Towards a logical model of social agreement for agent societies

Emiliano Lorini[1] and Mario Verdicchio[2]

[1] Université de Toulouse, CNRS, Institut de Recherche en Informatique de Toulouse, France
[2] Università degli studi di Bergamo, Italy

Abstract. Multi-agent systems (MASs), comprised of autonomous entities with the aim to cooperate to reach a common goal, may be viewed as computational models of distributed complex systems such as organizations and institutions. There have been several model proposals in the agent literature with the aim to support, integrate, substitute human organizations, but no attempt has gone beyond the boundaries of this research context to become a mainstream software engineering implementation guideline, nor has it been adopted as a universal model of multi-agent interaction in economics or social sciences. In this work we counter top-down, operational organization specifications with a logical model of a fundamental concept: agreement, with the long-term aim to create a formal model of multi-agent organization that can serve as a universally accepted basis for implementation of collaborative distributed systems.

1 Introduction

Multi-agent systems (MASs) can provide an effective computational model of autonomous individuals interacting in a complex distributed system. The models that simulate the operations of multiple entities can show how agent technology can be exploited in economics and social sciences. The lack of a breakthrough so far is possibly paralleled by some lack of generality in the proposed MAS implementations. Several research works aim at proposing operational models of multi-agent organizations in the form of templates of norms, roles, interaction patterns, and so on, that have a significant impact on the agent community, but whose adoption by a wider audience may be hindered by a discrepancy between how organizations are conceived in this research context and how they actually emerge in the real world.

In this work we begin our attempt to formalize the concept of organization starting from what we consider its most fundamental component: agreement. We see an organization as a way to coordinate agent interaction that starts from an agreement between the relevant agents. Moreover, we adopt a bottom-up, formal approach to keep our analysis as general as possible, and, as a consequence, the application field of our current and future results as wide as possible.

The paper is organized as follows: Section 2 illustrates more in detail the motivations to our efforts; Section 3 presents the syntax and the semantics of our logical model, and some choices made in the model are discussed in Section 4, while Section 5 presents some theorems; Sections 6 and 7 illustrate how agreements are formed and how commitments and norms can be grounded on them, respectively; Section 8 provides some pointers to significant related literature, and, finally, Section 9 concludes.

2 Motivation

MASs can be seen as conceived with two distinct purposes. In the scenarios envisioned by the pioneers of this field, whose hopes were boosted also by the unprecedented success of Internet technologies, agents were viewed as a further development of the object-oriented paradigm, leading to the implementation of goal-driven, mobile programs that could cooperate with each other autonomously to reach a common objective. In a broader interpretation including social, economic, legal aspects, MASs are seen as a computational model of groups of interacting entities: Agents are programs that simulate a real-life complex system whose properties are to be analyzed by means of a computer system.

The lack (so far) of a so-called ‘killer application’ based on MAS technology does not mean that the latter interpretation traces the only viable path for agent researchers. Nevertheless, in our opinion, significant achievements in the simulation-oriented MAS research are a necessary step to finally reach a breakthrough also in mainstream software development. We agree with DeLoach [5]: MAS researchers have not yet demonstrated that the agent approach can yield competitive or even better solutions than other programming paradigms by providing reliable, complex, distributed systems.

We refer to virtual organizations, and think that the relevance of MAS technologies can be shown by a believable agent-based simulation of real-life, human organizations. Once agents are proven to be capable of delivering detailed models of complex organizations, then they can become a very appealing candidate for cutting-edge software solutions aiming at supporting, or even substituting, their human counterparts.

Several models of virtual organizations have been proposed in the literature [19], [8]. In particular, Electronic Institutions [17] have been introduced to regulate agent interaction in open environments. We see some issues rising from this research line: How really open are these environments with respect to the constraints introduced by the proposed organizational models? How does the operational nature of these models (as opposed to logical) affect their impact on the potential adopters? These questions are facets of our main concern: The affinity of virtual organizations with real ones is a key factor in MAS technology’s shift from research to practice. Although we can provide detailed specifications of virtual organizations in terms of roles, scenarios, interaction patterns, communication protocols and so on, we think that such approach inevitably narrows down the scope of a proposal to the researchers’ working hypotheses. The top-down specification of a predefined template is not the way organizations are born in the real world, and this distance between theoretical research and actual organizational dynamics might correspond to the gap between the agent-based proposals and the solutions adopted in the industry.

Our work has a rather different, if not opposite, starting point. We intend to provide a logical model (as opposed to operational) that allows for the formalization of the creation of organizations in a bottom-up fashion (as opposed to top-down). It might seem surprising that researchers who call for the elimination of the gap between theory and practice opt for a logic-based approach. However, this is a research field where universal models for basic concepts, including the very concept of ‘agent’, are still missing. We think that theoretical definitions of general concepts might work as wider and more solid foundations for the construction of a model of organizations that can eventually provide effective implementation guidelines. This is also the idea behind the choice of a bottom-up approach: To keep a model of organizations as general as possible, instead of trying to impose a standard template, which is a surely successful approach only in monopoly contexts, we aim at shedding some light on the basic mechanisms that lead a group of independent individuals (or autonomous agents) to form an organization.

In a top-down approach, agents join an organization with pre-established rules. In our bottom-up approach, we see an organization as the product of the agreement of several agents on how their future interactions should be regulated. Thus, the aim of this work is to formally define 'agreement' as a fundamental concept for the creation of multi-agent organizations, that is, we intend to propose a logic of social agreement.

3 A modal logic of social agreement

We present in this section the syntax and semantics of the modal logic $\mathcal{SAL}$ (*Social Agreement Logic*). The logic $\mathcal{SAL}$ specifies the conditions under which agreements are established and annulled. The main idea behind the formalism is to take *agreement* as a primitive object and to clarify its relationships with the concept of *preference* (i.e. how agreement formation depend on agents' preferences). We make a general assumption about rationality of agents in our logical approach to agreement. In particular, we suppose that the agents in a group I agree about a certain issue φ only if φ is something satisfactory for the agents in I. In other words, an agreement between certain agents is formed only if the content of agreement is something good for every agent.

3.1 Syntax

Let $ATM = \{p, q, \ldots\}$ be a nonempty set of atomic formulas, $AGT = \{i, j, \ldots\}$ a nonempty finite set of agents, and $ACT = \{\alpha, \beta, \ldots\}$ a nonempty set of atomic actions. We note $2^{ACT*} = 2^{ACT} \setminus \emptyset$ the set of all non-empty sets of actions, and $2^{AGT*} = 2^{AGT} \setminus \emptyset$ the set of all non-empty sets of agents.

We introduce a function REP that associates to every agent i in AGT a non-empty set of atomic actions called *action repertoire* of agent i:

$$REP : AGT \longrightarrow 2^{ACT*}.$$

For every agent $i \in AGT$ we define the set of i's *action tokens* of the form $i{:}\alpha$, that is,

$$\Delta_i = \{i{:}\alpha \mid \alpha \in REP(i)\}.$$

That is, $i{:}\alpha$ is an action token of agent i only if α is part of i's repertoire. We note

$$\Delta = \bigcup_{i \in AGT} \Delta_i$$

the pointwise union of the sets of possible action tokens of all agents.

The following abbreviations are convenient to speak about joint actions of groups of agents. For every non-empty set of agents I we note $JACT_I$ the set of all possible *combinations of actions* of the agents in I (or *joint actions* of the agents in I), that is,

$$JACT_I = \prod_{i \in I} \Delta_i.$$

For notational convenience we write $JACT$ instead of $JACT_{AGT}$. Elements in every $JACT_I$ are tuples noted δ_I, δ'_I, $\delta''_I, \ldots$. Elements in $JACT$ are simply noted δ, δ', $\delta'', \ldots$ For example suppose that $I = \{1, 2, 3\}$ and $\delta_I = \langle 1{:}\alpha, 2{:}\beta, 3{:}\gamma \rangle$. This means that δ_I is the joint action of the agents $1, 2, 3$ in which 1 does action α, 2 does action β

and 3 does action γ. For notational convenience, we write δ_i instead of $\delta_{\{i\}}$ for every $i \in AGT$.

The language of $\mathcal{SAL}$ is the set of formulas defined by the following BNF:

$$\varphi ::= p \mid \bot \mid \neg\varphi \mid \varphi \vee \varphi \mid \texttt{Agree}_I\varphi \mid \texttt{Do}_{i:\alpha}\varphi$$

where p ranges over ATM, i ranges over AGT, $i{:}\alpha$ ranges over Δ_i, and I ranges over 2^{AGT*}.

The classical Boolean connectives $\wedge$, $\rightarrow$, $\leftrightarrow$ and $\top$ (tautology) are defined from $\bot$, $\vee$ and $\neg$ in the usual manner.

The operators of our logic have the following reading.

- $\texttt{Agree}_I\varphi$: 'the agents in the group I agree that φ'.
- $\texttt{Do}_{i:\alpha}\varphi$: 'agent i is going to do α and φ will be true afterwards' (therefore $\texttt{Do}_{i:\alpha}\top$ is read: 'agent i is going to do α').

Operators of the form $\texttt{Agree}_I$ enable one to express those issues on which the agents in I agree, while forming a coalition. For example, $\texttt{Agree}_I\neg\mathit{smokePublic}$ expresses that the agents in I agree that people should not smoke in public spaces.

The formula $\texttt{Agree}_I\bot$ literally means that 'the agents in I agree on a contradiction'. We assign a special meaning to this formula by supposing that 'agreeing on a contradiction' means 'not being part of the same group' (or 'not forming a coalition'). This is because we assume that functioning as members of the same coalition is (at least in a minimal sense) a rational activity, and a rational group of agents cannot agree on a contradiction. Thus, $\texttt{Agree}_I\bot$ should be read 'the agents in I do not function as members of the same group' or 'the agents in I do not form a coalition' or 'the agents in I do not constitute a group'. Conversely, $\neg\texttt{Agree}_I\bot$ has to be read 'the agents in I function as members of the same group' or 'the agents in I form a coalition' or 'the agents in I constitute a group'. This concept of constituted group is expressed by the following abbreviation. For every $I \in 2^{AGT*}$:

$$\texttt{Group}(I) \stackrel{\text{def}}{=} \neg\texttt{Agree}_I\bot.$$

Note that this definition of group demands for some form of agreement, in particular if the agents in I form a coalition (i.e. $\texttt{Group}(I)$) then the agents in I agree that they form a coalition (i.e. $\texttt{Agree}_I\texttt{Group}(I)$). Indeed, as we will show in Section 3.3, our agreement operators satisfy the axiom $\neg\texttt{Agree}_I\varphi \rightarrow \texttt{Agree}_I\neg\texttt{Agree}_I\varphi$.

If I is a singleton then $\texttt{Agree}_I$ is used to express the individual preferences of agent i. That is, for every $i \in AGT$:

$$\texttt{Pref}_i\varphi \stackrel{\text{def}}{=} \texttt{Agree}_{\{i\}}\varphi.$$

Formula $\texttt{Pref}_i\varphi$ has to be read 'agent i prefers that φ is possible' (semantically this means that 'φ is true in all states that are preferred by agent i').

The following additional abbreviations will be useful to make more compact our notation in the sequel of the article. For every $i \in AGT$:

$$\texttt{Sat}_i\varphi \stackrel{\text{def}}{=} \neg\texttt{Pref}_i\neg\varphi.$$

Formula $\texttt{Sat}_i\varphi$ has to be read 'φ is a satisfactory state of affairs for agent i' (semantically this means that 'there exists at least one preferred state of agent i in which φ is true').

For every $I \in 2^{AGT*}$ and $\delta_I \in JACT_I$:

$$\mathtt{Do}_{\delta_I}\varphi \stackrel{\text{def}}{=} \bigwedge_{j \in I} \mathtt{Do}_{\delta_j}\varphi.$$

Formula $\mathtt{Do}_{\delta_I}\varphi$ has to be read 'the agents in I execute in parallel their individual actions δ_i in the vector δ_I and φ will be true after this parallel execution'. We shorten this to 'the joint action δ_I is going to be performed by group I and φ will be true afterwards'. In other words, we consider a weak notion of joint action δ_I as the parallel execution of the individual actions δ_i by every agent in I.

For every $I \in 2^{AGT*}$:

$$\mathtt{Pref}_I\varphi \stackrel{\text{def}}{=} \bigwedge_{j \in I} \mathtt{Pref}_j\varphi;$$

$$\mathtt{Sat}_I\varphi \stackrel{\text{def}}{=} \bigwedge_{j \in I} \mathtt{Sat}_j\varphi.$$

Formula $\mathtt{Pref}_I\varphi$ has to be read 'every agent in I prefers that φ is true', whilst $\mathtt{Sat}_I\varphi$ has to be read 'φ is satisfactory for every agent in I'.

3.2 Semantics

Frames of the logic $\mathcal{SAL}$ ($\mathcal{SAL}$-frames) are tuples $F = \langle W, R, A \rangle$ defined as follows.

- W is a non empty set of possible worlds or states.
- $R : \Delta \longrightarrow W \times W$ maps every possible action token $i{:}\alpha$ to a deterministic relation $R_{i:\alpha}$ between possible worlds in W.[3]
- $A : 2^{AGT*} \longrightarrow W \times W$ maps every non-empty set of agents I to a transitive[4] and Euclidean[5] relation A_I between possible worlds in W.

It is convenient to view relations on W as functions from W to 2^W; therefore we write $A_I(w) = \{w' : (w, w') \in A_I\}$ and $R_{i:\alpha}(w) = \{w' : (w, w') \in R_{i:\alpha}\}$. If $A_I(w) \neq \emptyset$ and $R_{i:\alpha}(w) \neq \emptyset$ then we say that A_I and $R_{i:\alpha}$ are defined at w.

Given a world $w \in W$, $A_I(w)$ is the set of worlds which are compatible with group I's agreements at world w. If I is a singleton $\{i\}$ then $A_{\{i\}}(w)$ is the set of worlds that agent i prefers. If $(w, w') \in R_{i:\alpha}$ then w' is the unique actual *successor* world of world w, that will be reached from w through the occurrence of agent i's action α at w. (We might also say that $R_{i:\alpha}$ is a partial function). Therefore, if $R_{i:\alpha}(w) = \{w'\}$ then at w agent i performs an action α resulting in the next state w'.

It is convenient to use $R_{\delta_I} = \bigcap_{i \in I} R_{\delta_i}$. If $R_{\delta_I}(w) \neq \emptyset$ then coalition I performs joint action δ_I at w. If $w' \in \bigcap_{i \in I} R_{\delta_i}(w)$ then world w' results from the performance of joint action δ_I by I at w.

Frames will have to satisfy some other constraints in order to be legal $\mathcal{SAL}$-frames. For every $i, j \in AGT$, $\alpha \in REP(i)$, $\beta \in REP(j)$ and $w \in W$ we have:

S1 if $R_{i:\alpha}$ and $R_{j:\beta}$ are defined at w then $R_{i:\alpha}(w) = R_{j:\beta}(w)$.

---

[3] A relation $R_{i:\alpha}$ is deterministic iff, if $(w, w') \in R_{i:\alpha}$ and $(w, w'') \in R_{i:\alpha}$ then $w' = w''$.

[4] A relation A_I is transitive iff for every $w \in W$, if $(w, w') \in A_I$ and $(w', w'') \in A_I$ then $(w, w'') \in A_I$.

[5] A relation A_I is Euclidean iff for every $w \in W$, if $(w, w') \in A_I$ and $(w, w'') \in A_I$ then $(w', w'') \in A_I$.

Constraint S1 says that if w' is the *next* world of w which is reachable from w through the occurrence of agent i's action α and w'' is also the *next* world of w which is reachable from w through the occurrence of agent j's action β, then w' and w'' denote the same world. Indeed, we suppose that every world can only have one *next* world. Note that S1 implies the determinism of every $R_{i:\alpha}$. Moreover, note that constraint S1 justifies the reading of formula $\mathtt{Do}_{i:\alpha}\varphi$ as 'agent i is going to do α and φ will be true afterwards'. Indeed, we intend to express in our logic what agents *will do* as the result of their agreement on what to do together, rather than what agents *will possibly do*.

We also suppose that every agent can perform at most one action at each world. That is, for every $i \in AGT$ and $\alpha, \beta \in REP(i)$ such that $\alpha \neq \beta$ we have:

S2 if $R_{i:\alpha}$ is defined at w then $R_{i:\beta}$ is not defined at w.

We impose the following semantic constraint for individual preferences by supposing that every relation $A_{\{i\}}$ is serial, i.e. an agent has always at least one preferred state. For every $w \in W$ and $i \in AGT$:

S3 $A_{\{i\}}(w) \neq \emptyset$.

The following semantic constraint concerns the relationship between agreements and individual preferences. For every $w \in W$ and $I, J \in 2^{AGT*}$ such that $J \subseteq I$:

S4 if $w' \in A_I(w)$ then $w' \in A_J(w')$.

According to the constraint S4, if w' is a world which is compatible with I's agreements at w and J is a subgroup of group I, then w' belongs to the set of worlds that are compatible with J's agreements at w'.

The last two semantic constraints we consider are about the relationships between preferred states of an agent and actions. For every $w \in W$, $i \in AGT$ and $\delta_i \in \Delta_i$:

S5 if R_{δ_i} is defined at w' for every $w' \in A_{\{i\}}(w)$ then R_{δ_i} is defined at w.

According to the constraint S5, if action δ_i of agent i occurs in every state which is preferred by agent i, then the action δ_i occurs in the current state.

For every $w \in W$ and $i \in AGT$:

S6 if R_{δ_i} is defined at w then there exists $I \in 2^{AGT*}$ such that $i \in I$ and R_{δ_i} is defined at w' for every $w' \in A_I(w)$.

According to the constraint S6, if agent i's action δ_i occurs at world w then there exists a group I to which i belongs such that, for every world w' which is compatible with I's agreements at w, i's action δ_i occurs at w'.

Models of the logic $\mathcal{SAL}$ ($\mathcal{SAL}$-models) are tuples $M = \langle F, V \rangle$ defined as follows.

- F is a $\mathcal{SAL}$-frame.
- $V : W \longrightarrow 2^{ATM}$ is a valuation function.

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. The rules defining the truth conditions of formulas are just standard for p, $\bot$, $\neg$ and $\vee$. The following are the remaining truth conditions for $\mathtt{Agree}_I\varphi$ and $\mathtt{Do}_{i:\alpha}$.

- $M, w \models \texttt{Agree}_I\varphi$ iff $M, w' \models \varphi$ for all w' such that $w' \in A_I(w)$
- $M, w \models \texttt{Do}_{i:\alpha}\varphi$ iff there exists $w' \in R_{i:\alpha}(w)$ such that $M, w' \models \varphi$

Note that $\texttt{Agree}_I$ is a modal operator of type necessity, whilst $\texttt{Do}_{i:\alpha}$ is of type possibility. The following section is devoted to illustrate the axiomatization of $\mathcal{SAL}$.

3.3 Axiomatization

The axiomatization of the logic $\mathcal{SAL}$ includes all tautologies of propositional calculus and the rule of inference *modus ponens* (**MP**).

(**MP**) From $\vdash_{\mathcal{SAL}} \varphi$ and $\vdash_{\mathcal{SAL}} \varphi \rightarrow \psi$ infer $\vdash_{\mathcal{SAL}} \psi$

We have the following four principles for the dynamic operators $\texttt{Do}_{i:\alpha}$.

(**K**$_{\texttt{Do}}$) $(\texttt{Do}_{i:\alpha}\varphi \wedge \neg\texttt{Do}_{i:\alpha}\neg\psi) \rightarrow \texttt{Do}_{i:\alpha}(\varphi \wedge \psi)$

(**Alt**$_{\texttt{Do}}$) $\texttt{Do}_{i:\alpha}\varphi \rightarrow \neg\texttt{Do}_{j:\beta}\neg\varphi$

(**Single**) $\texttt{Do}_{i:\alpha}\top \rightarrow \neg\texttt{Do}_{i:\beta}\top$ if $\alpha \neq \beta$

(**Nec**$_{\texttt{Do}}$) From $\vdash_{\mathcal{SAL}} \varphi$ infer $\vdash_{\mathcal{SAL}} \neg\texttt{Do}_{i:\alpha}\neg\varphi$

Dynamic operators of the form $\texttt{Do}_{i:\alpha}$ are modal operators which satisfy the axioms and rule of inference of the basic normal modal logic K (Axiom **K**$_{\texttt{Do}}$ and rule of inference **Nec**$_{\texttt{Do}}$). Moreover, according to Axiom **Alt**$_{\texttt{Do}}$, if i is going to do α and φ will be true afterwards, then it cannot be the case that j is going to do β and $\neg\varphi$ will be true afterwards. According to Axiom **Single**, an agent cannot perform more than one action at a time. This axiom makes perfectly sense in simplified artificial settings and in game-theoretic scenarios in which actions of agents and joint actions of groups never occur in parallel.

We have the following principles for the agreement operators and the preference operators, and for the relationships between agreement operators, preference operators and dynamic operators.

(**K**$_{\texttt{Agree}}$) $(\texttt{Agree}_I\varphi \wedge \texttt{Agree}_I(\varphi \rightarrow \psi)) \rightarrow \texttt{Agree}_I\psi$

(**D**$_{\texttt{Pref}}$) $\neg\texttt{Pref}_i\bot$

(**4**$_{\texttt{Agree}}$) $\texttt{Agree}_I\varphi \rightarrow \texttt{Agree}_I\texttt{Agree}_I\varphi$

(**5**$_{\texttt{Agree}}$) $\neg\texttt{Agree}_I\varphi \rightarrow \texttt{Agree}_I\neg\texttt{Agree}_I\varphi$

(**Subgroup**$_{\texttt{Agree}}$) $\texttt{Agree}_I(\varphi \rightarrow \neg\texttt{Agree}_J\neg\varphi)$ if $J \subseteq I$

(**Int1**$_{\texttt{Pref,Do}}$) $\texttt{Pref}_i\texttt{Do}_{\delta_i}\top \rightarrow \texttt{Do}_{\delta_i}\top$

(**Int2**$_{\texttt{Pref,Do}}$) $\texttt{Do}_{\delta_i}\top \rightarrow \bigvee_{i \in I} \texttt{Agree}_I\texttt{Do}_{\delta_i}$

(**Nec**$_{\texttt{Agree}}$) From $\vdash_{\mathcal{SAL}} \varphi$ infer $\vdash_{\mathcal{SAL}} \texttt{Agree}_I\varphi$

Operators for agreement of the form $\texttt{Agree}_I$ are modal operators which satisfy the axioms and rule of inference of the basic normal modal logic K45 [4] (Axioms **K**$_{\texttt{Agree}}$, **4**$_{\texttt{Agree}}$ and **5**$_{\texttt{Agree}}$, and rule of inference **Nec**$_{\texttt{Agree}}$). It is supposed that the agents in a coalition always agree on the contents of their agreements and on the contents of their disagreements (Axioms $\mathbf{4}_{\texttt{Agree}}$ and $\mathbf{5}_{\texttt{Agree}}$). That is, if the agents in I agree (resp. do not agree) that φ should be true then, they agree that they agree (resp. do not agree) that φ should be true.

We add a specific principle for individual preferences by supposing that an agent cannot have contradictory preferences (Axiom $\mathbf{D}_{\texttt{Pref}}$).

Axiom $\mathbf{Subgroup}_{\texttt{Agree}}$ is about the relationship between the agreements of a group and the agreements of its subgroups. The agents of a group I agree that φ should be true only if there is no subgroup J of I such that J agree that φ should be false. A specific instance of Axiom $\mathbf{Subgroup}_{\texttt{Agree}}$ is $\texttt{Agree}_I(\varphi \rightarrow \texttt{Sat}_i\varphi)$ if $i \in I$. This means that the agents of a group I agree that φ should be true only if φ is satisfactory for every agent in I. A more detailed explanation of the logical consequences of Axiom $\mathbf{Subgroup}_{\texttt{Agree}}$ is given in Section 5.

Axiom $\mathbf{Int1}_{\texttt{Pref},\texttt{Do}}$ and Axiom $\mathbf{Int2}_{\texttt{Pref},\texttt{Do}}$ are general principles of intentionality describing the relationship between an agent's action, his preferences, and the agreements of the group to which the agent belongs. According to Axiom $\mathbf{Int1}_{\texttt{Pref},\texttt{Do}}$, if agent i prefers that he performs action δ_i (δ_i occurs in all states that are preferred by agent i) then agent i starts to perform action δ_i. A similar principle for the relationship between individual intentions and action occurrences has been studied in [15]. According to Axiom $\mathbf{Int2}_{\texttt{Pref},\texttt{Do}}$, if an agent i starts to perform a certain action δ_i then it means that either agents i prefers to perform this action or there exists some group I to which agent i belongs such that the agents in I agree that i should perform action δ_i. In other terms, an agent i's action δ_i is intentional in a general sense: either it is driven by i's intention to perform action δ_i or it is driven by the collective intention that i performs action δ_i of a group I to which agent i belongs.

We call $\mathcal{SAL}$ the logic axiomatized by the axioms and rules of inference presented above. We write $\vdash_{\mathcal{SAL}} \varphi$ if formula φ is a theorem of $\mathcal{SAL}$ (i.e. φ is the derivable from the axioms and rules of inference of the logic $\mathcal{SAL}$). We write $\models_{\mathcal{SAL}} \varphi$ if φ is *valid* in all $\mathcal{SAL}$-models, i.e. $M, w \models \varphi$ for every $\mathcal{SAL}$-model M and world w in M. Finally, we say that φ is *satisfiable* if there exists a $\mathcal{SAL}$-model M and world w in M such that $M, w \models \varphi$. We can prove that the logic $\mathcal{SAL}$ is *sound* and *complete* with respect to the class of $\mathcal{SAL}$-frames. Namely:

Theorem 1 *$\mathcal{SAL}$ is determined by the class of $\mathcal{SAL}$-frames.*

Proof. It is a routine task to check that the axioms of the logic $\mathcal{SAL}$ correspond one-to-one to their semantic counterparts on the frames. In particular, Axioms $\mathbf{4}_{\texttt{Agree}}$ and $\mathbf{5}_{\texttt{Agree}}$ correspond to the transitivity and Euclideanity of every relation A_I. Axiom $\mathbf{D}_{\texttt{Pref}}$ corresponds to the seriality of every relation $A_{\{i\}}$ (constraint S3). Axiom $\mathbf{Alt}_{\texttt{Do}}$ corresponds to the semantic constraint S1. Axiom **Single** corresponds to the semantic constraint S2. Axiom $\mathbf{Subgroup}_{\texttt{Agree}}$ corresponds to the semantic constraint S4. Axiom $\mathbf{Int1}_{\texttt{Pref},\texttt{Do}}$ corresponds to the semantic constraint S5. $\mathbf{Int2}_{\texttt{Pref},\texttt{Do}}$ corresponds to the semantic constraint S6.

It is routine, too, to check that all of our axioms are in the Sahlqvist class. This means that the axioms are all expressible as first-order conditions on frames and that they are complete with respect to the defined frames classes, cf. [2, Th. 2.42]. □

4 Discussion

One might wonder why we did not include a principle of monotonicity of the form $\texttt{Agree}_I\varphi \to \texttt{Agree}_J\varphi$ for $J \subseteq I$ in our logic of agreement: for every sets of agents I and J such that $J \subseteq I$, if the agents in I agree that φ should be true then the agents in the subgroup J agree that φ should be true as well. We did not do include this principle because we think that it is not sufficiently general to be applied in all situations. Indeed, a minority group J of a larger group I might exist which does not have the same view than the larger group. For example, imagine I is the group of members of a political party who are choosing the leader of the party for the next years. All agents in I agree that a certain member of the party called Mr. Brown should be the leader for the next years. This is the official position of the party. At the same time, a small minority of I in conspiracy agree that Mr. Black should be the leader.

Consider now the following principle $\texttt{Agree}_I(\varphi \to \bigwedge_{i\in I}\texttt{Pref}_i\varphi)$ and even the weaker $\texttt{Agree}_I(\varphi \to \bigvee_{i\in I}\texttt{Pref}_i\varphi)$: every group of agents I agree that φ should be true only if all of them prefer φ, and every group of agents I agree that φ should be true only if some of them prefers φ. These two principles are also too strong. Indeed, the agents in a group I might agree that φ should be true, without claiming that φ must be preferred by every agent in I and without claiming that φ must be preferred by some agent in I. For example, the members of a community I might agree that taxes should be payed by every agent in I without claiming and agreeing that tax payment must be preferred by every agent in I, and without claiming and agreeing that tax payment must be preferred by some agent in I. The members of the community just agree that tax payment must be something preferable by the whole community.

Finally, let us explain why we did not include stronger versions of Axiom $\mathbf{Int1}_{\texttt{Pref},\texttt{Do}}$ and Axiom $\mathbf{Int2}_{\texttt{Pref},\texttt{Do}}$ of the form $\texttt{Agree}_I\texttt{Do}_{\delta_I}\top \to \texttt{Do}_{\delta_I}\top$ and $\texttt{Do}_{\delta_I}\top \to \texttt{Agree}_I\texttt{Do}_{\delta_I}\top$ in the axiomatization of our logic $\mathcal{SAL}$ for every $I \in 2^{AGT*}$. On the one hand $\texttt{Agree}_I\texttt{Do}_{\delta_I}\top \to \texttt{Do}_{\delta_I}\top$ is too strong because autonomous agents should be capable to violate norms and to decide not to conform to agreements with other agents (see Section 7). For example, agents might agree at the public level that each of them should pay taxes (i.e. $\texttt{Agree}_{\{1,\ldots,n\}}\texttt{Do}_{\langle 1:payTaxes,\ldots,n:payTaxes\rangle}\top$) but, in private, some of them does not pay taxes (i.e. $\neg\texttt{Do}_{\langle 1:payTaxes,\ldots,n:payTaxes\rangle}\top$). On the other hand, $\texttt{Do}_{\delta_I}\top \to \texttt{Agree}_I\texttt{Do}_{\delta_I}\top$ is too strong because there are situations in which the agents in a set I perform a joint action δ_I without agreeing that such a joint action should be performed. Each agent in I is doing his part in δ_I without caring what the other agents in I do. For example, i might be cooking while j is reading a book without reciprocally caring what the other does, and without agreeing that the action of cooking performed by i and the action of reading performed by j should occur together. One might say that i and j do not have *interdependent reasons* for jointly preferring that i cooks while j reads a book.

5 Some $\mathcal{SAL}$-theorems

Let us now discuss some $\mathcal{SAL}$-theorems. The first group of theorems present some generalizations of Axioms $\mathbf{Alt}_{\texttt{Do}}$ and **Single** for joint actions of groups.

Proposition 1. *For every $I, J \in 2^{AGT*}$ and $\delta_I, \delta'_I, \delta_J$ such that $\delta_I \neq \delta'_I$:*

(1a) $$\vdash_{\mathcal{SAL}} \mathtt{Do}_{\delta_I}\varphi \rightarrow \neg\mathtt{Do}_{\delta_J}\neg\varphi$$

(1b) $$\vdash_{\mathcal{SAL}} \mathtt{Do}_{\delta_I}\top \rightarrow \neg\mathtt{Do}_{\delta'_I}\top$$

According to Theorem 1a, if group I is going to perform the joint action δ_I and φ will be true afterwards, then it cannot be the case that group J is going to perform the joint action δ_J and φ is going to be false afterwards. According to Theorem 1b, every group of agents can never perform more than one joint action at a time.

The second group of theorems present some interesting properties of agreement. Theorems 2a and 2b are derivable from Axioms $\mathbf{4}_{\mathtt{Agree}}$, $\mathbf{5}_{\mathtt{Agree}}$ and $\mathbf{D}_{\mathtt{Pref}}$. According to these two theorems, the agents in I agree (resp. do not agree) that φ if and only if they agree that they agree (resp. do not agree) that φ. According to Theorem 2c, a group of agents I can intend to perform at most one joint action. Theorem 2d expresses a unanimity principle for agreement: for every set of agents I, the agents in I agree that if each of them prefers φ then φ should be the case. Theorem 2e expresses an interesting property about coalition formation and coalition disintegration: if the agents in I agree that a minority part J of I agrees that φ and another minority part J' of I agrees that $\neg\varphi$, then the agents in I do not form a coalition (i.e. I is not a constituted group).

Proposition 2. *For every $I, J, J' \in 2^{AGT*}$ and δ_I, δ'_I such that $\delta_I \neq \delta'_I$ and $J, J' \subseteq I$:*

(2a) $$\vdash_{\mathcal{SAL}} \mathtt{Agree}_I\varphi \leftrightarrow \mathtt{Agree}_I\mathtt{Agree}_I\varphi$$

(2b) $$\vdash_{\mathcal{SAL}} \neg\mathtt{Agree}_I\varphi \leftrightarrow \mathtt{Agree}_I\neg\mathtt{Agree}_I\varphi$$

(2c) $$\vdash_{\mathcal{SAL}} \mathtt{Agree}_I\mathtt{Do}_{\delta_I}\top \rightarrow \neg\mathtt{Agree}_I\mathtt{Do}_{\delta'_I}\top$$

(2d) $$\vdash_{\mathcal{SAL}} \mathtt{Agree}_I(\bigwedge_{i\in I}\mathtt{Pref}_i\varphi \rightarrow \varphi)$$

(2e) $$\vdash_{\mathcal{SAL}} \mathtt{Agree}_I(\mathtt{Agree}_J\varphi \wedge \mathtt{Agree}_{J'}\neg\varphi) \rightarrow \neg\mathtt{Group}(I)$$

At the current stage, our logic does not allow to deal with situations in which I is a constituted group, the agents I agree about a certain fact φ and, at the same time, they agree that some agents in I prefer $\neg\varphi$. Formally, by Theorem 2e and Axiom $\mathbf{4}_{\mathtt{Agree}}$, we can prove that formula $\mathtt{Agree}_I\varphi \wedge \mathtt{Agree}_I\mathtt{Pref}_i\neg\varphi$ implies $\neg\mathtt{Group}(I)$, if $i \in I$. This means that at the current stage our logic $\mathcal{SAL}$ does not allow to handle collective decisions based on special procedures like majority voting in which certain agents might find a collective agreement about something while agreeing that it is not based on unanimous preferences. For example, the agents in I might be the members of the Parliament of a certain country and form a coalition (i.e. $\mathtt{Group}(I)$). They might collectively decide by majority voting to declare war upon another country (i.e. $\mathtt{Agree}_I war$), although they agree that there is a (pacifist) minority $i, j \in I$ preferring that war is not declared upon another country (i.e. $\mathtt{Agree}_I(\mathtt{Pref}_i\neg war \wedge \mathtt{Pref}_j\neg war)$). Although we are aware that this is a limitation of our proposal, we think that our logic of agreement is still sufficiently general to model informal and non-structured groups in which there are no special voting procedures nor special roles (e.g. legislators, officials of the law, etc.) which are responsible for agreement creation. In fact, in such a kind of groups agreements are often about solutions to coordination (or cooperation) problems which are satisfactory for all agents in the group (e.g. some agents find an agreement to have dinner together at a Japanese restaurant rather than at an Indian restaurant).

6 Reaching an agreement on what to do together

We can provide in our logic $\mathcal{SAL}$ the formal specification of some additional principles explaining how some agents might reach an agreement on what to do together starting from their individual preferences. We do not intend to add these principles to the axiomatization of $\mathcal{SAL}$ presented in Section 3.3. We just show that $\mathcal{SAL}$ is sufficiently expressive to capture them both syntactically and semantically so that the can be easily integrated into our formal framework. The principles we intend to characterize are specified in terms of agreements about the conditions under which a certain joint action should be performed.

In certain circumstances, it is plausible to suppose that a group of agents I agree that if there exists a unique satisfactory joint action δ_I for all agents in I, then such a joint action should occur. In other terms, the agents in a group I agree on the validity of the following general principle: 'Do together the joint action δ_I, if it is the only joint action that satisfies every agent in I!'. This criteria is often adopted by groups of agents in order to find cooperative solutions which are satisfactory for all them. For example, in a Prisoner Dilemma scenario with two agents i and j the joint action $\langle i{:}\mathit{cooperate},j{:}\mathit{cooperate}\rangle$ is the only satisfactory solution for both agents. If the two agents i and j agree on the previous principle and face a PD game, then they will agree that $\langle i{:}\mathit{cooperate},j{:}\mathit{cooperate}\rangle$ is the joint action that they should perform. The previous principle of agreement creation is formally expressed in our logic as follows. For every $I \in 2^{AGT*}$ and $\delta_I \in JACT_I$:

(*) $$\mathtt{Agree}_I((\mathtt{Sat}_I\mathtt{Do}_{\delta_I}\top \wedge \bigwedge_{\delta'_I \neq \delta_I} \neg\mathtt{Sat}_I\ \mathtt{Do}_{\delta'_I}\top) \rightarrow \mathtt{Do}_{\delta_I}\top)$$

Principle * corresponds to the following semantic constraint over $\mathcal{SAL}$-frames. For every $w \in W$, $I \in 2^{AGT*}$ and $\delta_I \in JACT_I$:

S6 if $w' \in A_I(w)$ and $A_{\{i\}} \circ R_{\delta_I}(w') \neq \emptyset$ for all $i \in I$ and, for all $\delta'_I \neq \delta_I$ there exists $i \in I$ such that $A_{\{i\}} \circ R_{\delta'_I}(w') = \emptyset$ then, $R_{\delta_I}(w') \neq \emptyset$

where $A_{\{i\}} \circ R_{\delta_I}(w')$ is defined as $\bigcup\{R_{\delta_I}(v) \mid v \in S_i(w')\}$.

If we suppose that the Principle * is valid then the following consequence is derivable for every $I \in 2^{AGT*}$ and $\delta_I \in JACT_I$:

(3) $$\mathtt{Agree}_I(\mathtt{Sat}_I\mathtt{Do}_{\delta_I}\top \wedge \bigwedge_{\delta'_I \neq \delta_I} \neg\mathtt{Sat}_I\ \mathtt{Do}_{\delta'_I}\top) \rightarrow \mathtt{Agree}_I\mathtt{Do}_{\delta_I}\top$$

REMARK. Note that in the previous Principles * and 3 of agreement creation *mutual trust* between the agents in the group is implicitly supposed, that is, it is supposed that every agent i in I thinks it possible that the other agents in I will do their parts in the joint action δ_I. Indeed, trust between the members of the group is a necessary condition for agreement creation (on this point, see [1] for instance). We postpone to future work a formal analysis of the relationships between trust and agreement. To this aim, we will have to extend our logic $\mathcal{SAL}$ with doxastic modalities to express agents' beliefs.

Example 1. Imagine a situation of exchange of goods in EBay between two agents i and j. Agent i is the buyer and agent j is the seller. They have to perform a one-shot trade transaction. We suppose $AGT = \{i, j\}$. Agent i has the following two actions available: pay and $skip$ (do nothing). Agent j has the following two actions available: $send$ and $skip$ (do nothing). That is, $\Delta_i = \{i{:}send, i{:}skip\}$ and $\Delta_j = \{j{:}pay, j{:}skip\}$. Therefore, the set of possible joint actions of the two agents is

$$JACT = \{\langle i{:}skip, j{:}skip\rangle, \langle i{:}send, j{:}skip\rangle, \langle i{:}skip, j{:}pay\rangle, \langle i{:}send, j{:}pay\rangle\}.$$

The two agents i and j agree that the situation in which i sends the product and j pays is satisfactory for both of them.

(A) $\texttt{Agree}_{\{i,j\}}\texttt{Sat}_{\{i,j\}}\texttt{Do}_{\langle i:send, j:pay\rangle}\top$.

Moreover, agent i and agent j agree that the situation in which i does nothing and j pays the product, the situation in which i sends the product and j does not nothing, and the situation in which i and j do nothing, always leave one of them unhappy. Thus we have that agent i and agent j agree that there is no other situation different from $\langle i{:}send, j{:}pay\rangle$ that is a satisfactory situation for both of them:

(B) $\texttt{Agree}_{\{i,j\}} \bigwedge_{\delta'_{\{i,j\}} \neq \langle i:send, j:pay\rangle} \neg\texttt{Sat}_{\{i,j\}}\texttt{Do}_{\delta'_I}\top$.

From items A and B, by using Principle 3, we infer that agent i and agent j agree that they should perform the joint action $\langle i{:}send, j{:}pay\rangle$:

(C) $\texttt{Agree}_{\{i,j\}}\texttt{Do}_{\langle i:send, j:pay\rangle}\top$.

Other conditions under which the agents in a group can reach an agreement on what to do together could be studied in our logical framework. For instance, one might want to have general principles of the following form which can be used to find a solution in coordination problems. Suppose that δ_I and δ'_I are both satisfactory joint actions for all agents in group I. Moreover, there are no joint actions δ''_I different from δ_I and δ'_I which are satisfactory for all agents in I. Then, either the agents in I agree that δ_I should be performed or they agree that δ'_I should be performed. In other terms, if the agents in a group I face a coordination problem then they strive to find a solution to this problem.

7 Grounding norms and commitments on agreements

The logic of agreement $\mathcal{SAL}$ presented in the previous section provides not only a formal framework in which the relationships between individual preferences of agents in a group and group agreements can be studied, but also it suggests a different perspective on concepts traditionally studied in the field of deontic logic.

Consider for instance deontic statements of the following form "within the context of group I it is required that agent i will perform action δ_i" or "within the context of group I it is required that agent i will perform his part in the joint action δ_I together with the other agents in I". These statements just say that i has a *directed obligation* towards his group I to do a certain action as part of a joint plan of the group I (see e.g. [13, 14] for a different perspective on directed obligations). By way of example, imagine the situation in which agent i and agent j are trying to organize a party together. After a brief negotiation, they conclude that i will prepare the cake, while j will buy drinks for the party. In this situation, "within the context of group $\{i, j\}$ it is required that agent i will prepare the cake for the party and it is required that agent j will buy drinks for the party". The following abbreviation expresses the classical deontic notion of directed obligation in terms of the concept of agreement. For every $I \in 2^{AGT*}$, $i \in I$ and $\delta_i \in \Delta_i$:

$$\texttt{Oblig}_i(\delta_i, I) \stackrel{\text{def}}{=} \texttt{Agree}_I \texttt{Do}_{\delta_i} \top.$$

Formula $\texttt{Oblig}_i(\delta_i, I)$ has to be read 'within the context of group I it is required that agent i will perform action δ_i'.[6] It is to be noted that the notion of directed obligation represents an essential constituent of the notion of *social commitment*. Thus, in our approach, an essential aspect of an agent i's commitment with respect to his group I to do a certain action δ_i is the fact that all agents in I agree that i should perform action δ_i. Since all agents in the group I agree on this, they are entitled to require agent i to perform this action. Moving beyond the notion of directed obligation as an essential constituent of social commitment, our logic $\mathcal{SAL}$ can be used to provide a a formal characterization of the notion of *mutual (directed) obligation* in a group I, as 'every agent in I is required to perform his part in a joint action δ_I of the group'. Formally, for every $I \in 2^{AGT*}$ and $\delta_I \in JACT_I$:

$$\texttt{MutualObg}_I(\delta_I) \stackrel{\text{def}}{=} \bigwedge_{i \in I} \texttt{Oblig}_i(\delta_i, I).$$

Formula $\texttt{MutualObg}_I(\delta_I)$, which is equivalent to $\texttt{Agree}_I \texttt{Do}_{\delta_I} \top$, has to be read 'the agents in the group I are mutually obliged to perform their parts in the joint action δ_I'. This notion of mutual (directed) obligation is an essential constituent of the notion of *mutual (social) commitment*. As already emphasized in Section 4, in our logic agents can violate obligations assigned to them (breaking their social commitments). Violation of a directed obligation is expressed in our logic by the construction $\texttt{Oblig}_i(\delta_i, I) \wedge \neg \texttt{Do}_{\delta_i} \top$: within the context of group I it is required that agent i will perform action δ_i, but agent i does not perform action δ_i. The discussion on the notion of commitment will be extended in Section 8 where our approach will be compared with some formal approaches to agreement recently proposed in the MAS area.

8 Related work

The literature on agents, organizations, agreements is too vast to be given an exhaustive overview here: let us provide pointers to some significant works that relate to our effort or that are set against it in a way that stimulates discussion. We take inspiration from Garcia et al. [10] to determine the dimensions along which multi-agent organizational concepts are developed: structural, functional, dynamic, and normative.

[6] Note that $\texttt{Oblig}_i(\delta_i, I)$ captures a specific notion of obligation based on agreement. We are aware that other forms of obligations exist in social life like legal obligations or moral obligations (an agent may feel obliged to do something for his own moral reasons).

From a conceptual perspective, the formalization of agreements comes before any structural consideration. As an example, Dignum et al. [7] propose an attempt to describe minimum requirements for agents to be organized into an institution. The minimum requirements rely on an existing institution designed with the ISLANDER tool [8], and call for a middle agent. ISLANDER is probably the most complete tool to date for the specification of institutions in terms of roles and scenes. The tool is kept as general as possible to allow for the widest possible variety of definable institutions. Nevertheless, the specification of an institution is entirely performed by a human designer, and agents join an institution by assuming one or more roles, with no account of the process by which individuals look for and reach agreements that give rise to an institution. Thus, the platform is an effective means to translate an existing institution into a MAS, but the automatization of the creation of organizations is out of scope.

With respect to the functional dimension of organizations, that is, their goals and how to achieve them, we adopt the conceptual distinction drawn by Griffiths and Luck [12] between teamwork and coalition formation. The former is seen as focused on task assignment and action coordination among agents in the short term, whereas the latter is said to be dealing with the establishment in the long term of a group of agents with a common aim or goal. Although agreeing with the authors in viewing multi-agent organizations as a means to achieve long-termed objectives and in considering trust as a key concept for an organization that influences an agent's decisions on undertaking cooperations, our focus is slightly different in the context of this research: we consider an organization to be born when an agreement is made, so our efforts are on the formalization of agreements. An investigation on the relationship between trust and agreements lies ahead in our research path. Nevertheless, we share the authors' aim to determine the basic principles that lead to the creation of organizations, as opposed to several coalition formation research works optimizing match-making algorithms between a set of tasks and a set of agent capabilities (e.g.: [18]).

We can consider dynamic and normative dimensions as intrinsic to any attempt to formalize a concept like a multi-agent organization. Dynamic aspects include the formation and the evolution of coalitions of agents and, on a smaller scale, the preconditions and the consequences of an agent's action. When these conditions deal with deontic concepts the MAS is characterized also by a normative dimension. An important question is which normative concept or set of concepts to choose as the fundamental basis for the formalization of organizations. Dignum et al., for instance, choose violation as a fundamental concept to define deadlines in a MAS [6]. Violation is surely a very important concept in any normative context, and especially in those where deadlines are the central focus. Nevertheless, we argue that it does not play a primary role when one wants to deal with a more general overview of organizations, especially electronic ones. As pointed out by Cardoso et al. [3], while in real life a coercive action is eventually enforced against individuals not able or willing to abide by a sanction deriving by some misdemeanor, such coercions are not (yet?) implementable in a distributed information system, so that effectiveness of violations and relevant sanctions is somehow diminished. Boella and van der Torre [9] propose contracts, defined as a special type of beliefs ascribed to a group of agents, as a foundational means for the creation of legal institutions. We follow a similar approach, but consider agreement as a more basic concept that precedes contracts.

In [20] a formal approach to multiparty agreement between agents is proposed based on the notion of social commitment: A multiparty agreement among the agents in $\{1, \ldots, n\}$ is given by a set of commitments $\{C_1, \ldots, C_n\}$ where C_i the commitment that agent i has towards the other agents. In this approach the notion of commitment is taken as a primitive concept and the notion of agreement is built on it, which we counter by starting from a primitive notion of agreement in a group I (which depends on the individual preferences of the agents in I), on the top of which we built a notion of directed obligation, the essential constituent of the notion of social commitment. The logic of agreement $\mathcal{SAL}$ has some similarities with the logical framework based on the concept of acceptance we presented in [16] and [11], in which a logical analysis of the relationships between the rules and norms of an institution and the *acceptances* of these rules and norms by the members of the institution has been provided. However, these works do not analyze the relationships between individual preferences of agents and collective acceptances (or agreements) which we presented in this paper.

9 Conclusions

This work is a starting point of a long enterprise. Our long-term aim is to provide a formal specification of all the basic notions that characterize organizations in general, including those in the real world, and we started with what we consider to be a fundamental concept: agreement. The formalization of this concept with a logical approach aims at analyzing in detail both its static characteristics and its dynamic properties, that is, what is meant by the term agreement and how it is supposed to influence agents' behavior when cooperation is the common goal. Once a model is universally established which is formal and general enough to abstract from particular types of organizations or specific operational details, such a model may be used as a sound basis for an agent-based implementation that can really have a significant impact on economic or social scientific contexts. The relations between our formalization of agreement and the notions of norms and commitments have been investigated in this work, but other dimensions of multi-agent interaction, such as trust, are still to be tackled, which is what we intend to pursue in the future.

10 Acknowledgements

Emiliano Lorini is supported by the project ForTrust "Social trust analysis and formalization" ANR-06-SETI-006 financed by the French ANR.

References

1. G. Andrighetto, L. Tummolini, C. Castelfranchi, and R. Conte. A convention or (tacit) agreement between us. In V. F. van Benthem, J. Hendricks, J. Symons, and S. A. Pedersen, editors, *Between Logic and Intuition: David Lewis and the Future of Formal Methods*, Philosophy Synthese Library. Springer. to appear.

2. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.
3. H. L. Cardoso, A. P. Rocha, and E. Oliveira. Supporting virtual organizations through electronic institutions and normative multi-agent system. In J. P. Rennard, editor, *Handbook of Research on Nature Inspired Computing for Economy and Management*. Idea Group, 2006.
4. B. F. Chellas. *Modal logic: an introduction*. Cambridge University Press, Cambridge, 1980.
5. S. A. DeLoach. Moving multiagent systems from research to practice. In *Future of Software Engineering and Multi-Agent Systems (FOSE-MAS)*, 2008.
6. F. Dignum, J. Broersen, V. Dignum, and J. J. Meyer. Meeting the deadline: Why, when and how. In *Formal Approaches to Agent-Based Systems*, pages 30–40. Springer, 2005.
7. F. Dignum, V. Dignum, J. Thangarajah, L. Padgham, and M. Winikoff. Open agent systems??? In *Agent-Oriented Software Engineering VIII*, pages 73–87, 2007.
8. M. Esteva, D. de la Cruz, and C. Sierra. Islander: an electronic institutions editor. In *Proceedings of AAMAS '02*, pages 1045–1052. ACM Press, 2002.
9. G. Boella and L. van der Torre. Contracts as legal institutions in organizations of autonomous agents. In *Proceedings of AAMAS '04*, pages 948–955. ACM Press, 2004.
10. E. Garcia, Argente E., Giret A., and Botti V. Issues for organizational multiagent systems development. In *Sixth International Workshop From Agent Theory to Agent Implementation (AT2AI-6)*, pages 59–65, 2008.
11. B. Gaudou, D. Longin, E. Lorini, and L. Tummolini. Anchoring institutions in agents' attitudes: Towards a logical framework for autonomous mas. In *Proceedings of AAMAS '08*, pages 728–735. ACM Press, 2008.
12. N. Griffiths and M. Luck. Coalition formation through motivation and trust. In *Proceedings of AAMAS '03*, pages 17–24. ACM Press, 2003.
13. S. Kanger and H. Kanger. Rights and parliamentarism. *Theoria*, 6(2):85–115, 1966.
14. L. Lindahl. Stig Kanger's theory of rights. In G. Holmström-Hintikka, S. Lindström, and R. Sliwinski, editors, *Collected Papers of Stig Kanger with Essays on his Life and Work*, volume 2, pages 151–171. Kluwer, Dordrecht, 2001.
15. E. Lorini and A. Herzig. A logic of intention and attempt. *Synthese*, 163(1):45–77.
16. E. Lorini, D. Longin, B. Gaudou, and A. Herzig. The logic of acceptance: Grounding institutions on agents' attitudes. *Journal of Logic and Computation*. to appear.
17. P. Noriega and C. Sierra. Electronic institutions: Future trends and challenges. In *Proceedings of the 6th International Workshop on Cooperative Information Agents*, pages 14–17, 2002.
18. O. Shehory, K. P. Sycara, and S. Jha. Multi-agent coordination through coalition formation. In *Proceedings of ATAL '97*, pages 143–154. Springer-Verlag, 1998.
19. K. Sycara, M. Paolucci, N. van Velsen, and J. A. Giampapa. The Retsina MAS infrastructure. *Autonomous Agents and MAS*, 7(1-2), 2003.
20. F. Wan and M. P. Singh. Formalizing and achieving multiparty agreements via commitments. In *Proceedings of AAMAS '05*, pages 770–777. ACM Press, 2005.