Grounding power on actions and mental attitudes*

Emiliano Lorini[1,2], Andreas Herzig[2], Jan Broersen[3], and Nicolas Troquard[1,2]

[1] Inst. of Cognitive Sciences and Technologies (ISTC-CNR), Rome & Trento, Italy
[2] Institut de Recherche en Informatique de Toulouse (IRIT), France
[3] Universiteit Utrecht, Department of Information and Computing Sciences

Abstract. The main objective of this work is to develop a logical framework called $\mathcal{IAL}$ (*Intentional Agency Logic*) in which we can reason about mental states of agents, action occurrences, agentive and group powers. $\mathcal{IAL}$ will be exploited for a formal analysis of different forms of power such as an agent's *power of* achieving a certain result, an agent's *power to* do a certain action and an agent i's *power over* another agent j.

1 Introduction

Power is one of the most important concepts in social theory and multi-agent systems. In this work we aim at devising a general logical framework in which different forms of power can be specified and their intrinsic and relational properties investigated. A formal model of agentive power should clarify many subtle aspects of this individual and social phenomenon. It should characterize the most basic form of agentive power called *power to*. The *power to* of an agent i is relative to actions that i is capable to correctly perform at will (i.e. when having the intention to perform them). For example, for an agent to have the power to raise his arm, it has to be case that he will successfully raise his arm if intends to do this. This form of power has to be distinguished from an agent's *power of* achieving something. When looking at an agent's *power of* achieving a certain result, we discover that this is based on the interrelation between objective level and subjective level. In fact, i's *power of* achieving a certain result φ seems to involve not only i's objective opportunity of achieving φ but also i's awareness over such an opportunity. For example, for a thief to have the power of opening a safe, he must know the safe's combination. In the end, there are intrinsically social forms of *power of* which are commonly called *powers over*. These correspond to agentive powers to influence other agents to do or to refrain from doing certain actions. An agent i's *power over* another agent j consists in i's capacity to shape j's preferences in such a way that j will intend or will not intend to do a certain action. For example, for a politician to have the power over the electorate with regard to the action of voting him, he must have the power of inducing the electorate to vote him. It is evident from these few observations that a comprehensive formal model and ontology of power should allow to:

* Emiliano Lorini, Andreas Herzig and Nicolas Troquard are partially supported by the Interdisciplinary Laboratory on Interacting Knowledge Systems (ILIKS) and the french project ForTrust.

– specify the relationship between an agent's intention and the agent's action performance in order to assess whether the agent has the *power to* do a certain action at will or not;
– reason about beliefs of agents in order to study the discretional aspect of their *powers of*;
– clarify the true nature of an agent i's *power over* another agent j as i's capacity to affect j's intentions in such a way that j will do or will refrain from doing a certain action.

In the literature of applied logic both in philosophy and computer science, several authors have developed very sophisticated logics of social interaction [4, 14, 2, 19]. For instance, Alternating-time temporal logic (ATL) [2] is a logic in which one can express what coalitions can achieve by cooperating. ATL has coalition modalities $\langle\langle G \rangle\rangle$ where G is an arbitrary group of agents G. The ATL formula $\langle\langle G \rangle\rangle X\varphi$ means that coalition G has a collective strategy to ensure that, no matter what the other agents do, φ will be true in the next state. In STIT logic [4, 14] modal operators of the form $[i\ cstit :]$, called $Chellas$ STIT operators, and a modal operator of historical necessity of the form $\Box$, whose dual is $\Diamond$, are given. In STIT formulas $[i\ cstit : \varphi]$ and $\Diamond\,[i\ cstit : \varphi]$ respectively mean that i sees to it that φ and i can see to it that φ. There are extensions of such logics of social interaction in which knowledge modalities for agents and coalitions of agents are introduced [24, 7]. Moreover, there are extensions in which actions are promoted to first-class citizens in the formal language [22] and the properties of interaction between action and knowledge of agents can be expressed [1]. In our view all these approaches are still insufficient to formalize many relevant forms and properties of agentive and group power. What is still missing in the logical literature is an integration of the expressiveness of such logics of social interaction with the expressivenesses of a logic of mental attitudes (so-called BDI logic[4]) and dynamic logic [13] in which actions of agents are explicit.[5] In this work we will try to fill this gap by developing a logic which allows to reason about mental states of agents, action occurrences, agentive and group powers and to capture some interesting properties of *power to, power of* and *power over*.

The paper is organized as follows. In section 2 we will present the syntax and the semantics of a logic of powers and mental states called $\mathcal{IAL}$ (*Intentional Agency Logic*). In section 3 the axiomatization of $\mathcal{IAL}$ will be given and some of its properties will be studied. In the second part of the paper (section 5) we will exploit $\mathcal{IAL}$ to formalize and study the properties of different forms of agentive power.

2 A logic of powers and mental states: syntax and semantics

The logic $\mathcal{IAL}$ (*Intentional Agency Logic*) combines the expressiveness of a logic of actions and mental states with the expressiveness of a logic of social interaction. On the top of a logic which allows to specify what agents and groups of agents can bring about and to talk about occurrences of actions of single agents, we introduce modal operators

[4] See [25, 28] for a survey on BDI logics.
[5] For a similar attempt to introduce mental attitudes in a logic of strategic interaction, see [17].

for beliefs and goals of agents. We here consider intentional actions only. The syntactic primitives of the logic are the following:

- a nonempty finite set of agents $AGT = \{1, 2, ..., n\}$;
- a nonempty finite set of *atomic actions* $ACT = \{a, b, ...\}$;
- a set of atomic formulas $\Pi = \{p, q, ...\}$.

Given an arbitrary agent $i \in AGT$ we denote with $Act(i)$ the set of all possible couples $i : a$, that is, $Act(i) = \{i : a | a \in ACT\}$. Besides, we denote with Δ the set of all possible combinations of actions by the agents in AGT, that is, $\Delta = \prod_{i \in AGT} Act(i)$. Elements in Δ are tuples denoted by δ, δ', δ'',... For notational convenience, given a certain $\delta \in \Delta$, we denote with δ_i the element in δ corresponding to agent i. For example, if $AGT = \{1, 2, 3\}$, and $\delta = (1 : a, 2 : b, 3 : c)$, then $\delta_1 = 1 : a$. Moreover, we denote with $\delta_C := (\delta_i)_{i \in C}$ the tuple which consists of all δ_i for $i \in C$. For example, if $AGT = \{1, 2, 3\}$, $C = \{1, 2\}$ and $\delta = (1 : a, 2 : b, 3 : c)$, then $\delta_C = (1 : a, 2 : b)$. The language $\mathcal{L}_{\mathcal{IAL}}$ is given by the following BNF:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid [i : a]\, \varphi \mid Does_C\varphi \mid \Box\varphi \mid Bel_i\varphi \mid Goal_i\varphi$$

where $p \in \Pi$, $a \in ACT$, $i \in AGT$ and $C \subseteq AGT$. $Bel_i\varphi$ is read "agent *i* believes that φ" whereas $Goal_i\varphi$ is read "agent *i* has the chosen goal that φ" or simply "agent *i* has decided to pursue φ". For the sake of simplicity, we will often use the expression "agent *i* wants that φ" as a reading of $Goal_i\varphi$. An agent's chosen goals are supposed to be consistent and compatible with his beliefs. The standard reading of $[i : a]\, \varphi$ is "φ holds after every occurrence of action a done by agent i". Hence $[i : a] \bot$ expresses "agent *i* does not do action a". If C is a coalition of two or more agents $Does_C\varphi$ is read "agents in coalition C bring it about that φ by doing something together" or simply "coalition C brings it about that φ". For the individual case, $Does_i\varphi$ is read "agent i brings it about that φ by doing some action" or simply "agent i brings it about that φ". The operator $\Box$ is used to quantify over choices of agents. Thus, $\Box\varphi$ has to be read "φ is true in each world corresponding to a choice for every agent" or simply "φ is necessarily true". Several abbreviations are used in our logic. The classical Boolean connectives $\wedge$, $\rightarrow$, $\leftrightarrow$, $\top$ (tautology) and $\bot$ (contradiction) are defined from $\vee$ and $\neg$ in the usual manner. Moreover, $\langle i : a\rangle\, \varphi$ abbreviates $\neg\, [i : a]\, \neg\varphi$, $\Diamond\varphi$ abbreviates $\neg\Box\neg\varphi$. As we will show in section 3, under some assumptions of our logic the more natural readings of $\langle i : a\rangle\, \varphi$ is "agent *i* does a and φ is true after a's occurrence". In fact, we suppose that an action performed by an agent at a certain moment is responsible for producing all outcomes that the agent brings about at that moment and produces only those outcomes.[6] Finally, $\Diamond\varphi$ is read "there exists a world corresponding to a choice for every agent in which φ is true" or simply "φ can/may be true". The operators $\Diamond$ and $Does_C$ (viz. $Does_i$) can be exploited for expressing what a coalition C (viz. a single agent i) can bring about. $\Diamond Does_C\varphi$ has to be read "there exists a world corresponding to a choice for every agent in which coalition C brings it about that φ" or simply " coalition C can bring it about that φ".[7]

[6] Note that this is different from supposing linear time or action determinism.

[7] $\Box Does_C\varphi$ might be read "coalition C necessarily brings it about that φ".

2.1 Model definition

$\mathcal{IAL}$ models are tuples $M = (W, R, R_\Box, S, B, G, \pi)$ where:

- W is a set of possible worlds or states;
- $R_\Box$ is an equivalence relation on W;
- R is a collection of binary relations $R_{i:a}$ on W one for every couple $i : a$ where $i \in AGT$ and $a \in ACT$;
- S is a collection of serial relations S_C on W one for every coalition $C \subseteq AGT$;
- B and G are collections of binary relations B_i and G_i on W one for every agent $i \in AGT$. We suppose that every B_i is transitive, euclidean and serial, whilst every G_i is serial;
- $\pi : \Pi \longrightarrow 2^W$ is a valuation function.

We suppose that all $\mathcal{IAL}$ models satisfy the following additional semantic conditions. For any $a \in ACT$, $i \in AGT$, $w \in W$:

($S.1$) $S_\emptyset(w) \subseteq (S_{AGT} \circ R_\Box(w))$
($S.2$) if $w' \in R_\Box(w)$ then $S_\emptyset(w') \subseteq S_\emptyset(w)$
($S.3$) if $R_{i:a}(w) \neq \emptyset$ then $S_i(w) = R_{i:a}(w)$
($S.4$) if $w' \in S_{AGT}(w)$ and $w'' \in S_{AGT}(w)$ then $w' = w''$
($S.5$) if $R_{i:a}(w) \neq \emptyset$ then $\forall\, w'$ if $w' \in G_i(w)$ then $R_{i:a}(w') \neq \emptyset$
($S.6$) $\forall w' \in G_i(w), R_{i:a}(w') = \emptyset$ or $\forall w' \in G_i(w), R_{i:a}(w') \neq \emptyset$
($S.7$) $B_i(w) \cap G_i(w) \neq \emptyset$
($S.8$) if $w' \in B_i(w)$ then $G_i(w') = G_i(w)$
($S.9$) if $v \in B_i(w)$ and $v' \in R_\Box(w)$ then $\exists w'$ s.t. $w' \in R_\Box(v)$ and $w' \in B_i(v')$

For any $w \in W$ and $B, C \subseteq AGT$:

($S.10$) $S_{B \cup C}(w) \subseteq S_B(w)$

For any $w \in W$, $C \subseteq AGT$ and $\delta \in \Delta$:

($S.11$) if $w' \in R_\Box(w)$ and $\forall i \in C$, $R_{\delta_i}(w) \neq \emptyset$ and $R_{\delta_i}(w') \neq \emptyset$ then $S_C(w') \subseteq S_C(w)$

For any $w \in W$, $B, C \subseteq AGT$ such that $B \cap C = \emptyset$ and $\delta, \delta' \in \Delta$:

($S.12$) if $\exists w' \in R_\Box(w)$ such that $\forall i \in C$, $R_{\delta_i}(w') \neq \emptyset$ and $\exists w'' \in R_\Box(w)$ such that $\forall j \in B$, $R_{\delta'_j}(w'') \neq \emptyset$ then $\exists w''' \in R_\Box(w)$ such that $\forall i \in C$, $R_{\delta_i}(w''') \neq \emptyset$ and $\forall j \in B$, $R_{\delta'_j}(w''') \neq \emptyset$

For any $w \in W$, $i \in AGT$:

($S.13$) $\bigcup_{a \in ACT} R_{i:a}(w) \neq \emptyset$

According to Property $S.1$, the set of outcomes that the empty coalition brings about is a subset of the set of all outcomes that the biggest coalition AGT can bring about. According to $S.2$, the set of outcomes that the empty coalition can bring about is independent from what the agents in AGT do. According to property $S.3$, if action a done by i produces an outcome then all outcomes brought about by i are outcomes that i

bring about by doing a and all outcomes that i brings about by doing a are outcomes that i brings about. An interesting consequence of property $S.3$ is that, if an agent i does n actions in parallel (for any $n \geq 1$) then the n actions are responsible for producing the outcomes that i brings about and only produce those outcomes. More generally, for every world w the relation $S_i(w)$ (which corresponds to an abstract action performed by agent i) is assigned a subset of the finite set of action names ACT which represents all concrete actions that i does in parallel at world w. Property $S.4$ says that the biggest coalition AGT brings about exactly one outcome. Properties $S.5$ and $S.6$ characterize the relationship between actions and goals. $S.5$ says that if action a is not performed by i then for every i's goal-accessible world, a is not executed by i. $S.6$ says that if action a is performed by i then for every i's goal-accessible world, a is performed by i. $S.7$ is a condition of weak realism, according to which, the set of i's belief-accessible worlds and the set of i's goal-accessible worlds are never disjoint. Property $S.8$ says that worlds that are compatible with i's goals are compatible with i's goals from those worlds which are compatible with i's beliefs. $S.9$ is a semantic condition of confluence which describes the relationship between $R_\Box$ and every B_i. According to Property $S.10$, the set of outcomes brought about by the union of coalitions B and C is a subset of the set of outcomes brought about by coalition B. Since $R_\Box$ is an equivalence relation, Property $S.11$ can be rewritten as follows. For any $w \in W$, $C \subseteq AGT$ and $\delta \in \Delta$: if $w' \in R_\Box(w)$ and $\forall i \in C$, $R_{\delta_i}(w) \neq \emptyset$ and $R_{\delta_i}(w') \neq \emptyset$ then $S_C(w') = S_C(w)$. This means that, the set of outcomes that agents in C can bring about by doing a combination of actions $\delta_C := (\delta_i)_{i \in C}$ is independent from what the other agents in AGT/C do. Property $S.12$ says that, given two disjoint coalitions B and C, if agents in C can do together a combination of actions $\delta_C := (\delta_i)_{i \in C}$ and agents in B can do together a combination of actions $\delta'_B := (\delta'_i)_{i \in B}$, then agents in $B \cup C$ can do together the combination of actions (δ_C, δ'_B). Property $S.13$ says that for any world w and agent i there is at least one action done by i at w (i.e. agents are never passive).

2.2 Truth conditions

Given a model M, a world w and a formula φ, we write $M, w \models \varphi$ to mean that φ is true at world w in M, under the basic semantics. The rules defining the truth conditions of formulas of our logic are inductively defined as follows.

– $M, w \models p \Longleftrightarrow w \in V(p)$.
– $M, w \models \neg\varphi \Longleftrightarrow$ not $M, w \models \varphi$.
– $M, w \models \varphi \vee \psi \Longleftrightarrow M, w \models \varphi$ or $M, w \models \psi$.
– $M, w \models \Box\varphi \Longleftrightarrow \forall w'$ if $w' \in R_\Box(w)$ then $M, w' \models \varphi$.
– $M, w \models [i:a]\,\varphi \Longleftrightarrow \forall w'$ if $w' \in R_{i:a}(w)$ then $M, w' \models \varphi$.
– $M, w \models Bel_i\varphi \Longleftrightarrow \forall w'$ if $w' \in B_i(w)$ then $M, w' \models \varphi$.
– $M, w \models Goal_i\varphi \Longleftrightarrow \forall w'$ if $w' \in G_i(w)$ then $M, w' \models \varphi$.
– $M, w \models Does_C\varphi \Longleftrightarrow \forall w'$ if $w' \in S_C(w')$ then $M, w' \models \varphi$.

We write $\models_{\mathcal{IAL}} \phi$ if formula ϕ is *valid* in all $\mathcal{IAL}$ models, i.e. $M, w \models \phi$ for every $\mathcal{IAL}$ model *M* and world w in *M*. Finally, we say that a formula ϕ is *satisfiable* if there exists a $\mathcal{IAL}$ model *M* and world w in *M* such that $M, w \models \phi$.

3 Axiomatization

The series of axiom schemes of $\mathcal{IAL}$ are given in Fig. 1. **S5**$_\Box$ corresponds to the

(**ProTau**)	All tautologies of propositional calculus
(**S5**$_\Box$)	All S5-theorems for $\Box$
(**KD**$_{Stit}$)	All KD-theorems for every $Does_C$
(**KD45**$_{Bel}$)	All KD45-theorems for every Bel_i
(**KD**$_{Goal}$)	All KD-theorems for every $Goal_i$
(**K**$_{Act}$)	All K-theorems for every $[i:a]$
(**Alt**$_{Stit}$)	$\neg Does_{AGT}\neg\varphi \rightarrow Does_{AGT}\varphi$
(**Incl**$_{Stit}$)	$\Box Does_{AGT}\varphi \rightarrow Does_{\emptyset}\varphi$
(**4**$_{Does_\emptyset,\Box}$)	$Does_{\emptyset}\varphi \rightarrow \Box Does_{\emptyset}\varphi$
(**SP**)	$\bigwedge_{i\in C}\langle\delta_i\rangle\top \wedge Does_C\varphi \rightarrow \Box(\bigwedge_{i\in C}\langle\delta_i\rangle\top \rightarrow Does_C\varphi)$
(**Indep**)	$\Diamond(\bigwedge_{i\in C}\langle\delta_i\rangle\top) \wedge \Diamond(\bigwedge_{j\in B}\langle\delta'_j\rangle\top) \rightarrow \Diamond(\bigwedge_{i\in C}\langle\delta_i\rangle\top \wedge \bigwedge_{j\in B}\langle\delta'_j\rangle\top)$ if $B\cap C=\emptyset$
(**Active**)	$\bigvee_{a\in ACT}\langle i:a\rangle\top$
(**Mon**)	$Does_B\varphi \rightarrow Does_{B\cup C}\varphi$
(**StitAct**)	$\langle i:a\rangle\top \rightarrow ([i:a]\varphi \leftrightarrow Does_i\varphi)$
(**IntAct1**)	$\langle i:a\rangle\top \rightarrow Goal_i\langle i:a\rangle\top$
(**IntAct2**)	$Goal_i\langle i:a\rangle\top \vee Goal_i[i:a]\bot$
(**D**$_{Bel,Goal}$)	$Goal_i\varphi \rightarrow \neg Bel_i\neg\varphi$
(**PosIntr**)	$Goal_i\varphi \rightarrow Bel_iGoal_i\varphi$
(**NegIntr**)	$\neg Goal_i\varphi \rightarrow Bel_i\neg Goal_i\varphi$
(**Confl**$_{Bel,\Box}$)	$\Diamond Bel_i\varphi \rightarrow Bel_i\Diamond\varphi$

Fig. 1. Axiomatization of $\mathcal{IAL}$

fact that $R_\Box$ is an equivalence relation; **KD**$_{Stit}$ corresponds to the seriality of every S_C; **KD45**$_{Bel}$ to the seriality, euclideanity and transitivity of every B_i; **KD**$_{Goal}$ to the seriality of every G_i. Moreover the following correspondence relations exist between the previous axioms and the semantic properties given in the previous section: Axiom **Alt**$_{Stit}$ corresponds to property $S.4$, **Incl**$_{Stit}$ to $S.1$, **4**$_{Does_\emptyset,\Box}$ to $S.2$, **SP** to $S.11$, **Indep** to $S.12$, **Active** to $S.13$, **Mon** to $S.10$, **StitAct** to $S.3$, **IntAct1** to $S.5$, **IntAct2** to $S.6$, **D**$_{Bel,Goal}$ to $S.7$, **PosIntr** and **NegIntr** to $S.8$ and **Confl**$_{Bel,\Box}$ to $S.9$.

Axioms **K**$_{Act}$, **S5**$_\Box$, **KD45**$_{Bel}$, **KD**$_{Goal}$ correspond to standard axiomatizations for the operators $[i:a]$ and $\Box$, the belief and goal operators. Axiom **D**$_{Bel,Goal}$ is a weak realism axiom which relates an agent's beliefs with his goals, whereas **PosIntr** and **NegIntr** are principles of positive and negative introspection for goals [11]. According to Axiom **Alt**$_{Stit}$ the biggest coalition AGT always produces deterministic effects, whilst according to Axioms **Incl**$_{Stit}$ and **4**$_{Does_\emptyset,\Box}$, if the biggest coalition AGT necessarily brings it about that φ then the empty coalition brings it about that φ and, if the empty coalition brings it about that φ then it necessarily brings it about that φ. We suppose that modal operators $Does_C$ "see one step forward". Thus we simply adopt a KD logic for every $Does_C$ (Axiom **KD**$_{Stit}$). Axiom **Mon** corresponds to a monotonicity property: if coalition B ensures φ then φ is ensured by all coalitions of which B is a subset.[8] According to Axiom **SP**, given a combination δ of actions of agents in AGT if every agent i in C does his part by executing the corresponding δ_i in δ and the coalition C brings it about that φ then, necessarily, if every agent i in C does his part by executing the corresponding δ_i in δ, coalition C brings it about that φ. Axiom **SP** characterizes a strong notion of power. Indeed, given Axiom **SP**, saying "coalition C can ensure φ by doing a certain combination of actions" is equivalent to say that "if coalition C does a certain combination of actions then it will ensure φ, no matter what the other agents in AGT do".[9] Axiom **Indep** says that if B and C are two disjoint coalitions, agents in C can do together a certain combination of actions δ_C and agents in B can do together a certain combination of actions δ'_B then agents in $B \cup C$ can do together a combination of actions (δ_C, δ'_B). This axiom is the "actional" counterpart of the axiom of *independence of agents* (called AIA_k) given in [4].[10] Axiom **Active** says that an agent always performs at least one action. According to Axiom **IntAct1**, an agent does action a only if he intends to do a. Thus, in our formal model the actions performed by an agent are intentional actions. According to Axiom **IntAct2**, at each moment an agent either decides (intends) to do an action or decides (intends) not to do it. Axiom $\mathbf{Confl}_{Bel,\Box}$ says that if there exists a world corresponding to a choice for every agent in which i believes that φ is true then i believes that there exists a world corresponding to a choice for every agent in which φ is true. Imagine there are two agents called Bill and Bob, and there exists a world corresponding to a choice of Bill and a choice of Bob in which Bob believes that he will meet Bill (i.e. $\Diamond Bel_{Bob} Does_{\{Bob,Bill\}} BobMeetsBill$). This is the world where both Bill and Bob decide to go to the same place Y and Bob believes that Bill has decided to go to Y. It seems reasonable to say that at the actual world - suppose this is the world where Bob decides to go to Y and Bill decides not to go-, Bob believes that there exists a world corresponding to Bill's choice and Bob's choice to go to Y in which Bill and Bob will meet (i.e. $Bel_{Bob} \Diamond Does_{\{Bob,Bill\}} BobMeetsBill$). In section 5.1 we will clarify why, on the other hand, property $Bel_i \Diamond \varphi \rightarrow \Diamond Bel_i \varphi$ cannot be accepted. Axiom **StitAct** is an interaction axiom between action occurrences and agentive causation. According to this axiom, an action performed by an agent at a certain moment is responsible for producing all outcomes that the agent brings about at that moment and produces only those outcomes. As noted in section 2, due to the semantic property $S.3$ corresponding to **StitAct** (i.e. if $R_{i:a}(w) \neq \emptyset$ then $S_i(w) = R_{i:a}(w)$), formula $\langle i : a \rangle \top$ has to be read "agent i does a", and $\langle i : a \rangle \top \wedge [i : a] \varphi$ has to be read "agent i brings it about that φ by doing a". This means that, in $\mathcal{IAL}$, $\langle i : a \rangle \top$ has not the standard dynamic logic reading "it is possible that i does a". In fact, according to property $S.3$, if action a done by i produces an outcome then all outcomes brought

[8] Note that **Mon** is equivalent to $Does_B \varphi \wedge Does_C \psi \rightarrow Does_{B \cup C}(\varphi \wedge \psi)$.

[9] At the single-agent level, this axiom corresponds to Weber's concept of power [27] as the capacity of an individual to resist to all interferences of other individuals, that is, "...the probability that one actor within a social relationship will be in a position to carry out his own will despite resistance..." (p. 152).

[10] Xu's Axiom AIA_k is a family of axiom schemes for *independence of agents* parameterized by the integer k of the form: $\Diamond [i_0 \; cstit: \; \varphi_0] \wedge ... \wedge \Diamond [i_k \; cstit: \; \varphi_k] \rightarrow \Diamond([i_0 \; cstit: \; \varphi_0] \wedge ... \wedge [i_k \; cstit: \; \varphi_k])$ (for $1 \leq k < k$).

about by i are outcomes that i bring about by doing a and all outcomes that i brings about by doing a are outcomes that i brings about. It follows that the more natural readings of formulas $\Diamond \langle i:a \rangle \top$ and $\Diamond(\langle i:a \rangle \top \wedge [i:a]\varphi)$ are respectively "agent i can do a" and "agent i can bring it about that φ by doing a". Furthermore, we have to note that, due to the fact that $Does_{AGT}$ is deterministic, it is reasonable to conceive state w' such that $w' = S_{AGT}(w)$ as the unique temporal successor of w and to read $Does_{AGT}\varphi$ "φ will be true in the next state". Thus, $Does_{AGT}$ can be interpreted as a standard operator X (*next*) of temporal logic.

Definition 1. $X\varphi =_{def} Does_{AGT}\varphi$

We call $\mathcal{IAL}$ the logic axiomatized by the twenty principles given in Fig. 1 and we write $\vdash_{\mathcal{IAL}} \phi$ if formula ϕ is a theorem of $\mathcal{IAL}$. Since the set of agents AGT and the set of atomic actions ACT is supposed to be finite, we can prove that $\mathcal{IAL}$ is *sound* and *complete* with respect to the class of $\mathcal{IAL}$ models.

Theorem 1. *$\mathcal{IAL}$ is determined by the class of models of $\mathcal{IAL}$.*

Proof. It is a routine to prove soundness, whereas completeness is obtained by Sahlqvist's completeness theorem [5]. □

4 Some properties of $\mathcal{IAL}$

The following theorems highlight some interesting properties of $\mathcal{IAL}$.

Theorem 2. *For any $i \in AGT$, $a \in ACT$ and $B, C \subseteq AGT$ such that $B \cap C = \emptyset$*

1. $\vdash_{\mathcal{IAL}} \Diamond Does_B\varphi \wedge \Diamond Does_C\psi \rightarrow \Diamond Does_{B\cup C}(\varphi \wedge \psi)$
2. $\vdash_{\mathcal{IAL}} X\varphi \leftrightarrow \neg X \neg\varphi$
3. $\vdash_{\mathcal{IAL}} \langle i:a \rangle \top \wedge [i:a]\varphi \rightarrow Does_i\varphi$
4. $\vdash_{\mathcal{IAL}} \Diamond(\langle i:a \rangle \top \wedge [i:a]\varphi) \rightarrow \Diamond Does_i\varphi$
5. $\vdash_{\mathcal{IAL}} \Diamond Does_B\varphi \wedge \Diamond Does_C\neg\varphi \rightarrow \bot$
6. $\vdash_{\mathcal{IAL}} \langle i:a \rangle \top \rightarrow Bel_i Goal_i \langle i:a \rangle \top$
7. $\vdash_{\mathcal{IAL}} \neg Goal_i \langle i:a \rangle \top \rightarrow Bel_i [i:a]\bot$
8. $\vdash_{\mathcal{IAL}} Goal_i\varphi \wedge Bel_i(\varphi \rightarrow \langle i:a \rangle \top) \rightarrow Goal_i \langle i:a \rangle \top$

Proof. Here we only prove Theorem 2.1 and Theorem 2.8. Let us start with Theorem 2.1. $\Diamond Does_B\varphi$ and $\Diamond Does_C\psi$ together imply $\exists \delta, \delta' \in \Delta$ such that $\Diamond(\bigwedge_{i\in B} \langle \delta_i \rangle \top \wedge Does_B\varphi)$ and $\Diamond(\bigwedge_{j\in C} \langle \delta'_j \rangle \top \wedge Does_C\psi)$ (by Axiom **Active**). From this, it follows that $\Diamond((\bigwedge_{i\in B, j\in C} \langle \delta_i \rangle \top \wedge \langle \delta'_j \rangle \top) \wedge Does_B\varphi \wedge Does_C\psi)$ (by Axiom **Indep**, Axiom **SP** and the fact that B and C are disjoint). We can conclude that $\Diamond Does_{B\cup C}(\varphi \wedge \psi)$ (by Axiom **Mon**).
In order to prove Theorem 2.8 it is sufficient to note that $Goal_i\varphi \wedge Bel_i(\varphi \rightarrow \langle i:a \rangle \top)$ implies $\neg Goal_i \neg \langle i:a \rangle \top$ (by Axiom $\mathbf{D}_{Bel,Goal}$) which in turn implies $Goal_i \langle i:a \rangle \top$ (by Axiom **IntAct2**). □

Theorem 2.1 says that two disjoint coalitions can combine their efforts to ensure a conjunction of outcomes. This corresponds to the *superadditivity* axiom of Coalition Logic [19] of the form $[B]\,\varphi \wedge [C]\,\psi \rightarrow [B \cup C]\,(\varphi \wedge \psi)$ (when $B \cap C = \emptyset$). Theorem 2.2 shows the tight correspondence between our definition of $next$ and the standard operator $next$ of linear temporal logic. According to Theorems 2.3 and 2.4 if an agent brings it about that φ by doing a then he brings it about that φ and, if an agent can bring it about that φ by doing a then he can bring it about that φ. Theorem 2.5, which is a direct consequence of Theorem 2.1, says that two disjoint coalitions can never bring about conflicting effects. Theorems 2.6 and 2.7 are about the relation between intentions, beliefs and action occurrences: if an agent does action a then he believes that he intends to do a; if an agent does not intend to do action a now then he believes that he will not perform a. Theorem 2.8 expresses a sort of generative principle for intentions according to which, if i wants φ to be true and believes that φ will be true only if he does a then i comes to intend to do a. It has to be noted that $Does_i\varphi \rightarrow Bel_i Does_i\varphi$ and $\neg Does_i\varphi \rightarrow Bel_i \neg Does_i\varphi$ are not valid here, that is, an agent is not necessarily aware of what he will bring about.

4.1 Related works

Main differences between STIT and $\mathcal{IAL}$ Some substantial differences exist between $\mathcal{IAL}$ and STIT logic [4, 14]. Formulas in STIT logic can be built by means of the boolean connectives together with the modal operator $\Box$ of historic necessity, whose dual is $\Diamond$, and the so-called $Chellas$ STIT operator $[i\ cstit :]$. The modal construction $\Box\varphi$ is read "φ is true in all possible histories", whereas $[i\ cstit : \varphi]$ is read "agent i sees to it that φ". Thus, $\Diamond\,[i\ cstit : \varphi]$ and $\Box\,[i\ cstit : \varphi]$ have to be read "agent i can see to it that φ" and "agent i necessarily sees to it that φ". Space restrictions prevent presenting STIT semantics, the interpretation of operators $\Box$ and $[i\ cstit :]$ and their semantic relationships. Let us only remark that in STIT theory formulas of type $[i\ cstit : \varphi]$ are interpreted according to equivalence relations. Thus, $[i\ cstit : \varphi] \rightarrow \varphi$ is valid in STIT. This means that in STIT theory actions are supposed to be instantaneous. In $\mathcal{IAL}$ we suppose that an agent (viz. a coalition of agents) brings about something as an effect of his actions (viz. joint actions) and that actions (viz. joint actions) are not instantaneous. For these reasons, for every $C \subseteq AGT$ S_C is simply a serial relation and for every $C \subseteq AGT$ $Does_C\varphi \wedge \neg\varphi$ is satisfiable.[11] Moreover, it has to be noted that under STIT semantics the following formulas are valid: $[i\ cstit : [j\ cstit : \varphi]] \leftrightarrow \Box\,[j\ cstit : \varphi]$; $[i\ cstit : [j\ cstit : \varphi]] \leftrightarrow [i\ cstit : \Box\,[j\ cstit : \varphi]]$.[12] This means that in STIT logic an agent can never really induce another agent to ensure some state of affairs φ. In fact, in STIT an agent always acts independently from what other agents do and cannot be induced to bring about something that he would not bring about without being in-

[11] Note that in STIT theory $\Box\varphi \rightarrow [i\ cstit : \varphi]$ is also valid. This is not the case in $\mathcal{IAL}$ where for every $C \subseteq AGT$ $\Box\varphi \wedge \neg Does_C\varphi$ is satisfiable.

[12] $[i\ cstit : [j\ cstit : \varphi]] \leftrightarrow \Box\varphi$ is also valid and if we refine STIT logic by supposing that time is discrete (as done in [6]) then even $[i\ cstit : X\,[j\ cstit : \varphi]] \leftrightarrow [i\ cstit : X\Box\varphi]$ becomes valid.

duced.[13] This is a serious limitation of this logic since it prevents expressing crucial aspects of sociality such as indirect power (see section 5.4). The nice aspect of $\mathcal{IAL}$ is that it does not incur these limitations. For instance, due to the temporal properties of the modal operators $Does_i$, in $\mathcal{IAL}$ the formulas $Does_i Does_j \varphi \wedge \neg \Box Does_j \varphi$ and $Does_i Does_j \varphi \wedge \neg Does_i \Box Does_j \varphi$ are satisfiable. Thus, $\mathcal{IAL}$ allows to express the fact that an agent j is induced by i to bring about φ while φ is something that j would not bring about without being induced by i.

Relationship between Coalition Logic and $\mathcal{IAL}$ Pauly's Coalition Logic (CL) [19] is one of the most well-known logics for multiagent systems. CL has been introduced to reason about what single agents and groups of agents are able to achieve. CL has coalition modalities of the form $[C]$ where C is an arbitrary coalition of agents $C \subseteq AGT$ (where AGT is the set of all agents). The CL formula $[C]\,\varphi$ is read "the coalition C can bring about (can enforce an outcome state satisfying) φ". In an extended version of this paper [16] it is proved that $\mathcal{IAL}$ subsumes Coalition Logic. More precisely the following translation $tr(.)$ from formulas of CL to formulas of $\mathcal{IAL}$ is given:

- $tr(p) = p$
- $tr(\neg\varphi) = \neg tr(\varphi)$
- $tr(\varphi \vee \psi) = tr(\varphi) \vee tr(\psi)$
- $tr([C]\,\varphi) = \Diamond Does_C tr(\varphi)$

and the following three theorems are proved:

- If φ is a theorem of CL then $tr(\varphi)$ is a theorem of $\mathcal{IAL}$.
- If φ is CL-satisfiable then $tr(\varphi)$ is satisfiable in $\mathcal{IAL}$.
- φ is CL-satisfiable if and only if $tr(\varphi)$ is satisfiable in $\mathcal{IAL}$.

A general observation Modal operators of logics of agency typically have three components: historic necessity, agent's choice, and time. In Pauly's Coalition Logic and in ATL these three components are *fused* and make up a single non-normal modal operator. We have seen that in STIT logic, these three ingredients are separated, and each has its own modal operator. In $\mathcal{IAL}$ we explore the middle ground: we fuse the choice and the temporal *next* operator.

5 Varieties of Power

5.1 Power of

The aim of this section is to provide a formal characterization of the concept of *power of* by exploiting the expressiveness of $\mathcal{IAL}$. We will start with a general definition of *power of* and we will progressively refine it. As argued in [8, 3], for an agent i to have

[13] In [4, 14] it is argued that the deliberative STIT construction $[i\ dstit : \varphi] =_{def} [i\ cstit : \varphi] \wedge \neg\Box\varphi$ provides a better approximation of agentive causation. It can be proved that $[i\ dstit : [j\ dstit : \varphi]] \leftrightarrow \bot$ is valid.

the power of achieving φ: i must have the objective opportunity to achieve φ and, he must be aware of this.[14] In fact, without i's discretion over his objective opportunity, i would not be capable of exploiting it in order to ensure φ. A first rough pre-formal definition of i's *power of* achieving φ is given by the two conditions:

1. i can bring it about that φ (*objective opportunity*);
2. i believes that he can bring it about that φ (*discretion over the opportunity*).

In $\mathcal{IAL}$ the former condition is expressed by the formula $\Diamond Does_i\varphi$, while the latter condition is expressed by the formula $Bel_i\Diamond Does_i\varphi$. Here we denote with $K_i\varphi$ i's correct belief that φ holds.

Definition 2. *For any* $i \in AGT$
$K_i\varphi =_{def} Bel_i\varphi \wedge \varphi$

Then, i's *power of* achieving φ can be expressed by the formula $K_i\Diamond Does_i\varphi$. If we look carefully at the semantics of $K_i\Diamond Does_i\varphi$, we can easily discover that it is insufficient to express a notion of genuine power. An evident problem with such a formal definition is the absence of a condition which guarantees that φ is not something which would happen in any case (independently from i's intervention). A more precise definition of objective opportunity would require a negative condition of the form $\neg\Box X\varphi$. In fact, it is counterintuitive to say that i has the opportunity of achieving φ when $\Box X\varphi$ holds, that is, when φ is going to be true whatever i does. For example, the fact "2+2=4" is something which is going to be true whatever i does (i.e. $\Box X$ "2+2=4"). For this reason, it is quite odd to say that i has the opportunity of ensuring that "2+2=4". This observation leads to the following refined formal characterization of objective opportunity: $\Diamond Does_i \wedge \neg\Box X\varphi$. From this, one might try to formalize the concept i's *power of* achieving φ by the formula $K_i(\Diamond Does_i\varphi \wedge \neg\Box X\varphi)$. But again this is not sufficient to formalize a genuine concept of power. In fact, $K_i(\Diamond Does_i\varphi \wedge \neg\Box X\varphi)$ simply says "i correctly believes that there exists some action whose execution can ensure φ and that φ is not something that is going to be true whatever i does". It does not say "there exists some action such that i may correctly believe that he will ensure φ by doing that action".[15] To see why $K_i(\Diamond Does_i\varphi \wedge \neg\Box X\varphi)$ is insufficient to capture the concept of power consider the scenario in Fig. 2. Agent i is at w_1 and is in front of two doors A and B. Behind door A there is a treasure, behind door B there is nothing. Besides, i believes that behind one of the two doors there is a treasure whereas behind the other there is nothing, but he is not sure whether the treasure is behind door A or B. The agent can either open door A or open door B. In world w_1 and in each world which is compatible with i's beliefs at w_1 (worlds w_7 and w_9) it is the case that he can get the treasure and that getting the treasure is not something that is going to necessarily happen. From this,

[14] A similar argument is given in [26] where the notion of *practical possibility* is distinguished from the notion of *power* (formalized by the operator CAN).

[15] The necessity to distinguish *de dicto* sentences of the form "i knows that there exists some action by doing which he can ensure φ" from *de re* sentences of the form "there exists some action such that i may correctly believe that he will ensure φ by doing it" (or "there exists some action such that i correctly believes that he may ensure φ by doing that action") has also been stressed in [1, 15, 7, 21].

we conclude that at w_1 i correctly believes that he can get the treasure and that getting the treasure is not something that is going to necessarily happen: $K_i(\Diamond Does_i t \wedge \neg\Box X t)$ holds at w_1. Unfortunately, there is no action that i may correctly believe that it will ensure φ. So, it is reasonable to say that in the example i does not have the power of getting the treasure. At w_1 i cannot correctly believe that he will get the treasure by opening door A nor correctly believe that he will get the treasure by opening door B: $\neg\Diamond K_i(\langle i:a\rangle \top \wedge [i:a]\, t)$ and $\neg\Diamond K_i(\langle i:b\rangle \top \wedge [i:b]\, t)$ hold at w_1. More generally, at w_1 i cannot correctly believe that he will get the treasure: $\neg\Diamond K_i Does_i t$ holds at at w_1. From the previous example, we have to conclude that an agent i does not have the

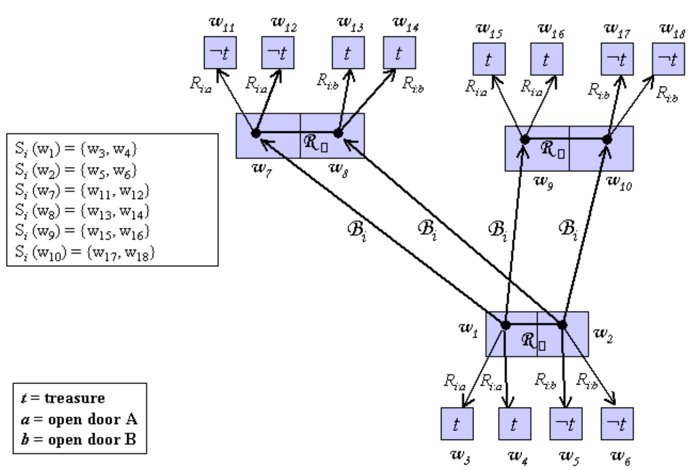

Fig. 2. Scenario

power of achieving φ by doing action a unless:

1* i may correctly believe that he will ensure φ by doing a and i correctly believes that φ is not something that is going to be true whatever he does.

More generally, an agent i does not have the *power of* achieving φ unless:

2* i may correctly believe that he will ensure φ and i correctly believes that φ is not something that is going to be true whatever he does.

1* and 2* correspond to a notion of *power of* which can be formalized in $\mathcal{IAL}$.

Definition 3. *For any* $i \in AGT$, $a \in ACT$

A. $Opp(i,a,\varphi) =_{def} \Diamond([i:a]\,\varphi \wedge \langle i:a\rangle \top) \wedge \neg\Box X\varphi$

B. $Power(i,a,\varphi) =_{def} \Diamond K_i([i:a]\,\varphi \wedge \langle i:a\rangle\,\top) \wedge K_i\neg\Box X\varphi$
C. $Power(i,\varphi) =_{def} \Diamond K_i Does_i\varphi \wedge K_i\neg\Box X\varphi$

Definitions 3.A, 3.B and 3.C respectively characterize i's opportunity of achieving φ by doing action a (i.e. i can ensure φ by doing a and φ is not something that is going to be true whatever i does), i's power of achieving φ by doing action a (definition 1^*) and i's power of achieving φ (definition 2^*).[16] It is straightforward to prove that $Power(i,a,\varphi)$ implies $Power(i,\varphi)$. Moreover, by Axiom $\mathbf{Confl}_{Bel,\Box}$, we can show that both $Power(i,a,\varphi)$ and $Power(i,\varphi)$ imply $K_i(\Diamond Does_i\varphi \wedge \neg\Box X\varphi)$[17] which, as discussed above, characterizes a situation of uncertainty in which i cannot determine what action must be taken to ensure φ.
The following theorems highlight some properties of the previous notions of power.

Theorem 3. *For any $i \in AGT$ and $a \in ACT$*

1. $\vdash_{\mathcal{IAL}} Power(i,a,\varphi) \leftrightarrow K_i Power(i,a,\varphi)$
2. $\vdash_{\mathcal{IAL}} Power(i,a,\varphi) \rightarrow K_i\Box\,[i:a]\,\varphi$

Proof. We only prove $\rightarrow$ direction of Theorem 3.1. The other direction is trivially satisfied by definition of $K_i\varphi$. By definition of $Power(i,a,\varphi)$ and $K_i\varphi$, we have that $Power(i,a,\varphi)$ implies $\Diamond K_i([i:a]\,\varphi \wedge \langle i:a\rangle\,\top) \wedge K_i\neg\Box X\varphi$. Moreover $\Diamond K_i\varphi \rightarrow K_i\Diamond\varphi$ and $K_i\varphi \rightarrow K_iK_i\varphi$ are theorems of $\mathcal{IAL}$ (by definition of $K_i\varphi$, Axiom **4** for Bel_i and Axiom $\mathbf{Confl}_{Bel,\Box}$). Therefore, $\Diamond K_i([i:a]\,\varphi \wedge \langle i:a\rangle\,\top) \wedge K_i\neg\Box X\varphi$ implies $\Diamond K_iK_i([i:a]\,\varphi \wedge \langle i:a\rangle\,\top) \wedge K_iK_i\neg\Box X\varphi$ which in turn implies $K_i\Diamond K_i([i:a]\,\varphi \wedge \langle i:a\rangle\,\top) \wedge K_iK_i\neg\Box X\varphi$. From this and the definition of $Power(i,a,\varphi)$ we can infer $K_iPower(i,a,\varphi)$. $\Box$

According to Theorem 3.1, an agent has the power of achieving φ by doing a if and only if he correctly believes this. The same principle holds for the general concepts of power without action argument. In fact, $Power(i,\varphi) \leftrightarrow K_iPower(i,\varphi)$ is a theorem of $\mathcal{IAL}$ as well. According to Theorem 3.2 if an agent has the power of achieving φ by doing a then he correctly believes that if he does action a then he will ensure φ, no matter what the other agents will do.

5.2 Power to

As stressed in the introduction of the paper, the most basic form of agentive power is the so-called *power to*. The *power to* of an agent i concerns an action that i can do at will. One might try to formalize such a concept by constructions of the form $\Diamond\,\langle i:a\rangle\,\top$ (i.e. i can do action a). Nevertheless, these constructions are insufficient to characterize a true notion of *power to*. Suppose that i can do the action of raising an arm,

[16] The condition $\Diamond K_i Does_i\varphi$ in definition 3.C corresponds to the property "being able to conformantly bring it about that φ" studied in [7]. As definition 3.B shows, $\mathcal{IAL}$ allows us to refine this property by specifying the action on the basis of which an agent is able to conformantly bring it about that φ. In fact, the condition $\Diamond K_i([i:a]\,\varphi \wedge \langle i:a\rangle\,\top)$ could be read "i is able to conformantly bring it about that φ by doing a".

[17] But not vice versa.

i.e. $\Diamond \langle i : raiseArm \rangle \top$, and it is possible that i intends to raise his arm and he does not succeed in doing this since j blocks i's movement, i.e. $\Diamond(Goal_i \langle i : raiseArm \rangle \top \wedge \langle j : block \rangle \top \wedge [i : raiseArm] \bot)$. In this scenario, we would not say that i has the power to raise the arm. In fact, the possibility that i will raise an arm in a successful way heavily depends on what j will decide do. This example leads us to conclude that for an agent i to have the power to do action a it has to be case that:

1. i can do a;
2. if i intends to do a now then i will do a in a successful way, no matter what the other agents do.

Definition 4. *For any $i \in AGT$, $a \in ACT$*
$PowerTo(i, a) =_{def} \Box(Goal_i \langle i : a \rangle \top \rightarrow \langle i : a \rangle \top) \wedge \Diamond \langle i : a \rangle \top$

5.3 Exercise of power

There are two different and equally important views of power. On the one hand, power can be conceived as a capacity, as a potential power. On the other hand, power can be conceived as the exercise of a capacity, as the exercise of power. The notion of power defined in section 5.1 is a power in the former sense. Now, we want to look at power in the latter sense. To this end, we provide a quite general definition of *exercise of power of*. We suppose that an agent i exercises his power of achieving φ by doing a if and only if i has the power of achieving φ by doing action a and he does a. Thus, in our account i's *exercise of power of* corresponds to the fact i has a certain power *plus* the fact that i does the action on which his power is based.

Definition 5. *For any $i \in AGT$, $a \in ACT$*

A. $ExPower(i, a, \varphi) =_{def} Power(i, a, \varphi) \wedge \langle i : a \rangle \top$
B. $ExPower(i, \varphi) =_{def} \bigvee_{a \in ACT} ExPower(i, a, \varphi)$

Definitions 5.A and 5.B respectively express that i exercises his power of achieving φ by doing a and i exercises his power of achieving φ. The following theorem clarifies the relationship between exercise of power, beliefs and power to do.

Theorem 4. *For any $i \in AGT$ and $a \in ACT$*

1. $\vdash_{\mathcal{IAL}} ExPower(i, \varphi) \rightarrow Does_i \varphi$
2. $\vdash_{\mathcal{IAL}} ExPower(i, a, \varphi) \wedge Bel_i PowerTo(i, a) \rightarrow K_i Does_i \varphi$

Proof. Theorem 4.1 easily follows from **StitAct**. Here we only prove Theorem 4.2. By definition of $ExPower(i, a, \varphi)$ and
$PowerTo(i, a)$, we have that $ExPower(i, a, \varphi) \wedge Bel_i PowerTo(i, a)$ implies
$\langle i : a \rangle \top \wedge Bel_i \Box(Goal_i \langle i : a \rangle \top \rightarrow \langle i : a \rangle \top) \wedge Power(i, a, \varphi)$.
Moreover, $\langle i : a \rangle \top \wedge Bel_i \Box(Goal_i \langle i : a \rangle \top \rightarrow \langle i : a \rangle \top) \wedge Power(i, a, \varphi)$ implies $\langle i : a \rangle \top \wedge Bel_i \Box(Goal_i \langle i : a \rangle \top \rightarrow \langle i : a \rangle \top) \wedge Power(i, a, \varphi) \wedge Bel_i Goal_i \langle i : a \rangle \top$ (by Theorem 2.6) which in turn implies $\langle i : a \rangle \top \wedge Power(i, a, \varphi) \wedge Bel_i \langle i : a \rangle \top$ (by Axiom K for Bel_i and Axiom T for $\Box$). From this, by Theorem 3.2 and the definition of K_i, we can infer $K_i \langle i : a \rangle \top \wedge K_i \Box [i : a] \varphi$. Finally, from $K_i \langle i : a \rangle \top \wedge K_i \Box [i : a] \varphi$ we can conclude that $K_i Does_i \varphi$ (by Axiom **StitAct** and Axiom T for $\Box$). □

According to Theorems 4.1 and 4.2, if i exercises his power of achieving φ then he brings it about that φ and if i exercises his power of achieving φ by a and believes that he has the power to do a then he correctly believes that he brings it about that φ.

5.4 Power over

An interesting form of power on which many authors have focused is the intrinsically social power called *power over*. There is no consensus on the meaning of the expression "an agent has power over another agent with respect a given issue, fact, etc...". Several alternative definitions have been proposed. A major point of disagreement is whether i's *power over* j should be based on j's dependence on i for the achievement of his goals (*dependence-based power over*) or whether it should be based on i's ability, to affect the behavior of j by inducing j to intend to do a certain action or to refrain from doing a certain action (*power of influencing*). [18] In this section we only focus on that form of *power over* called *power of influencing*. That is, we suppose that an agent i has a power over agent j when i is in a position to induce j to intend to do certain action or in a position to induce j to want to refrain from doing a certain action. In the former case i has the power of shaping j's preferences in such a way that j will intend to do a certain action, in the latter case i has the power of shaping j's preferences in such a way that j will intend not to do a certain action. Thus, in the present analysis i's *power over* is conceived as the particular type of i's *power of* relative to the intentions of other agents in the social world. More precisely, we say that i has the *positive (viz. negative) power over* j with respect to action a if and only if i has the power of ensuring that j will intend to do (viz. will intend not to do) action a. Formally:

Definition 6. *For any* $i, j \in AGT$, $a \in ACT$

A. $PosPowerOver(i, j, a) =_{def} Power(i, Goal_j \langle j : a \rangle \top)$
B. $NegPowerOver(i, j, a) =_{def} Power(i, Goal_j [j : a] \bot)$

As in [20], we distinguish *power over* from *indirect power*. In our view, agent i has the *indirect power* of achieving φ via agent j if and only if i has the power of ensuring that j will bring it about that φ. Formally: $IndPower(i, j, \varphi) =_{def} Power(i, Does_j \varphi)$.

5.5 Effective power

In section 5.1 we have only focused on what Lukes [18] calls *operative* sense of power, that is, the power *sufficient* to produce a certain result. A more radical form of power is the so-called power in an *effective* sense, that is, the power *necessary* and *sufficient* to produce a certain result. For example, the judge in a court has the effective power of imprisoning the defendant by sentencing him to imprisonment given that he can imprison the defendant by sentencing him to imprisonment (*sufficiency*) and as long as he does not sentence the defendant, the defendant will not be imprisoned (*necessity*).

[18] As shown in [12] and [8] the two views are closely interdependent. In fact, if i has a dependence-based power over j and he knows this, then he is in a position to make threats or offers to i in order to affect his behavior thereby having a power of influencing j.

In order to formalize the notion of effective power, the definition of power given in section 5.1 has to be refined in an appropriate way. First, we have to add the conjunct $\Box([i:a]\bot \rightarrow \neg X\varphi)$ to definition 3.A (section 5.1) of opportunity. This new conjunct expresses the fact that the occurrence of action a performed by i is necessary to ensure that φ will be true next. This operation leads us to a formal definition of i's *effective opportunity of* achieving φ by doing action a, that is, i's capacity of achieving φ by a which resists to all interferences of other agents and which must be necessarily exercised by i to ensure φ (definition 7.A). From the concept of effective opportunity, it is straightforward to come up with a formal definition of *effective power*. We suppose that i has the effective power of achieving φ by doing action a if and only if i has the power of achieving φ by doing action a and correctly believes that he *must* do a in order to ensure that φ will be true next (definition 7.B). Again this can be generalized to a notion of i's effective power of achieving φ (definition 7.C).

Definition 7. *For any* $i, j \in AGT$, $a \in ACT$

A. $EffOpp(i,a,\varphi) =_{def} Opp(i,a,\varphi) \wedge \Box([i:a]\bot \rightarrow \neg X\varphi)$
B. $EffPower(i,a,\varphi) =_{def} Power(i,a,\varphi) \wedge K_i\Box([i:a]\bot \rightarrow \neg X\varphi)$
C. $EffPower(i,\varphi) =_{def} \bigvee_{a\in ACT} EffPower(i,a,\varphi)$

Theorem 5 captures an interesting relationship between effective power, exercise of power and power to.

Theorem 5. *For any* $i \in AGT$ *and* $a \in ACT$
$\vdash_{\mathcal{IAL}} EffPower(i,a,\varphi) \wedge Goal_i X\varphi \wedge PowerTo(i,a) \rightarrow ExPower(i,a,\varphi)$

According to Theorem 5, if i has the effective power of achieving φ by a, i wants φ to be true next and has the power to do a then i exercises his power of achieving φ by doing a. Moreover, from Theorems 5, 4.1 and 4.2, and the fact that $Does_i\varphi$ implies $X\varphi$, we can derive the following interesting theorem: $EffPower(i,a,\varphi) \wedge Goal_i X\varphi \wedge K_i PowerTo(i,a) \rightarrow K_i X\varphi$. Thus, if i has the effective power of achieving φ by doing a, wants φ to be true and correctly believes that he has the power to do a, then i correctly believes that he will achieve φ. The present analysis of effective power can also be extended to those forms of power called *power over* studied in section 5.4. Here, we conceive i's *effective positive* (viz. *negative*) *power over* j with respect to action a as i's effective power of inducing j to intend to do action a (viz. i' power of inducing j to intend not to do action a). The former concept is expressed by the formula $EffPower(i, Goal_j \langle j:a\rangle \top)$, the latter is expressed by $EffPower(i, Goal_j [j:a]\bot)$. $EffPower(i, Goal_j \langle j:a\rangle \top)$ corresponds to i's power to induce j to intend do a certain action a that i would not otherwise intend to do. This is similar to Dahl's concept of *power over* [10].

6 Conclusion

There are several ways in which the work presented in this paper can be advanced. An interesting direction of application is social trust theory. As in [9], we accept a definition of social trust with four arguments, that is, we would say that an agent i trusts j with

respect to a given task φ and action a, when the former wants to solve task φ and thinks that the latter has the opportunity to solve the task by doing action a, is willing and has the power to do action a. As the following abbreviation shows, such a conceptual core of trust can be formalized by exploiting the formal constructions studied in this paper: $Trust(i,j,a,\varphi) =_{def} Goal_i X\varphi \wedge Bel_i(Opp(j,a,\varphi) \wedge Goal_j \langle j : a\rangle \top \wedge PowerTo(j,a))$.
From $Trust(i,j,a,\varphi)$, a definition of trust with three arguments can be given:
$Trust(i,j,\varphi) =_{def} \bigvee_{a \in ACT} Trust(i,j,a,\varphi)$.
This is the formal translation of the expression "i trusts j with respect to a given task φ". The two definitions correspond to a form of strong trust. In fact, we can prove that in our logic both $Trust(i,j,a,\varphi)$ and $Trust(i,j,\varphi)$ imply $Bel_i Does_j \varphi$: if i trusts j with respect to a given task φ then i thinks that j is going to solve the task.

Another interesting direction of application is the theory of collective powers [8]. $\mathcal{IAL}$'s constructions for groups of agents of the form $Does_C \varphi$ can be useful for understanding how powers of coalitions interact with powers and mental attitudes of individuals. We have argued that, for an agent i to have the power of achieving φ, i must have both the objective opportunity to achieve φ and, being aware of such an opportunity. The same argument applies to collective powers. Indeed, it seems reasonable to suppose that, for a group of agents C to have the power of achieving φ, agents in C must be able to perform a joint action that will ensure φ and must be collectively aware of this, where being collectively aware of something seems to require some group belief notions such as common belief.

We also think that $\mathcal{IAL}$ is a suitable framework for studying games in strategic form and for clarifying the epistemic foundations of some game theoretic notions such as *Nash equilibrium*.[19] In the future we will investigate such an intriguing issue and try to understand how an agent's preferences are related with his goals.

References

1. T. Agotnes. Action and knowledge in alternating-time temporal logic. *Synthese*, 149(2):377–409, 2006.
2. R. Alur, T. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49:672–713, 2002.
3. B. Barnes. *The Nature of Power*. Polity Press, Cambridge, 1988.
4. N. Belnap, M. Perloff, and M. Xu. *Facing the future: agents and choices in our indeterminist world*. Oxford University Press, New York, 2001.
5. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.
6. J. Broersen, A. Herzig, and N. Troquard. Embedding alternating-time temporal logic in strategic STIT logic of agency. *Journal of Logic and Computation*, 16(5):559–578, 2006.
7. J. Broersen, A. Herzig, and N. Troquard. Normal coalition logic and its conformant extension. In *Proceedings of Eleventh Conference on Theoretical Aspects of Rationality and Knowledge (TARK XI)*, 2007.
8. C. Castelfranchi. The micro-macro constitution of power. *Protosociology*, 18-19, 2003.
9. C. Castelfranchi and R. Falcone. Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In *Proceedings of ICMAS'98*, pages 72–79, 1998.

[19] See [23] for a survey on modal logics for games.

10. R. Dahl. The concept of power. *Behavioral Science*, 2:201–215, 1957.
11. B. Dunin-Keplicz and R. Verbrugge. Collective intentions. *Fundamenta Informaticae*, 51(3):271–295, 2002.
12. A. I. Goldman. Toward a theory of social power. *Philosophical studies*, 23:221–268, 1972.
13. D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, Cambridge, 2000.
14. J. F. Horty. *Agency and Deontic Logic*. Oxford University Press, Oxford, 2001.
15. W. Jamroga and W. Van der Hoek. Agents that know how to play. *Fundamenta Informaticae*, 63(2-3):185–219, 2004.
16. E. Lorini, A. Herzig, J. Broersen, and N. Troquard. Grounding power on actions and mental attitudes. Technical report, ISTC-CNR (Rome) and IRIT (Toulouse), 2007. (http://www.istc.cnr.it/createhtml.php?nbr=83).
17. E. Lorini, N. Troquard, A. Herzig, and C. Castelfranchi. Delegation and mental states. In *Proceedings of Sixth International Joint Conference on Autonomous Agents in Multi-Agent Systems (AAMAS'07)*, ACM Press, 2007. New York.
18. S. Lukes. *Power: A Radical View (2nd Edition)*. Palgrave Macmillan, London, 2005.
19. M. Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.
20. I. Porn. *Action Theory and Social Science: Some Formal Models*. Synthese Library 120, D. Reidel,, Dordrecht, 1977.
21. J. Van Benthem. Games in dynamic-epistemic logic. *Bulletin of Economic Research*, 53(4):219–248, 2001.
22. W. Van der Hoek, W. Jamroga, and M. Wooldridge. A logic for strategic reasoning. In *Proceedings of Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05)*, ACM Press, 2005. New York.
23. W. Van der Hoek and M. Pauly. Modal logic for games and information. In J. van Benthem, P. Blackburn, and F. Wolter, editors, *Handbook of Modal Logic (3)*. Elsevier, Amsterdam, 2006.
24. W. Van der Hoek and M. Wooldridge. Cooperation, knowledge, and time: Alternating-time temporal epistemic logic and its applications. *Studia Logica*, 75:125–157, 2003.
25. W. Van der Hoek and M. Wooldridge. Towards a logic of rational agency. *Logic Journal of the IGPL*, 11:133–157, 2003.
26. B. Van Linder, W. van der Hoek, and J.-J. Ch. W., Meyer. Formalising abilities and opportunities. *Fundamenta Informaticae*, 34:53–101, 1998.
27. M. Weber. *The Theory of Social and Economic Organization*. Oxford University Press, New York, 1947.
28. M. Wooldridge. *Reasoning about rational agents*. MIT Press, Cambridge, 2000.