, ψ is logical consequence of φ. This last action operates directly on the working memory without retrieving anything from the background knowledge.

Formulas of the forms $exec_i(\alpha)$ and $exec_G(\alpha)$ express executability of inferential actions either by agent i, or by group G of agents (which is a consequence of any of the group members being able to execute the action). It has to be read as: "α is an inferential action that agent i (resp. an agent in G) can perform".

REMARK 2.2. In the mental actions $\vdash(\varphi,\psi)$ and $\downarrow(\varphi,\psi)$, the formula ψ which is inferred and asserted as a new belief can be $can\_do_i(\phi_A)$ or $do_i(\phi_A)$, which denotes the actual [possibility of] execution of physical action ϕ_A. In fact, we assume that when inferring $do_i(\phi_A)$ (from $can\_do_i(\phi_A)$ and possibly other conditions) the action is actually executed (and the corresponding belief $do_i^P(\phi_A)$ is asserted, possibly augmented with a time-stamp). Actions are supposed to succeed by default, in case of failure a corresponding failure event will be perceived by the agent. The do_i^P beliefs constitute a *history* of the agent's operation, so they might be useful for the agent to reason about its own past behavior, and/or, importantly, they may be useful to provide *explanations* to human users.

REMARK 2.3. Explainability in our approach can be directly obtained from proofs. Let us assume for simplicity that inferential actions can be represented in infix form as $\varphi_n\ OP\ \varphi_{n+1}$. Also, $exec_i(\alpha)$ means that the mental action α is executable by agent i and is indeed executed. If, for instance, the user wants an explanation of why the action ϕ_A has been performed, the system can respond by exhibiting the proof that has lead to ϕ_A, put in the explicit form:
$(exec_i(\varphi_1 OP_1\ \varphi_2) \wedge \ldots\ \wedge (exec_i(\varphi_{n-1} OP_n\ \varphi_n) \wedge$
$(exec_i(\varphi_n OP_n\ can\_do_i(\phi_A)) \wedge (exec_i(intend_i(\phi_A) \wedge can\_do_i(\phi_A) \vdash do_i(\phi_A))$
where each OP_i is one of the (mental) actions discussed above. The proof can possibly be translated into natural language, and declined either top-down or bottom-up.

As said in the Introduction, we model agents which, to execute an action, may have to pay a cost, so they must have a consistent budget available. Our agents, moreover, are entitled to perform only those physical actions that they conclude they can do. In our approach, agents belong to groups (where the smallest possible group is the single agent), where agents belonging to a group are by definition cooperative. With respect to action execution, an action can be executed by the group if at least one agent in the group is able to execute it, and the group has the necessary budget available, sharing the cost according to some policy. The cooperative nature of our agents manifests itself also in selecting, among the agents that are able to do some physical action, the one(s) which best prefer to perform that action. We do not have introduced costs and budget, feasibility of actions and willingness to perform them in the language for two reasons: to keep the complexity of the logic reasonable, and to make such features customizable in a modular way. In fact, we intend to use this logic in practice, to formalize memory in DALI agents, where DALI is a

logic-based agent-oriented programming language [5, 6, 15]. So, computational effectiveness and modularity are crucial. Assuming that agents share the cost is reasonable when agents share resources, or cooperate to a common goal, as discussed, e.g., in [7, 8]. In fact, by making the assumption that agents are cooperative, we also assume that they are aware of and agree with the cost-sharing policy. So, as seen below, costs and budget are coped with at the semantic level. Variants of the logic can be easily worked out, where the modalities of cost sharing are different from the one shown here, where the group members share an action cost in equal parts. Below we indicate which are the points that should be modified to change the cost-sharing policy. Moreover, for brevity we introduce a single budget function, and thus, implicitly, a single resource to be spent. Several budget functions, each one concerning a different resource, might be plainly defined.

2.2. Semantics

Definition 2.4 introduces the notion of *L-INF model*, which is then used to introduce semantics of the static fragment of the logic. As before let Agt be the set of agents.

DEFINITION 2.4. *A model is a tuple* $M = (W, N, \mathcal{R}, E, B, C, A, P, V)$ *where:*

- W *is a set of worlds (or situations);*
- $\mathcal{R} = \{R_i\}_{i \in Agt}$ *is a collection of equivalence relations on* W*:* $R_i \subseteq W \times W$ *for each* $i \in Agt$*;*
- $N : Agt \times W \longrightarrow 2^{2^W}$ *is a neighborhood function such that, for each* $i \in Agt$*, each* $w, v \in W$*, and each* $X \subseteq W$ *these conditions hold:*

  (**C1**) *if* $X \in N(i, w)$ *then* $X \subseteq \{v \in W \mid wR_iv\}$,

  (**C2**) *if* wR_iv *then* $N(i, w) = N(i, v)$*;*
- $E : Agt \times W \longrightarrow 2^{\mathcal{L}_{\mathsf{ACT}}}$ *is an executability function of mental actions such that, for each* $i \in Agt$ *and* $w, v \in W$*, it holds that:*

  (**D1**) *if* wR_iv *then* $E(i, w) = E(i, v)$*;*
- $B : Agt \times W \longrightarrow \mathbb{N}$ *is a budget function such that, for each* $i \in Agt$ *and* $w, v \in W$*, the following holds*

  (**E1**) *if* wR_iv *then* $B(i, w) = B(i, v)$*;*
- $C : Agt \times \mathcal{L}_{\mathsf{ACT}} \times W \longrightarrow \mathbb{N}$ *is a cost function such that, for each* $i \in Agt$*,* $\alpha \in \mathcal{L}_{\mathsf{ACT}}$*, and* $w, v \in W$*, it holds that:*

(**F1**) *if* wR_iv *then* $C(i, \alpha, w) = C(i, \alpha, v)$*;*

- $A : Agt \times W \longrightarrow 2^{Atm_A}$ *is an executability function for physical actions, such that, for each* $i \in Agt$ *and* $w, v \in W$*, it holds that:*

  (**G1**) *if* wR_iv *then* $A(i, w) = A(i, v)$*;*

- $P : Agt \times W \times Atm_A \longrightarrow Int$ *is a preference function for physical actions* α *such that, for each* $i \in Agt$ *and* $w, v \in W$*, it holds that:*

  (**H1**) *if* wR_iv *then* $P(i, w, \alpha) = P(i, v, \alpha)$*;*

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

To simplify the notation, let $R_i(w)$ denote the set $\{v \in W \mid wR_iv\}$, for $w \in W$. The set $R_i(w)$ identifies the situations that agent i considers possible at world w: i.e., it represents the *epistemic state* of agent i at 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.

While $R_i(w)$ concerns background knowledge, $N(i, w)$ is the set of all facts that agent i explicitly believes at world w, a fact being identified with a set of worlds. Hence, if $X \in N(i, w)$ then, the agent i has fact X under the focus of her attention and believes it. We say that $N(i, w)$ is the explicit *belief set* of agent i at world w.

The executability of inferential actions is determined by function E. For an agent i, $E(i, w)$ is the set of inferential actions that agent i can execute at world w. The value $B(i, w)$ is the budget the agent has available to perform inferential actions. Similarly, the value $C(i, \alpha, w)$ is the cost to be paid by agent i to execute the inferential action α in the world w.

The executability of physical actions is determined by function A. For an agent i, $A(i, w)$ is the set of physical actions that agent i can execute at world w.

The agent's preference on executability of physical actions is determined by function P. For an agent i, and a physical action α, $P(i, w, \alpha)$ is an integer value d indicating the degree of willingness of agent i to execute such action at world w.

Constraint (**C1**) imposes that agent i can have explicit in her mind only facts which are compatible with her current epistemic state. Moreover, according to constraint (**C2**), if a world v is compatible with the epistemic state of agent i at world w, then agent i should have the same explicit beliefs at w and v. In other words, if two situations are equivalent as concerns background knowledge, then they cannot be distinguished through the explicit belief set. This

aspect of the semantics can be extended in future work to allow agents to make plausible assumptions. Analogous properties are imposed by constraints (**D1**), (**E1**), and (**F1**). Namely, (**D1**) imposes that agent i always knows which inferential actions she can perform and those she cannot. (**E1**) states that agent i always knows the available budget in a world (potentially needed to perform actions). (**F1**) determines that agent i always knows how much it costs to perform an inferential action. (**G1**) and (**H1**) determine that an agent i always knows which physical actions she can perform and those she cannot, and with which degree of willingness.

Truth values for formulas of *L-DINF* are inductively defined as follows. Given a model $M = (W, N, \mathcal{R}, E, B, C, A, P, V)$, $i \in Agt$, $G \subseteq Agt$, $w \in W$, and a formula $\varphi \in \mathcal{L}_{L\text{-}INF}$, we introduce the following shorthand notation:

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

whenever $M, v \models \varphi$ is well-defined (see below). Then, we set:

- $M, w \models p$ iff $p \in V(w)$
- $M, w \models exec_i(\alpha)$ iff $\alpha \in E(i, w)$
- $M, w \models exec_G(\alpha)$ iff there exists $i \in G$ with $\alpha \in E(i, w)$
- $M, w \models can\_do_i(\phi_A)$ iff $\alpha \in A(i, w)$
- $M, w \models can\_do_G(\phi_A)$ iff there exists $i \in G$ with $\alpha \in A(i, w)$
- $M, w \models pref\_do_i(\phi_A, d)$ iff $\phi_A \in A(i, w)$ and $P(i, w, \phi_A) = d$
- $M, w \models pref\_do_G(i, \phi_A)$ iff $M, w \models pref\_do_i(\phi_A, d)$ and $d = \max\{P(j, w, \phi_A) | j \in G \wedge \phi_A \in A(j, w)\}$
- $M, w \models \neg\varphi$ iff $M, w \not\models \varphi$
- $M, w \models \varphi \wedge \psi$ iff $M, w \models \varphi$ and $M, w \models \psi$
- $M, w \models \mathbf{B}_i\,\varphi$ iff $||\varphi||^M_{i,w} \in N(i, w)$
- $M, w \models \mathbf{K}_i\,\varphi$ iff $M, v \models \varphi$ for all $v \in R_i(w)$

As seen above, a physical action can be performed by a group of agents if at least one agent of the group can do it, and the level of preference for performing this action is set to the maximum among those of the agents enabled to do this action. For any inferential action α performed by any agent i, we set:

- $M, w \models [G : \alpha]\varphi$ iff $M^{[G:\alpha]}, w \models \varphi$

where we put $M^{[G:\alpha]} = \langle W; N^{[G:\alpha]}, \mathcal{R}, E, B^{[G:\alpha]}, C, A, P, V\rangle$, representing the fact that the execution of an inferential action α affects the sets of beliefs of agent i and modifies the available budget. Such operation can add new beliefs by direct perception, by means of one inference step, or as a conjunction of previous beliefs. Hence, when introducing new beliefs (i.e., performing mental actions), the neighborhood must be extended accordingly.

A key aspect in the definition of the logic is the following, which states under which conditions, and by which agent(s), an action may be performed.

$$\mathit{enabled}_w(G,\alpha) \leftrightarrow \exists j \in G\,(\alpha \in E(j,w) \wedge \tfrac{C(j,\alpha,w)}{|G|} \leq \min_{h\in G} B(h,w)).$$

This condition, as defined above, expresses the fact that an an inferential action is enabled when: at least one agent can perform it; and the "payment" due by each agent, obtained by dividing the action's cost equally among all agents of the group, is within each agent's available budget. In case more than one agent in G can execute an action, we implicitly assume that the agent j performing the action is the one corresponding to the lowest possible cost. Namely, j is such that $C(j,\alpha,w) = \min_{h\in G} C(h,\alpha,w)$. This definition reflects a parsimony criterion reasonably adoptable by cooperative agents sharing a crucial resource such as, e.g., energy or money. Other choices might be viable, so variations of this logic can be easily defined simply by devising some other enabling condition and, possibly, introducing differences in neighborhood update. Notice that the definition of the enabling function basically specifies the "**role**" that agents take while concurring with their own resources to actions' execution. Also, in case of specification of different resources, different corresponding enabling functions should be defined.

Our contribution to modularity is that functions A and P, i.e., executability of physical actions and preference level of an agent concerning physical action execution are not meant to be built-in. Rather they can be defined via separate sub-theories, possibly defined using different logics, or, in a practical approach, via pieces of code. This approach can be extended to function C, i.e., the cost of mental actions instead of being fixed (like in our previous work) may vary and computed upon need.

2.3. Belief Update

In this kind of logic, updating an agent's beliefs accounts to modify the neighborhood of the present world. The updated neighborhood $N^{[G:\alpha]}$ resulting from execution of a mental action α by a group of agents is as follows. A key point is that of the update of each agent's budget, which decreases when part of it is spent to perform α.

$$N^{[G:\downarrow(\psi,\chi)]}(i,w) = \begin{cases} N(i,w) \cup \{||\chi||^M_{i,w}\} & \text{if } i \in G \text{ and } \mathit{enabled}_w(G,\downarrow(\psi,\chi)) \\ & \text{and } M,w \models \mathbf{B}_i\psi \wedge \mathbf{K}_i(\psi \rightarrow \chi) \\ N(i,w) & \text{otherwise} \end{cases}$$

$$N^{[G:\cap(\psi,\chi)]}(i,w) = \begin{cases} N(i,w) \cup \{||\psi \wedge \chi||^M_{i,w}\} & \text{if } i \in G \text{ and } \mathit{enabled}_w(G,\cap(\psi,\chi)) \\ & \text{and } M,w \models \mathbf{B}_i\psi \wedge \mathbf{B}_i\chi \\ N(i,w) & \text{otherwise} \end{cases}$$

$$N^{[G:\dashv(\psi,\chi)]}(i,w) = \begin{cases} N(i,w) \setminus \{||\chi||^M_{i,w}\} & \text{if } i \in G \text{ and } \mathit{enabled}_w(G,\vdash(\psi,\chi)) \\ & \text{and } M,w \models \mathbf{B}_i\psi \wedge \mathbf{K}_i(\psi \rightarrow \neg\chi) \\ N(i,w) & \text{otherwise} \end{cases}$$

$$N^{[G:\vdash(\psi,\chi)]}(i,w) = \begin{cases} N(i,w) \cup \{||\chi||^M_{i,w}\} & \text{if } i \in G \text{ and } \mathit{enabled}_w(G,\vdash(\psi,\chi)) \\ & \text{and } M,w \models \mathbf{B}_i\psi \wedge \mathbf{B}_i(\psi \rightarrow \chi) \\ N(i,w) & \text{otherwise} \end{cases}$$

Notice that after an inferential action α has been performed by an agent $j \in G$, all agents $i \in G$ see the same update in the neighborhood. Conversely, for any agent $h \notin G$ the neighborhood remains unchanged (i.e., $N^{[G:\alpha]}(h,w) = N(h,w)$). However, even for agents in G, the neighborhood remains unchanged if the required preconditions, on explicit beliefs, knowledge, and budget, do not hold (and hence the action is not executed). Notice also that we might devise variations of the logic by making different decisions about neighborhood update to implement, for instance, partial visibility within a group.

Since each agent in G has to contribute to cover the costs of execution by consuming part of its available budget, an update of the budget function is needed. We assume however that only inferential actions that add new beliefs have a cost. I.e., forming conjunction and performing belief revision are actions with no cost. As before, for an action α, we require $\mathit{enabled}_w(G,\alpha)$ to hold and assume that $j \in G$ executes α. Then, depending on α, we have:

$$B^{[G:\downarrow(\psi,\chi)]}(i,w) = \begin{cases} B(i,w) - \frac{C(j,\downarrow(\psi,\chi),w)}{|G|} & \text{if } i \in G \text{ and } \mathit{enabled}_w(G,\downarrow(\psi,\chi)) \text{ and} \\ & M,w \models \mathbf{B}_i\psi \wedge \mathbf{K}_i(\psi \rightarrow \chi) \\ B(i,w) & \text{otherwise} \end{cases}$$

$$B^{[G:\vdash(\psi,\chi)]}(i,w) = \begin{cases} B(i,w) - \frac{C(j,\vdash(\psi,\chi),w)}{|G|} & \text{if } i \in G \text{ and } \mathit{enabled}_w(G,\vdash(\psi,\chi)) \text{ and} \\ & M,w \models \mathbf{B}_i\psi \wedge \mathbf{B}_i(\psi \rightarrow \chi) \\ B(i,w) & \text{otherwise} \end{cases}$$

We write $\models_{\mathit{L\text{-}DINF}} \varphi$ to denote that $M,w \models \varphi$ holds for all worlds w of every model M.

We introduce below relevant consequences of our formalization. For lack of space we omit the proof, that can be developed analogously to what done in previous work [10].

Property As consequence of previous definitions, for any set of agents G and each $i \in G$, we have the following:

- $\models_{L\text{-}INF} (\mathbf{K}_i(\varphi \to \psi)) \wedge \mathbf{B}_i\, \varphi) \to [G : \downarrow(\varphi, \psi)]\, \mathbf{B}_i\, \psi$.

  Namely, if an agent has φ among beliefs and $\mathbf{K}_i(\varphi \to \psi)$ in its background knowledge, then as a consequence of the action $\downarrow(\varphi, \psi)$ the agent and any group G to which it belongs start believing ψ.

- $\models_{L\text{-}INF} (\mathbf{K}_i(\varphi \to \neg\psi)) \wedge \mathbf{B}_i\, \varphi) \to [G : \dashv(\varphi, \psi)]\, \neg\mathbf{B}_i\, \psi$.

  Namely, if an agent has φ among beliefs and $\mathbf{K}_i(\varphi \to \psi)$ in its background knowledge, then as a consequence of the action $\downarrow(\varphi, \psi)$ the agent and any group G to which it belongs stop believing ψ.

- $\models_{L\text{-}INF} (\mathbf{B}_i\varphi \wedge \mathbf{B}_i\psi) \to [G : \cap(\varphi, \psi)]\mathbf{B}_i(\varphi \wedge \psi)$.

  Namely, if an agent has φ and ψ as beliefs, then as a consequence of the action $\cap(\varphi, \psi)$ the agent and any group G to which it belongs starts believing $\varphi \wedge \psi$.

- $\models_{L\text{-}INF} (\mathbf{B}_i(\varphi \to \psi)) \wedge \mathbf{B}_i\, \varphi) \to [G : \vdash(\varphi, \psi)]\, \mathbf{B}_i, \psi$.

  Namely, if an agent has φ among its beliefs and $\mathbf{B}_i(\varphi \to \psi)$ in its working memory, then as a consequence of the action $\vdash(\varphi, \psi)$ the agent and any group G to which it belongs starts believing ψ.

2.4. Axiomatization

Below we introduce the axiomatization of our logic.

The *L-INF* and *L-DINF* axioms and inference rules are the following:

1. $(\mathbf{K}_i\, \varphi \wedge \mathbf{K}_i(\varphi \to \psi)) \to \mathbf{K}_i\, \psi$;
2. $\mathbf{K}_i\, \varphi \to \varphi$;
3. $\neg\mathbf{K}_i(\varphi \wedge \neg\varphi)$;
4. $\mathbf{K}_i\, \varphi \to \mathbf{K}_i\, \mathbf{K}_i\, \varphi$;
5. $\neg\mathbf{K}_i\, \varphi \to \mathbf{K}_i\, \neg\mathbf{K}_i\, \varphi$;
6. $\mathbf{B}_i\, \varphi \wedge \mathbf{K}_i\, (\varphi \leftrightarrow \psi) \to \mathbf{B}_i\, \psi$;
7. $\mathbf{B}_i\, \varphi \to \mathbf{K}_i\, \mathbf{B}_i\, \varphi$;
8. $\frac{\varphi}{\mathbf{K}_i\, \varphi}$;
9. $[G : \alpha]p \leftrightarrow p$;
10. $[G : \alpha]\neg\varphi \leftrightarrow \neg[G : \alpha]\varphi$;
11. $exec_G(\alpha) \to \mathbf{K}_i\, (exec_G(\alpha))$;
12. $[G : \alpha](\varphi \wedge \psi) \leftrightarrow [G : \alpha]\varphi \wedge [G : \alpha]\psi$;

13. $[G:\alpha]\mathbf{K}_i\,\varphi \leftrightarrow \mathbf{K}_i\,([G:\alpha]\varphi)$;
14. $[G:\downarrow(\varphi,\psi)]\mathbf{B}_i\,\chi \leftrightarrow \mathbf{B}_i\,([G:\downarrow(\varphi,\psi)]\chi) \vee \big((\mathbf{B}_i\,\varphi \wedge \mathbf{K}_i\,(\varphi \rightarrow \psi)) \wedge \mathbf{K}_i\,([G:\downarrow(\varphi,\psi)]\chi \leftrightarrow \psi)\big)$;
15. $[G:\cap(\varphi,\psi)]\mathbf{B}_i\,\chi \leftrightarrow \mathbf{B}_i\,([G:\cap(\varphi,\psi)]\chi) \vee \big((\mathbf{B}_i\,\varphi \wedge \mathbf{B}_i\,\psi) \wedge \mathbf{K}_i\,[G:\cap(\varphi,\psi)]\chi \leftrightarrow (\varphi \wedge \psi)\big)$;
16. $[G:\vdash(\varphi,\psi)]\mathbf{B}_i\,\chi \leftrightarrow \mathbf{B}_i\,([G:\vdash(\varphi,\psi)]\chi) \vee \big((\mathbf{B}_i\,\varphi \wedge \mathbf{B}_i\,(\varphi \rightarrow \psi)) \wedge \mathbf{B}_i\,([G:\vdash(\varphi,\psi)]\chi \leftrightarrow \psi)\big)$;
17. $intend_G(\phi_A) \leftrightarrow \forall i \in G\; intend_i(\phi_A)$;
18. $do_G(\phi_A) \rightarrow can\_do_G(\phi_A)$;
19. $do_i(\phi_A) \rightarrow can\_do_i(\phi_A)$;
20. $\frac{\psi \leftrightarrow \chi}{\varphi \leftrightarrow \varphi[\psi/\chi]}$.

We write *L-DINF*$\vdash \varphi$ to denote that φ is a theorem of *L-DINF*.

It is easy to verify that the above axiomatization is sound for the class of *L-INF* models, namely, all axioms are valid and inference rules preserve validity. In particular, soundness of axioms 14–16 immediately follows from the semantics of $[G:\alpha]\varphi$, for each inferential action α, as previously defined.

Notice that, by abuse of notation, we have axiomatized the special predicates concerning intention and action enabling. Axioms 17–19 concern in fact physical actions, stating that: what is intended by a group of agents is intended by them all; and, neither an agent nor a group of agents can do what they are not enabled to do. Such axioms are not enforced by the semantics, but are supposed to be enforced by a designer's/programmer's encoding of parts of an agent's behaviour. In fact, axiom 17 enforces agents in a group to be cooperative. Axioms 18 and 19 ensure that agents will attempt to perform actions only of their preconditions are satisfied, i.e., if they can do them. We do not handle such properties in the semantics as done, e.g., in dynamic logic, because we want agents' definition to be independent of the practical aspect, and vice versa we intend to introduce flexibility in the definition of such parts.

3. Problem Specification and Inference: an Example

In this section we propose an example of problem specification and inference in *L-DINF*. Consider a group of n agents, e.g., three, who are siblings or friends, who decide to act together in order to renovate some property, e.g., a cottage where to spend weekends. In order to save money and time they aim to contribute at practical work, to the extent of their capabilities. Prior to starting the activities, they agree upon sustaining costs in equal parts. They know that one of them is able to repair the roof, while the other two are both able to redecorate the walls and replace the carpet, but one of the two would clearly prefer the former task. Below we show how our logic is able to represent

the situation, and the proceedings of this work. For the sake of simplicity of illustration and of brevity, the example is in "skeletal" form.

Each agent will initially have the fact $\mathbf{K}_i(intend_G(renovate))$ (implicitly, the cottage) in its knowledge base. The physical actions that agents can perform are the following:

$$buy\text{-}material, \quad redecorate\text{-}walls, \quad repair\text{-}roof, \quad replace\text{-}carpet. \tag{1}$$

Assume that the knowledge base of each agent i contains the following rule, that specifies how to reach the intended goal in terms of actions to perform:

$$\begin{aligned}\mathbf{K}_i(intend_G(renovate) \quad \rightarrow \quad & intend_G(buy\text{-}material) \wedge intend_G(repair\text{-}roof) \wedge \\ & intend_G(replace\text{-}carpet) \wedge intend_G(redecorate\text{-}walls))\end{aligned}$$

By axiom 17 listed in previous section, every agent will also have the specialized rule

$$\begin{aligned}\mathbf{K}_i(intend_i(renovate) \quad \rightarrow \quad & intend_i(buy\text{-}material) \wedge intend_i(repair\text{-}roof) \wedge \\ & intend_i(replace\text{-}carpet) \wedge intend_i(redecorate\text{-}walls))\end{aligned}$$

Therefore, the following is entailed for each of the agents ($1 \leq i \leq 3$):

$$\begin{aligned}\mathbf{K}_i(intend_i(renovate) &\rightarrow intend_i(buy\text{-}material)) \\ \mathbf{K}_i(intend_i(renovate) &\rightarrow intend_i(repair\text{-}roof)) \\ \mathbf{K}_i(intend_i(renovate) &\rightarrow intend_i(replace\text{-}carpet)) \\ \mathbf{K}_i(intend_i(renovate) &\rightarrow intend_i(redecorate\text{-}walls))\end{aligned}$$

Assume now that the knowledge base of each agent i contains also the following general rules, stating that the group is available to perform each of the necessary actions.

$$\begin{aligned}\mathbf{K}_i(intend_G(buy\text{-}material) \wedge can\_do_G(buy\text{-}material) \wedge \\ pref\_do_G(i, buy\text{-}material) \rightarrow do_G(buy\text{-}material)) \\ \mathbf{K}_i(intend_G(repair\text{-}roof) \wedge can\_do_G(repair\text{-}roof) \wedge \\ pref\_do_G(i, repair\text{-}roof) \rightarrow do_G(repair\text{-}roof)) \\ \mathbf{K}_i(intend_G(replace\text{-}carpet) \wedge can\_do_G(replace\text{-}carpet) \wedge \\ pref\_do_G(i, replace\text{-}carpet) \rightarrow do_G(replace\text{-}carpet)) \\ \mathbf{K}_i(intend_G(redecorate\text{-}walls) \wedge can\_do_G(redecorate\text{-}walls) \wedge \\ pref\_do_G(i, redecorate\text{-}walls) \rightarrow do_G(redecorate\text{-}walls))\end{aligned}$$

As before, by axiom 17 such rules can be specialized to each single agent $1, 2, 3$.

$$\begin{aligned}\mathbf{K}_i(intend_i(buy\text{-}material) \wedge can\_do_i(buy\text{-}material) \wedge \\ pref\_do_G(i, buy\text{-}material) \rightarrow do_i(buy\text{-}material)) \\ \mathbf{K}_i(intend_i(repair\text{-}roof) \wedge can\_do_i(repair\text{-}roof) \wedge \\ pref\_do_G(i, repair\text{-}roof) \rightarrow do_i(repair\text{-}roof)) \\ \mathbf{K}_i(intend_i(replace\text{-}carpet) \wedge can\_do_i(replace\text{-}carpet) \wedge \\ pref\_do_G(i, replace\text{-}carpet) \rightarrow do_i(replace\text{-}carpet)) \\ \mathbf{K}_i(intend_i(redecorate\text{-}walls) \wedge can\_do_i(redecorate\text{-}walls) \wedge \\ pref\_do_G(i, redecorate\text{-}walls) \rightarrow do_i(redecorate\text{-}walls))\end{aligned}$$

As previously stated, whenever an agent derives $do_i(\phi_A)$ for any physical action ϕ_A, the action is supposed to have been performed via some kind of *semantic attachment* which links the agent to the external environment. However, $do_i(\phi_A)$ will be derived by means of a mental action based upon the available rules. Such mental action can have a cost, that can be payed either by the agent itself or by the group (according to the adopted policy of cost-sharing for this group). The reason to attribute the cost to the mental action is exactly to avoid that some agent tries to execute physical actions that it cannot support. According to the above rules, an agent can execute an action ϕ_A if it is allowed to performed that action ($can\_do_i(\phi_A)$) and if it is the one most willing to do it ($pref\_do_G(i, \phi_A)$). In our approach, such conclusion will be drawn on the basis of the assessment performed in external modules. Such modules will provide the decision according to some kind of reasoning process in some formalism, with respect to which our logic is completely agnostic, and they will add the corresponding facts to each agent's knowledge base.

In order to have our agents do the actions listed in (1) (note that one agent will have to perform two of them, as there are three agents and four actions), four sequences of mental actions will have to be executed, yielding, respectively, conclusions of the forms

$$\begin{array}{ll} do_G(\textit{buy-material}), & do_G(\textit{repair-roof}), \\ do_G(\textit{replace-carpet}), & do_G(\textit{redecorate-walls}). \end{array}$$

and causing their addition to agents' working memory. Such reasoning would consist in mental actions of kind $\cap$ to form conjunctions from single facts, and mental actions of kind $\downarrow$ to apply knowledge rule, i.e., given their preconditions, draw the conclusions. In particular, given the initial general intention by the group, it will be possible to derive the practical goal, in terms of the conjunction of actions to be performed by the group. From its own specialized rules and from the available facts about enabling and willingness, the execution of each action by some agent i will be hopefully derived. Note that, there can be the unlucky situation where no agent is enabled to perform some action, or that the one allowed is not willing, or that there is not enough budget. In this case, the goal fails.

Let α_1–α_4 be the last mental actions performed at the end of the mentioned four sequences of mental inferences (that lead to derive the $do_i(\phi_A)$, for some $i \leq 3$ and for ϕ_A among the actions in 1), respectively. Assume, moreover, that the costs of α_1–α_4 are the following (and, for simplicity, assume all other mental actions to have no cost):

$$C(i, \alpha_1, w) = 18, \quad C(i, \alpha_2, w) = 15, \quad C(i, \alpha_3, w) = 3, \quad C(i, \alpha_4, w) = 20.$$

and that $\alpha_j \in E(i, w), j \leq 3$ holds, for each world w, each agent i, and each action α_j.

Assume that in world w_1 the three agents have the following budgets to perform mental actions:

$$B(1, w_1) = 11,\ B(2, w_1) = 21,\ B(3, w_1) = 20$$

Assume, e.g., that all agents are enabled in w_1 to go and buy material. Suppose that agent 1 is the best wishing to go to buy, i.e., under the current model (which remains implicit)

$$w_1 \models \mathit{can\_do}_1(\mathit{buy\text{-}material}) \wedge \mathit{pref\_do}_G(1, \mathit{buy\text{-}material})\,.$$

However, (s)he alone cannot perform the action, because (s)he does not have enough budget. But, using the inferential action $[G : \alpha_1]$, with $G = \{1, 2, 3\}$, the other agents can devote part of their budgets to share the cost, so the group can perform α_1, because $\frac{C(1,\alpha_1,w_1)}{|G|} \leq \min_{h\in G} B(h, w_1)$.

Hence, $\mathbf{B}_i(\mathit{do}_G(\mathit{buy\text{-}material}))$ can be inferred by each agent i (in consequence, the *past event* $\mathbf{B}_i(\mathit{do}^P_G(\mathit{buy\text{-}material}))$ will also be asserted). Indeed, the inferential action is considered as performed by the whole group G, so each agent of G updates its neighborhood. After the execution of the action the budget of each agent is updated as well (cf., Section 2.2). The new budgets, given that we are assuming the policy to divide expenses into equal parts, are:

$$B(1, w_2) = 5,\ B(2, w_2) = 15,\ B(3, w_2) = 14$$

where we name w_2 the situation reached after executing the action.

Assume that only agent 3 is enabled in w_2 to repair the roof. Suppose that agent 3 is the best wishing to go to repair, i.e., under the current model (which remains implicit) $w_2 \models \mathit{can\_do}_3(\mathit{repair\text{-}roof}) \wedge \mathit{pref\_do}_G(3, \mathit{repair\text{-}roof})$. (S)he alone however cannot perform the action, because (s)he does not have enough budget. But, using the inferential action $[G : \alpha_2]$, with $G = \{1, 2, 3\}$, the other agents can devote part of their budgets to share the cost, so the group can perform α_2, because $\frac{C(3,\alpha_2,w_2)}{|G|} \leq \min_{h\in G} B(h, w_2)$. Hence, $\mathbf{B}_i(\mathit{do}_G(\mathit{repair\text{-}roof}))$ can be inferred by each agent i (in consequence, also $\mathbf{B}_i(\mathit{do}^P_G(\mathit{repair\text{-}roof}))$ will be asserted). Again, after the execution of the action the budget of each agent is updated. The new budgets, given that we are assuming the policy to divide expenses into equal parts, are:

$$B(1, w_3) = 0,\ B(2, w_3) = 10,\ B(3, w_3) = 9$$

where we name w_3 the situation reached after executing the action.

Assume that only agent 2 is enabled in w_3 to replace the carpet. (S)he can perform the action alone because (s)he has enough budget. So, (s)he can perform $[G : \alpha_3]$, with $G = \{1, 2, 3\}$ obtaining $\mathbf{B}_i(\mathit{do}_G(\mathit{replace\text{-}carpet}))$ (and, in consequence, $\mathbf{B}_i(\mathit{do}^P_G(\mathit{replace\text{-}carpet}))$). Indeed, the inferential action

is considered as performed by the whole group G so each agent of G updates its neighborhood. After the execution of the action only the budget of agent 2 is updated: $B(2, w_4) = 7$. Summing up budgets:

$$B(1, w_4) = 0,\ B(2, w_4) = 7,\ B(3, w_1) = 9$$

where we name w_4 the situation reached after executing the action.

There would be the last goal ($intend_G$(*redecorate-walls*)) but no agent has the necessary budget, so they cannot perform α_4 and the last goal cannot be achieved, and so the overall goal fails. Only some injection of new budget (maybe a loan from another group) might save the situation. Interaction among groups is a subject of future work.

It is relevant to comment about the role of past events. If the set of past events, which is a part of an agent's short-term memory, is made available to the external modules defining actions enabling and degree of willingness, past events might be used, for instance, to define constraints concerning actions execution. Referring to our example, it would be reasonable, e.g., to state that no repair can take place if the material has not been bought yet, and then, e.g., that repairing the roof should be performed as first thing.

4. Canonical Model and Strong Completeness

In this section we introduce the notion of canonical model of our logic, and we outline the proof of strong completeness w.r.t. the proposed class of models (by means of a standard canonical-model argument). As before, let *Agt* be a set of agents.

Definition 4.1. *The* canonical *L-INF model is a tuple*

$$M_c = \langle W_c, N_c, \mathcal{R}_c, E_c, B_c, C_c, A_c, P_c, V_c \rangle$$

where:

- W_c *is the set of all maximal consistent subsets of* $\mathcal{L}_{L\text{-}INF}$*;*
- *For* $w \in W_c$, $\varphi \in \mathcal{L}_{L\text{-}INF}$ *let* $A_\varphi(i, w) = \{v \in R_{c,i}(w) \mid \varphi \in v\}$. *Then, we put* $N_c(i, w) = \{A_\varphi(i, w) \mid \mathbf{B}_i\, \varphi \in w\}$.
- $\mathcal{R}_c = \{R_{c,i}\}_{i \in Agt}$ *is a collection of equivalence relations on* W_c *such that, for every* $i \in Agt$ *and* $w, v \in W_c$, $wR_{c,i}v$ *if and only if (for all* φ, $\mathbf{K}_i\, \varphi \in w$ *implies* $\varphi \in v$*)*
- $E_c : Agt \times W_c \longrightarrow 2^{\mathcal{L}_{\mathsf{ACT}}}$ *is such that, for each* $i \in Agt$ *and* $w, v \in W_c$, *if* $wR_{c,i}v$ *then* $E_c(i, w) = E_c(i, v)$*;*
- $B_c : Agt \times W_c \longrightarrow \mathbb{N}$ *is such that, for each* $i \in Agt$ *and* $w, v \in W_c$, *if* $wR_{c,i}v$ *then* $B_c(i, w) = B_c(i, v)$*;*
- $C_c : Agt \times \mathcal{L}_{\mathsf{ACT}} \times W_c \longrightarrow \mathbb{N}$ *is such that, for each* $i \in Agt$, $\alpha \in \mathcal{L}_{\mathsf{ACT}}$, *and* $w, v \in W_c$, *if* $wR_{c,i}v$ *then* $C_c(i, \alpha, w) = C_c(i, \alpha, v)$*;*
- $A_c : Agt \times W_c \longrightarrow 2^{Atm_A}$ *is such that, for each* $i \in Agt$ *and* $w, v \in W_c$, *if* $wR_{c,i}v$ *then* $A_c(i, w) = A_c(i, v)$*;*

- $P_c : Agt \times W_c \times Atm_A \longrightarrow Int$ *is such that, for each* $i \in Agt$ *and* $w, v \in W$*, if* $wR_{c,i}v$ *then* $P_c(i, w, \alpha) = P_c(i, v, \alpha)$*;*
- $V_c : W_c \longrightarrow 2^{Atm}$ *is such that* $V_c(w) = Atm \cap w$*.*

Note that, analogously to what done before, $R_{c,i}(w)$ denotes the set $\{v \in W_c \mid wR_{c,i}v\}$, for each $i \in Agt$. It is easy to verify that M_c is an *L-INF* model as defined in Definition 2.4, since, it satisfies conditions (**C1**), (**C2**), (**D1**), (**E1**), (**F1**), (**G1**), (**H1**). Hence, it models the axioms and the inference rules 1–16 and 20 introduced before (while, as mentioned in Section 2.4, axioms 17–19 are assumed to be enforced by an external specification of some aspects of agents' behaviour). Consequently, the following properties hold too. Let $w \in W_c$, then

- given $\varphi \in \mathcal{L}_{L\text{-}INF}$, it holds that $\mathbf{K}_i \varphi \in w$ if and only if $\forall v \in W_c$ such that $wR_{c,i}v$, we have $\varphi \in v$;
- for $\varphi \in \mathcal{L}_{L\text{-}INF}$, if $\mathbf{B}_i \varphi \in w$ and $wR_{c,i}v$ then $\mathbf{B}_i \varphi \in v$;

Thus, $R_{c,i}$-related worlds have the same knowledge and N_c-related worlds have the same beliefs, i.e. there can be $R_{c,i}$-related worlds with different beliefs.

By proceeding similarly to what done in [1] we obtain the proof of strong completeness. We list the main theorems but omit lemmas and proofs, that we have however developed analogously to what done in previous work [10].

THEOREM 4.2. *L-INF is strongly complete for the class of L-INF models.*

THEOREM 4.3. *L-DINF is strongly complete for the class of L-INF models.*

5. Future Extensions

We intend in future work to enhance our language by introducing the expression $\Diamond\varphi$, which has to be read "the agent can ensure φ by executing some (inferential and/or physical) action in her repertoire". Specifically, we intend to inductively define: $\Diamond^0\varphi = \varphi$, $\Diamond^{k+1} = \Diamond\Diamond^k\varphi$. The formula $\Diamond^k B\,\varphi$ represents the fact that the agent is capable of inferring φ in k steps. We might easily extend our semantics by stating

$$M, w \models \Diamond\varphi \leftrightarrow \exists\alpha \in E(w) \text{ s.t. } M^\alpha, w \models \varphi$$

(where we are denoting by $E(w)$ the executability function for the single agent under consideration).
A tentative axiomatization could be

- $exec(\alpha) \wedge [\alpha]\varphi \rightarrow \Diamond\varphi$;
- $p \rightarrow \Diamond p$;
- $\Diamond(\varphi \wedge \psi) \rightarrow \Diamond\varphi \wedge \Diamond\psi$;
- $\Diamond\varphi \rightarrow \Diamond\Diamond\varphi$;

- $\Diamond B\,\varphi \rightarrow B\,\Diamond\varphi$;
- $\Diamond K\,\varphi \rightarrow K\,\Diamond\varphi$.
- $([\alpha]\varphi) \rightarrow \Diamond^1\varphi$.
- $([\alpha_1]([\alpha_2]\varphi)) \rightarrow \Diamond^2\varphi$.
- $([\alpha_1]([\alpha_2]([\alpha_3]\varphi))) \rightarrow \Diamond^3\varphi$

...

So far in fact, we have been able to consider only a limited number of iterations of the $\Diamond$ operator, in a specific (though analogous) way for each case.

Yet, even in the bounded form such operator would allow us to better formalize many practical situations, including the one in the example discussed in Section 3. There, one could take into account the expected duration of each action. For instance, the rule expressing the goal of our group of agents could be reformulated as follows, where $\Diamond^v\phi_A$ means that we expect action ϕ_A to take (at most) v steps for its completion:

$$\mathbf{K}_i(\mathit{intend}_G(\mathit{renovate}) \rightarrow \mathit{intend}_G(\Diamond^1\ \mathit{buy\_material}) \wedge \mathit{intend}_G(\Diamond^5\ \mathit{repair\_roof}) \wedge \mathit{intend}_G(\Diamond^4\ \mathit{replace\_carpet}) \wedge \mathit{intend}_G(\Diamond^4\ \mathit{redecorate\_walls}))$$

Intelligent software agents are usually modelled and programmed (via available agent-oriented programming languages) in terms of the BDI (Belief, Desire, Intention) modal logic [24], that however is limited to the representation of the mental state of the agent itself, but is too weak to represent Theory of Mind (ToM), which is understood as the ability to attribute mental states not only to oneself but also to others. I.e., it is the intuitive theory, developed during childhood, by which people understand others' actions in terms of their beliefs, desires, emotions, and supposed intentions. Such ability is crucial to interpret and predict other persons' behavioural responses. Recent research [16] has claimed that epistemic logic could be a suitable formalism for representing essential aspects of ToM for an autonomous agent. In our logic, the capability of agents in a group to be aware of other agents' (of the group) beliefs and intentions is already an embryonic form of ToM.

However, in developmental psychology, one of the standard methods to test the capabilities of a human child's ToM is "false-belief tasks". It is a class of tests in which the child is told a story involving multiple characters, where one or more of the characters necessarily develop, under the circumstances, some false belief. The child should then answer questions indicating whether she has correctly modelled the mental states (beliefs) of the characters, identifying the false beliefs and their motivation.

A common false belief task is the "Sally-Anne" task in which the child is shown a story about two girls, Sally and Anne, who are in a room with a basket and a box. Sally puts the marble into the basket, leaves the room, and then Anne moves the marble to the box in her absence. The child is then asked: "where does Sally believe the marble to be?". To pass the test, the child must

answer "in the basket", since Sally did not see Anne moving the marble, and therefore Sally has the false belief that the marble is still in the basket.

In our logic it is easy to model the consequences of actions, i.e., if moving an object from a container to another one, the mental operations $\downarrow$ or $\vdash$ allow an agent to conclude that the marble is in the second container, and the mental operation $\dashv$ can remove the (no longer valid) belief that the marble is in the original container.

As we have seen before, what is inferred or removed from the working memory via a mental action is common knowledge of all agents of any group to which the agent which does the action belongs. So, the Sally-Ann task might be solved in our logic by reconfiguring the group. I.e., Sally, Ann and the observer child can be assumed to belong to the group called, e.g., "Room1". So, all of them observe the action of Sally putting the marble into the basket. However, when Sally leaves the room she can be assumed to leave the group "Room1". Thus, she cannot "observe" the action where Anne moves the marble to the box, and in consequence she still retains the belief that the marble is in the basket. Since all past beliefs are common knowledge in a group, the child (that we consider as an agent in the group) can answer the question correctly.

Therefore, what is to be extended in our logic is to model explicitly that there are actions that lead an agent to leave or join a group. For Sally, leaving the room leads to leave the group, and re-entering the room leads to re-joining. So, all new beliefs formed or removed by the group in the meanwhile are not known to her. Reasonably enough, to suitably cope with these aspects a concept of time and time intervals might be needed, that we have already treated in past work [9, 23] and might be suitably exploited in this context.

6. Conclusions

In this paper, we have reported about a line of work concerning how to exploit a logical formulation for providing the semantics of MAS, covering not only single agents, but also groups of cooperative agents. We aimed to consider to some extent practical aspects concerning actions' executability. So, we introduced beliefs about physical actions concerning whether they could, are, or have been executed. These beliefs can be potentially useful for explainability, but also to model complex group dynamics. We introduced costs of actions, and agents' preferences in performing actions. We introduced single agent's and group's intentions so that, as shown by means of an example, a group of agent can devise a joint plan to reach a goal step by step taking into account composing agents' capabilities and preferences, and the available resources. We tried to make our semantics modular, thus allowing engineers to encode some customizable aspects separately from the 'main' agent code. To model these aspects we have extended our previously-proposed epistemic logic *L-DINF*. We

have introduced dedicated syntax to represent actions' feasibility and preferences, aiming to introduce a connection among the 'abstract' agents and the external environment in which they will be situated, and we have shown that the new syntax facilitates (at least in principle) the explainability of an agent's internal logical processes, since a natural-language explanation can in principle be directly extracted from proofs.

We have proved some useful properties of the extended logic, among which strong completeness. We have provided a significant example, and we have outlined further extensions to the logic to better represent this and other examples.

The complexity of the extended logic needs by no means be higher than that of the original *L-DINF*, which is the same as that of other similar logics. So, we can safely claim that, in the proposed new logic, the satisfiability problem is PSPACE-complete in the multi-agent case for *L-INF*, and is decidable for *L-DINF* (though there are conjectures that it might be PSPACE-complete as well).

In future work, we mean to extend our logic so as to represent the number of steps needed to reach a goal, and relevant aspects of Theory of Mind, so as to define agents able to cope with "false-belief tasks", i.e., capable of attributing correct mental states to other agents also in presence of ambiguous situations. To this aim, we intend to integrate temporal aspects, i.e., in which instant or time interval an action has been or should be performed, and how this may affect resource usage, and agent's and group's functioning.

References

[1] Ph. Balbiani, D. Fernández Duque, and E. Lorini, *A logical theory of belief dynamics for resource-bounded agents*, in: 15th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), ACM, 2016, pp. 644–652.

[2] Ph. Balbiani, D. Fernández-Duque, and E. Lorini, *The dynamics of epistemic attitudes in resource-bounded agents*, Studia Logica **107** (2019), no. 3, 457–488.

[3] R. H. Bordini, L. Braubach, M. Dastani, A. El Fallah Seghrouchni, J. J. Gómez-Sanz, J. Leite, G. M. P. O'Hare, A. Pokahr, and A. Ricci, *A survey of programming languages and platforms for multi-agent systems*, Informatica (Slovenia) **30** (2006), no. 1, 33–44.

[4] R. Calegari, G. Ciatto, V. Mascardi, and A. Omicini, *Logic-based technologies for multi-agent systems: a systematic literature review*, Auton. Agents Multi Agent Syst. **35** (2021), no. 1, 1.

[5] S. Costantini and A. Tocchio, *A logic programming language for multi-agent systems*, in: Logics in Artificial Intelligence. JELIA 2002, Lecture Notes in Comput. Sci., vol 2424, Springer, 2002, pp. 1–13.

[6] S. Costantini and A. Tocchio, *The DALI logic programming agent-oriented language*, in: Logics in Artificial Intelligence. JELIA 2004, Lecture Notes in Comput. Sci., vol 3229. Springer, 2004, pp. 685–688.
[7] S. Costantini and G. De Gasperis, *Flexible goal-directed agents' behavior via DALI mass and ASP modules*, 2018 AAAI Spring Symposia, Stanford University, Palo Alto, California, USA, March 26-28, 2018, AAAI Press, 2018.
[8] S. Costantini, G. De Gasperis, and G. Nazzicone, *DALI for cognitive robotics: Principles and prototype implementation*, in: Practical aspects of declarative languages, Lecture Notes in Comput. Sci., vol. 10137, Springer, 2017, pp. 152–162.
[9] S. Costantini, A. Formisano, and V. Pitoni, *Timed memory in resource-bounded agents*, in: AI*IA 2018 - Advances in Artificial Intelligence. AI*IA 2018, Lecture Notes in Comput. Sci., vol. 11298, Springer, 2018, pp. 15–29.
[10] S. Costantini, A. Formisano, and V. Pitoni, *An epistemic logic for multi-agent systems with budget and costs*, in: Logics in Artificial Intelligence, JELIA 2021, Lecture Notes in Comput. Sci., vol. 12678, Springer, 2021, pp. 101–115.
[11] S. Costantini, A. Formisano, and V. Pitoni, *An epistemic logic for multi-agent systems with budget and costs*, in: EMAS 2021: 9th International Workshop on Engineering Multi-Agent Systems, Lecture Notes in Comput. Sci., Springer, 2021, to appear.
[12] S. Costantini and V. Pitoni, *Memory management in resource-bounded agents*, in: AI*IA 2019 – Advances in Artificial Intelligence. AI*IA 2019, Lecture Notes in Comput. Sci., vol 11946. Springer, 2019, pp. 46–58.
[13] S. Costantini and V. Pitoni, *Towards a logic of "inferable" for self-aware transparent logical agents*, in: Italian Workshop on Explainable Artificial Intelligence, CEUR Workshop Proceedings, vol. 2742, 2020, pp. 68–79.
[14] S. Costantini, A. Tocchio, and A. Verticchio, *Communication and trust in the DALI logic programming agent-oriented language*, Intelligenza Artificiale **2** (2005), no. 1, 39–46.
[15] G. De Gasperis, S. Costantini, and G. Nazzicone, *Dali multi agent systems framework*, doi: 10.5281/zenodo.11042, DALI GitHub Software Repository, July 2014, DALI: `http://github.com/AAAI-DISIM-UnivAQ/DALI`.
[16] L. Dissing and T. Bolander, *Implementing theory of mind on a robot using dynamic epistemic logic*, Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, ijcai.org, 2020, pp. 1615–1621.
[17] H. Van Ditmarsch, J. Y. Halpern, W. Van Der Hoek, and B. Kooi, *Handbook of epistemic logic*, College Publications, 2015, Editors.
[18] H. N. Duc, *Reasoning about rational, but not logically omniscient, agents*, J. Logic Comput. **7** (1997), no. 5, 633–648.
[19] A. Garro, M. Mühlhäuser, A. Tundis, M. Baldoni, C. Baroglio, F. Bergenti, and P. Torroni, *Intelligent agents: Multi-agent systems*, Encyclopedia of Bioinformatics and Computational Biology - Volume 1, Elsevier, 2019, pp. 315–320.
[20] A.I. Goldman et al., *Theory of mind*, The Oxford Handbook of Philosophy of Cognitive Science, vol. 1, Oxford University Press, 2012.
[21] A. Herzig, E. Lorini, and D. Pearce, *Social intelligence*, AI Soc. **34** (2019),

no. 4, 689.
[22] A. C. Kakas, P. Mancarella, F. Sadri, K. Stathis, and F. Toni, *Computational logic foundations of KGP agents*, J. Artificial Intelligence Res. **33** (2008), 285–348.
[23] V. Pitoni and S. Costantini, *A temporal module for logical frameworks*, Proceedings of ICLP 2019 (Tech. Comm.), EPTCS, vol. 306, 2019, pp. 340–346.
[24] A. S. Rao and M. Georgeff, *Modeling rational agents within a BDI architecture*, Proceedings of the Second Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'91), Morgan Kaufmann, 1991, pp. 473–484.
[25] R. W. Weyhrauch, *Prolegomena to a theory of mechanized formal reasoning*, Artif. Intell. **13** (1980), no. 1-2, 133–170.

Author's address:

Stefania Costantini
Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica
Università degli Studi dell'Aquila
Via Vetoio snc Loc. Coppito, I-67100 L'Aquila, Italy
E-mail: stefania.costantini@univaq.it

Received June 9, 2021
Revised September 20, 2021
Accepted September 22, 2021