Anchoring Institutions in Agents' Attitudes: Towards a Logical Framework for Autonomous Multi-Agent Systems

Benoit Gaudou
IRIT, Toulouse (France)
Benoit.Gaudou@irit.fr

Dominique Longin
IRIT, Toulouse (France)
Dominique.Longin@irit.fr

Emiliano Lorini
IRIT, Toulouse (France)
Emiliano.Lorini@irit.fr

Luca Tummolini
ISTC-CNR, Rome (Italy)
luca.tummolini@istc.cnr.it

ABSTRACT

The aim of this paper is to provide a logical framework for the specification of *autonomous* Multi-Agent Systems (MAS). A MAS is autonomous in so far as it is capable of binding ('nomos') itself ('auto') independently of any external normative constraint specified by a designer. In particular, a MAS is autonomous if it is able to maintain its social institutions (*i.e.* rule-governed social practices) only by way of the agents' attitudes. In order to specify an autonomous MAS, we propose the logic $\mathcal{AL}$ (*Acceptance Logic*) in which the acceptance of a proposition by the agents *qua* group members (*i.e.* group acceptance) is introduced. Such propositions are true w.r.t. an institutional context and correspond to facts that are instituted in an attitude-dependent way (*i.e.* normative and institutional facts). Finally, we contend that the present approach paves the way for a foundation of legal institutions, for studying the interaction between social and legal institutions and, eventually, for understanding and modeling institutional change.

Categories and Subject Descriptors

I.2.4 [**Knowledge Representation Formalisms and Methods**]: Modal logic; I.2.11 [**Distributed Artificial Intelligence**]: Intelligent agents; I.2.0 [**General**]: Philosophical foundations.

General Terms

Theory.

Keywords

Normative systems, logics for agent systems, modal logic.

1. INTRODUCTION

Autonomous agents that interact with each other (and with human beings) pose at least two general problems: they should be able to achieve some level of coordination in order to accomplish their distributed tasks and, notwithstanding their autonomy and self-interest, they should be somehow influenceable towards the fulfillment of some collective goal. One possible way to tackle these problems is to devise artificial *institutions* ([23, 9]). Following the classical work of Douglass North artificial institutions are usually conceived as human-like: "the rules of the game in a society or the humanly devised constraints that structure agents' interaction" ([24, p. 3]). With this model in mind, AI practitioners have interpreted their task as that of advancing logical or computational frameworks to represent institutions, while leaving to the agents' autonomy the decision whether to comply or not with the specified rules ([1, 8]). This approach, however, has at least three strong limitations. First of all, the institutions are not only constraints but also 'enablements' ([26]): new possibilities of actions (*i.e.* institutional actions like paying, marrying, promising *etc.*) are possible when an institution is in place. Secondly, artificial institutions are usually inspired by human legal ones which, however, are only a small part of the institutionalized human interactions. Moreover, to work effectively, legal institutions should interact with informal ones.[1] Finally, and more importantly, institutions should be constructed by the agents themselves and not imposed from the outside.

More precisely, while it is a widely shared that, in order to face complex and dynamical problems, the individual agents must be autonomous, less emphasis is devoted to the fact that the multi-agent systems (MAS) themselves (for exactly the same reasons) should be conceived and designed to be autonomous. In fact, etymologically, autonomous means self-binding ('auto' and 'nomos'), and an autonomous MAS is the vision of an artificial society that is able to create, maintain, and eventually change its own institutions by itself, without the intervention of the external designer in this process.

This challenge is also strongly tied to the new trend of designing self-organizing MASs but, in contrast to many efforts in the area, we are after a notion of self-organization that is amenable for, and can make profit of, more complex cognitive agents (*i.e.* BDI-like; see [7] for the general approach). In fact, quoting North again [21, p. 77]:

> "Only because institutions are anchored in peoples minds do they ever become behaviorally relevant. The *elucidation of the internal aspect is the crucial step* in adequately explaining the emergence, evolution, and effects of institutions (emphasis added)."

In this paper, we aim to provide a logical framework for the specification of autonomous MASs, that is, MASs whose agents are ca-

Cite as: Anchoring Institutions in Agents' Attitudes: Towards a Logical Framework for Autonomous MAS, Gaudou, B., Longin, D., Lorini, E. and Tummolini, L., *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. XXX-XXX.

[1] Following [24], we consider informal such institutions as social norms and social practices (*i.e.* promise). In his seminal book North explicitly states the relevance of this informal layer but still this component is widely neglected in the MAS literature. On the importance of informal normative relations in social contexts see [5].

pable of creating and maintaining their institutions by themselves (Section 3). The focus of this contribution is on modeling social or informal institutions, rather than legal ones. Social institutions are the basic structures of a society on top of which more complex legal ones are constructed. By social or informal institutions, we refer to *rule-governed social practices* in which no member with 'special' powers is introduced.[2] More specifically, we will introduce the notion of an agent's acceptance of a proposition *qua* group member in a given institutional context (Section 2), and we will study its interaction with different notions such that of common belief and private belief (Section 4). On the basis of these attitudes *qua* group members, we will specify how a group can create and maintain normative and institutional facts which hold only in an attitude-dependent way. That is, it is up to the agents, and not to the external designer, to support such facts (Section 5). We will compare our proposal with related logical works on the issues of collective belief and institutions (Section 6). In conclusion we will identify directions for future work on the basis of our framework (Section 7). Anchoring institutions, and their facts, in agents' minds is just the first step towards a more complete characterization of the "internal aspect" of normative systems and towards the vision of autonomous MASs.

2. ACCEPTANCE QUA GROUP MEMBER

Although in this paper the notion of acceptance *qua* group member is a primitive (*i.e.* it is not analyzed in more specific mental attitudes), some conceptual clarification is needed because of the crucial role it plays in explaining the maintenance of social institutions. Whereas beliefs have been studied for decades [16] as representative of doxastic mental states, acceptances have only been examined since [27] and [6] while studying the nature of argument premises or reformulating Moore's paradox [6]. If a belief that p is an attitude constitutively aimed at the truth of p, an acceptance is the output of "a decision to treat p as true in one's utterances and actions" [15] without being necessarily connected to the actual truth of the proposition. In order to better specify this distinction, it has been suggested [15] that while beliefs are not subject to the agent's will, acceptances are voluntary; while beliefs aim at truth, acceptance are sensitive to pragmatic considerations; while beliefs are shaped by evidence, acceptances need not be; while beliefs come in degrees, acceptances are qualitative; finally, while beliefs are context-independent, acceptance depends on context.

For the aims of this paper we are particularly interested in the last feature, namely the fact that acceptances can be context-dependent. In fact, one can decide (say for prudential reasons) to reason and act by "accepting" the truth of a proposition in a specific context, and possibly rejecting the very same proposition in a different one. Although, usually, this aspect of the acceptance state is studied in private contexts (*i.e.* when an agent, in order not to take too many risks, accepts that the total cost of her house restructuring will be beyond her reasonable expectations; see [4]), we will explore the role of this attitude in institutional contexts. Institutional contexts are rule-governed social practices on the background of which the agents reason. For example, take the case of a game like Clue. The institutional context is the rule-governed social practice which the agents conform to in order to be competent players.

On the background of such contexts, we are interested in the *explicit* mental states (the acceptances) that can be formally captured. In the context of Clue, for instance, an agent accepts that something has happened (see Example 3) *qua* player of Clue. The state of acceptance *qua* group member in an institutional context is the kind of acceptance one is committed to when one is "functioning as a group member" [29]. Although space restrictions prevent a full analysis of this notion, it is important to stress that we consider this attitude as one that is held by an agent. Nevertheless, there are specific consequences deriving from the agent's functioning as a group member: *e.g.* the acceptance of a proposition *qua* group member is always a public fact (see Section 4.1).

[2] It is in fact specific to legal institutions to have specialized agents empowered to change the institution itself on behalf of everybody else (see Section 5.1).

3. THE LOGIC OF ACCEPTANCE

3.1 Syntax

The syntactic primitives of our logic $\mathcal{AL}$ (*Acceptance Logic*) are the following: – a finite set of $n > 0$ agents $AGT = \{1, 2, ..., n\}$; – a nonempty finite set of *atomic actions* $ACT = \{a, b, ...\}$; – a set of atomic formulas $ATM = \{p, q, ...\}$; – a finite set of labels denoting institutional contexts $INST = \{inst_1, inst_2, ..., inst_m\}$; – a symbol λ denoting the private context. For notational convenience we note $2^{AGT\star} = 2^{AGT} \setminus \{\emptyset\}$ the set of all non empty subsets of agents, $\Delta_1 = \{C{:}x | C \in 2^{AGT\star}, x \in INST\}$ the set of all couples of non empty subsets of agents and institutional contexts, $\Delta_2 = \{i{:}\lambda | i \in AGT\}$ the set of all couples of single agents and private context, and $i{:}x$ for $\{i\}:x$. Finally, $\Delta = \Delta_1 \cup \Delta_2$.

The language $\mathcal{L}_{\mathcal{AL}}$ is defined as the smallest superset of ATM such that: if $\varphi, \psi \in \mathcal{L}_{\mathcal{AL}}$, $i \in AGT$ and $C{:}x \in \Delta$ then $\neg\varphi$, $\varphi \vee \psi$ and $[C{:}x]\,\varphi \in \mathcal{L}_{\mathcal{AL}}$. The classical boolean connectives $\wedge$, $\rightarrow$, $\leftrightarrow$, $\top$ (tautology) and $\perp$ (contradiction) are defined from $\vee$ and $\neg$ in the usual manner.

Formula $[C{:}x]\,\varphi$ has to be read "the agents in C accept that φ while functioning as group members in the institutional context x".

EXAMPLE 1. $[C{:}Greenpeace]\,protectEarth$ *is read "the agents in* C *accept that the mission of Greenpeace is to protect the Earth while functioning as activists in the context of Greenpeace" and* $[i{:}Catholic]\,PopeInfallibility$ *is read "the agent* i *accepts that the Pope is infallible while functioning as a Catholic in the context of the Catholic Church".*

For $C{:}x \in \Delta_1$: $[C{:}x]\perp$ has to be read "agents in C are not functioning as group members in the institutional context x" because we assume that functioning as a group member is, at least in this minimal sense, a rational activity; conversely, $\neg[C{:}x]\perp$ has to be read "agents in C are functioning as group members in the institutional context x"; $\neg[C{:}x]\perp \wedge [C{:}x]\,\varphi$ stands for "agents in C are functioning as group members in the context x and they accept that φ while functioning as group members" or simply "agents in C accept that φ *qua* group members in the institutional context x" which, for us, is tantamount to "The group C accepts that φ in the institutional context x" (*i.e.* group acceptance). Similarly, the formula $\neg[C{:}x]\,\varphi$ has to be read "agents in C are functioning as group members in the institutional context x and they do not accept that φ while functioning as group members in x" or simply "agents in C do not accept that φ *qua* group members in x" (*i.e.* "The group C does not accept that φ in the institutional context x).

EXAMPLE 2. $\neg[\{i,j\}{:}Europe]\perp \wedge [\{i,j\}{:}Europe]\,EuroMeansOfExchange$ *stands for "*i *and* j *accept* qua *Europeans that the Euro is the official means of exchange in the context of* $Europe$*", whereas* $\neg[\{i,j\}{:}Europe]\,DollarMeansOfExchange$ *stands for "*i *and* j qua *Europeans do not accept that dollar is the official means of exchange".*

Modal operators of the form $[i{:}\lambda]$ correspond to standard doxastic operators.[3] Hence a formula $[i{:}\lambda]\,\varphi$ has to be read "agent i believes that φ".

3.2 Semantics

We use a standard possible worlds semantics and a model is a triple $\mathcal{M} = \langle W, \mathscr{A}, \mathscr{V}\rangle$ where:

- W is a set of possible worlds;
- $\mathscr{A} : \Delta \longrightarrow (W \longrightarrow 2^W)$ associates each $C{:}x \in \Delta$ and possible world w with the set $\mathscr{A}_{C:x}(w)$ of possible worlds accepted by the group C in w, where agents in C are functioning as group members in the institutional context x;
- $\mathscr{V} : W \longrightarrow 2^{ATM}$ is a truth assignment which associates each world w with the set $\mathscr{V}(w)$ of atomic propositions true in w.

The rules defining the truth conditions of formulas of our logic are inductively defined as follows.

- $\mathcal{M}, w \models p$ iff $p \in \mathscr{V}(w)$;
- $\mathcal{M}, w \models \neg\varphi$ iff not $\mathcal{M}, w \models \varphi$;
- $\mathcal{M}, w \models \varphi \vee \psi$ iff $\mathcal{M}, w \models \varphi$ or $\mathcal{M}, w \models \psi$;
- $\mathcal{M}, w \models [C{:}x]\,\varphi$ iff for all $w' \in W$, if $w' \in \mathscr{A}_{C:x}(w)$ then $\mathcal{M}, w' \models \varphi$.

3.3 Axiomatization

The axiom system of $\mathcal{L}_{\mathcal{AL}}$ is made of all tautologies of propositional calculus and, the axioms and rules of inference of the basic normal modal logic for every operator $[C{:}x]$ where $C{:}x \in \Delta$. That is, we have all K-theorems for every $C{:}x \in \Delta$.

Moreover, we suppose that given a set of agents C, all $B \subseteq C$ have access to all the facts that are accepted (or that are not accepted) by agents in C while functioning as group members in the institutional context x. In particular, we suppose the following relations between the acceptances of the group members with respect to the institutional contexts: if agents in C (do not) accept that φ while functioning as group members in the institutional context x then for every subset B of C and institutional context y while functioning as group members in the institutional context y, agents in B accept that agents in C (do not) accept that φ while functioning as group members in the institutional context x. Furthermore, we suppose the following relations between the acceptance *qua* group member and individual beliefs: if agents in C (do not) accept that φ while functioning as group members in the institutional context x then, for every agent i in C, we have that i believes that agents in C (do not) accept that φ while functioning as group members in the institutional context x. Finally we suppose standard properties of introspection for beliefs: if agent i believes that φ then he believes that he believes that φ; if agent i does not believe that φ then he believes that he does not believe that φ. Such properties are captured by the following two axiom schemas. For every $C{:}x$, $B{:}y$ $\in \Delta$, if $B \subseteq C$ then:

$\mathbf{4}_{[C:x],[B:y]}$ $\quad [C{:}x]\,\varphi \rightarrow [B{:}y]\,[C{:}x]\,\varphi$

$\mathbf{5}_{[C:x],[B:y]}$ $\quad \neg\,[C{:}x]\,\varphi \rightarrow [B{:}y]\,\neg\,[C{:}x]\,\varphi$

[3]For the sake of compactness we prefer to adopt this non-standard notation for doxastic operators.

Axioms $\mathbf{4}_{[C:x],[B:y]}$ and $\mathbf{5}_{[C:x],[B:y]}$ together correspond to the following semantic property of Kripke models. For every $w \in W$ and $C{:}x$, $B{:}y \in \Delta$, if $B \subseteq C$ then:

S1 $\quad$ if $w' \in \mathscr{A}_{B:y}(w)$ then $\mathscr{A}_{C:x}(w') = \mathscr{A}_{C:x}(w)$

We also suppose that if agents in C accept that φ *qua* group members in the institutional context x then, for every subset B of C, it holds that agents in B accept φ *qua* group members in the institutional context x. This means that things accepted by the agents in a set C (*qua* group members) with respect to a certain institutional context x are also accepted by agents in all C's subsets with respect to the same context x. Formally, for every $C{:}x$, $B{:}x \in \Delta$, if $B \subseteq C$ then:

$\mathbf{Inc}_{[C:x],[B:x]}$ $\quad \neg\,[C{:}x]\,\bot \wedge [C{:}x]\,\varphi \rightarrow \neg\,[B{:}x]\,\bot \wedge [B{:}x]\,\varphi$

EXAMPLE 3. *Imagine three agents i, j, k that,* qua *players accept, in the context of Clue, that someone called Mrs. Red, has been killed:* $\neg\,[\{i,j,k\}{:}Clue]\,\bot \wedge [\{i,j,k\}{:}Clue]\,killedMrsRed$*. This implies that also the two agents i, j* qua *Clue players accept that someone called Mrs. Red has been killed in that context:*
$\neg\,[\{i,j\}{:}Clue]\,\bot \wedge [\{i,j\}{:}Clue]\,killedMrsRed$.

Axiom $\mathbf{Inc}_{[C:x],[B:x]}$ has the following semantic characterization. For every $w \in W$, $C{:}x$, $B{:}x \in \Delta$, if $B \subseteq C$ then:

S2 $\quad$ if $\mathscr{A}_{C:x}(w) \neq \emptyset$ then $\mathscr{A}_{B:x}(w) \neq \emptyset$ and $\mathscr{A}_{B:x}(w) \subseteq \mathscr{A}_{C:x}(w)$

As far as operators of type $[i{:}\lambda]$ for beliefs are concerned, we suppose that an agent cannot believe contradictions. Formally, for every $i{:}\lambda \in \Delta_2$:

$\mathbf{D}_{[i:\lambda]}$ $\quad \neg([i{:}\lambda]\,\varphi \wedge [i{:}\lambda]\,\neg\varphi)$

which corresponds to the following standard property of seriality. For every $w \in W$ and $i{:}\lambda \in \Delta_2$ we have:

S3 $\quad \mathscr{A}_{i:\lambda}(w) \neq \emptyset$

Thus, every doxastic operator $[i{:}\lambda]$ is $KD45$. (Indeed, besides satisfying Axiom D, it also satisfies Axioms 4 and 5 as particular instances of Axioms $\mathbf{4}_{[C:x],[B:y]}$ and $\mathbf{5}_{[C:x],[B:y]}$ where $C = B = \{i\}$ and $x = y = \lambda$.)

We call $\mathcal{AL}$ (Acceptance Logic) the logic axiomatized by the four principles $\mathbf{4}_{[C:x],[B:y]}$, $\mathbf{5}_{[C:x],[B:y]}$, $\mathbf{Inc}_{[C:x],[B:x]}$, $\mathbf{D}_{[i:\lambda]}$ and we write $\vdash_{\mathcal{AL}} \varphi$ iff formula φ is a theorem of $\mathcal{AL}$. Moreover, let $\mathcal{M}$ be a model such that $\mathcal{M} = \langle W, \mathscr{A}, \mathscr{V}\rangle$ as defined in Section 3.2 and satisfying the semantic constraints **S1**–**S3** given above. We write $\models_{\mathcal{AL}} \varphi$ iff formula φ is *valid* in all $\mathcal{AL}$ models, *i.e.* $\mathcal{M}, w \models \varphi$ for every $\mathcal{AL}$ model $\mathcal{M}$ and world w in $\mathcal{M}$. Finally, we say that a formula φ is *satisfiable* if there exists an $\mathcal{AL}$ model $\mathcal{M}$ and a world w in $\mathcal{M}$ such that $\mathcal{M}, w \models \varphi$.

4. GROUP ACCEPTANCE PROPERTIES

4.1 The public nature of group acceptance

In Section 3.1, we have analyzed the notion of group acceptance as the set of the acceptances of all the agents in the group while functioning as group members. This notion of acceptance *qua* group member however must not be confused with (nor reduced to) that of a private mental attitude. On the contrary we claim that group acceptances are always public so much that it is part of the concept of functioning as a group member that all the agents commonly believe that one is functioning in this way. In MAS literature, an operator to express common belief is given (see for instance [10]). The notion of common belief can be built on the concept of individual belief and on a particular kind of distributed belief of the form "every agent in C believes that φ". The former concept is expressed in our logic by operators of type $[i{:}\lambda]$. The latter concept is formally expressed by operators of type E_C where a formula $E_C\varphi$ is defined as follows:

$$E_C\varphi \stackrel{def}{=} \bigwedge_{i\in C} [i{:}\lambda]\,\varphi$$

Given a set of agents $C \subseteq AGT$, formula $CB_C\varphi$ is meant to stand for "there is common belief in C that φ", that is, "everyone in C believes that φ, everyone in C believes that everyone in C believes that φ, everyone in C believes that everyone in C believes that everyone in C believes that φ, and so on". If $E^1_C\varphi$ denotes $E_C\varphi$ and $E^k_C\varphi$ denotes $E_C(E^{k-1}_C\varphi)$, we can define $CB_C\varphi$ as follows:

$$CB_C\varphi \stackrel{def}{=} \bigwedge_{k>0} E^k_C\varphi$$

With the aim of making the public nature of group acceptance explicit, the following theorem highlights the relationship between our notion of group acceptance (*i.e.* acceptance by each of the agents *qua* group members) expressed by operator of type $[C{:}x]$ and the concept of common belief.

THEOREM. *For any $C{:}x \in \Delta$:*

(1) $$\vdash_{\mathcal{AL}} \quad [C{:}x]\,\varphi \leftrightarrow CB_C\,[C{:}x]\,\varphi$$

PROOF. Direction $\rightarrow$ can be proved from proving that $\forall k > 0$, $[C{:}x]\,\varphi \rightarrow E^k_C\,[C{:}x]\,\varphi$ by induction on k:

- $[C{:}x]\,\varphi \rightarrow E_C\,[C{:}x]\,\varphi$ (case $k = 1$)
- From $[C{:}x]\,\varphi \rightarrow E^k_C\,[C{:}x]\,\varphi$ infer $[C{:}x]\,\varphi \rightarrow E^{k+1}_C\,[C{:}x]\,\varphi$ (inductive case)

To prove the case $k = 1$, we just apply Axiom $\mathbf{4}_{[C:x],[B:y]}$ with $B{:}y = i{:}\lambda$ for each $i \in C$, which implies that $[C{:}x] \rightarrow \bigwedge_{i\in C}[i{:}\lambda]\,[C{:}x]\,\varphi$. The latter is the case $k = 1$ by definition of E_C.

Let us prove the inductive case. We suppose that $[C{:}x]\,\varphi \rightarrow E^k_C\,[C{:}x]\,\varphi$. By rule of necessitation on every $[i{:}\lambda]$, we infer $\bigwedge_{i\in C}[i{:}\lambda]\,([C{:}x]\,\varphi \rightarrow E^k_C\,[C{:}x]\,\varphi)$ which is (by definition of E_C) equivalent to: $E_C([C{:}x]\,\varphi \rightarrow E^k_C\,[C{:}x]\,\varphi)$. Thus from the latter, case $k = 1$ and definition of E^{k+1}_C we can deduce that $[C{:}x]\,\varphi \rightarrow E^{k+1}_C\,[C{:}x]\,\varphi$. This is enough to prove that $[C{:}x]\,\varphi \rightarrow E^k_C\,[C{:}x]\,\varphi$ (for $k > 0$) is a theorem. We can thus infer that $\bigwedge_{k>0}([C{:}x]\,\varphi \rightarrow E^k_C\,[C{:}x]\,\varphi)$ holds. By standard modal principles, $\bigwedge_{k>0}([C{:}x]\,\varphi \rightarrow E^k_C\,[C{:}x]\,\varphi)$ implies $[C{:}x]\,\varphi \rightarrow \bigwedge_{k>0} E^k_C\,[C{:}x]\,\varphi$ which is equivalent to $[C{:}x]\,\varphi \rightarrow CB_C\,[C{:}x]\,\varphi$. We leave to the reader the proof of $\leftarrow$ direction of the theorem. □

According to Theorem 1, the agents in C accept that φ while functioning as group members in the institutional context x if and only if there is common belief in C that they accept that φ while functioning as group members in the institutional context x. Hence, accepting a proposition while functioning as a group member is always a *public* fact which is out in the open and that is used by all the members to reason about each other in an institutional context.

4.2 Group acceptance and individual beliefs

As far as the relationship between acceptances *qua* group members and individual beliefs is concerned, it has to be noted that $\neg[C{:}x]\bot \wedge [C{:}x]\,\varphi \wedge \bigwedge_{i\in C}[i{:}\lambda]\neg\varphi$ where $C{:}x \in \Delta_1$ is satisfiable in our logic. This means that the attitudes privately endorsed by the agents and those entertained *qua* group members can diverge: one can privately disbelieve that which is accepted while functioning as a group member.

EXAMPLE 4. *Consider the discursive dilemma as elaborated in [25] in which a three-member court has to make a judgment on whether a defendant is liable for a breach of contract. If one assumes that the group accepts the majority rule to decide on the issue, it might happen that each judge can privately believe that the group ought to accept a certain conclusion (*e.g. *that the defendant is liable), while each is forced to accept the opposite* qua *group member (*i.e. qua *judge).*

5. ATTITUDE-DEPENDENT FACTS

Normative and institutional facts are a class of facts that are typical of institutional contexts [26]. Such facts have the peculiar feature of being dependent on the agents' attitudes in a way that we are now in the position to specify in detail. More precisely it has been noted that these facts are characterized at least by two features [19, 26, 28].

- Performativity: an attitude of certain type shared by a group of agents towards a normative or an institutional fact may contribute to the truth of a sentence describing the fact.
- Reflexivity: if a sentence describing a normative or an institutional fact is true, the relevant attitude is present.

EXAMPLE 5. *If the agents* qua *group members accept that a certain piece of paper as money (an institutional fact), then, in the appropriate context, this piece of paper is money for that group (performativity). At the same time, if it is true that a certain piece of paper is money for a group, then the agents* qua *group members accept the piece of paper as money (reflexivity).*

In order to represent in $\mathcal{AL}$ these kind of facts, we need first to define the concept of truth with respect to an institutional context in way that respects these two principles.

5.1 Truth in an institutional context

We formalize the notion of truth w.r.t. to a certain institutional context with the operator $[x]$. A formula $[x]\,\varphi$ is read "within the institutional context x, it is the case that φ". Here we suppose that "within the institutional context x it is the case that φ" if and only if "for every set of agents C, the agents in C accept that φ while functioning as group members in the institutional context x". Formally, for $x \in INST$:

$$[x]\,\varphi \stackrel{def}{=} \bigwedge_{C\in 2^{AGT\star}} [C{:}x]\,\varphi$$

It is straightforward to prove that $[x]$ are normal modal operators. Given the previous analysis, a fact is true w.r.t. an institutional context if and only if such fact is accepted by all the agents while they function as group members (hence the performativity and the reflexivity principles are maintained). Moreover, following Theorem 1, this group acceptance is the object of a common belief.

At this point, it might be objected that there are facts which are true in an institutional context but only "special" group members in the institution are aware of them. For instance, there are laws in every country which known only by the specialists of the domain (lawyers, judges, members of the parliament, *etc.*). Aren't these facts true notwithstanding that many group members are not aware of them?

In order to resist to this objection recall that, at this stage, our model applies to the basic informal institutions of a society. Relative to this restriction, the proposed assumption is justified because, w.r.t. these institutions, there is no other special institutional contexts in which the agents have the power to create and eliminate institutional facts characterizing the institution itself (*i.e.* nobody has the power to change the rules for promising). It is in fact peculiar of legal (formal) institutions to create such a specialized *meta*-context in which the agents have special powers to interpret and modify the institution itself. Given the aims of this paper, we leave this special case for future work.

Finally, the following abbreviation is defined:

$$[Univ]\,\varphi \stackrel{def}{=} \bigwedge_{x \in INST} [x]\,\varphi$$

which stands for "φ is universally accepted as true".

5.2 Contextual conditionals

From the concept of truth with respect to an institutional context a notion of *contextual conditional* can be defined. A contextual conditional is a material implication of the form $\varphi \rightarrow \psi$ in the scope of an operator $[x]$. A contextual conditional is a local one, that is, a conditional that is not universally valid while it is accepted by the group members in a specific institutional context. More precisely, we exclude the situation in which $[Univ]\,(\varphi \rightarrow \psi)$ is true.

EXAMPLE 6. *Let consider the institutional context of gestural language. There exists a contextual conditional in this language according to which, the nodding gesture "counts as" an endorsement of what the speaker is suggesting. This conditional is formally expressed by the construction* $[gesture]\,(nodding \rightarrow yes)$. *It is clear that this kind of conditional is not universally valid (*e.g. *in a different cultural context the same gesture may express exactly the opposite fact). Thus,* $\neg\,[Univ]\,(nodding \rightarrow yes)$ *holds.*

More generally, for every $x \in INST$ we define the following abbreviation:

$$\varphi \stackrel{x}{\rhd} \psi \stackrel{def}{=} [x]\,(\varphi \rightarrow \psi) \wedge \neg\,[Univ]\,(\varphi \rightarrow \psi)$$

$\varphi \stackrel{x}{\rhd} \psi$ stands for "in the institutional context x, if φ then ψ". Although space restrictions prevent from presenting and discussing all relevant properties of our construction $\varphi \stackrel{x}{\rhd} \psi$, it is interesting to note that $\varphi \stackrel{x}{\rhd} \psi$ satisfies some intuitive properties of count-as conditionals as isolated in [17].

THEOREM. *For every* $x \in INST$*:*

(2) *From* $\vdash_{\mathcal{AL}} (\varphi_2 \leftrightarrow \varphi_3)$ *infer* $\vdash_{\mathcal{AL}} (\varphi_1 \stackrel{x}{\rhd} \varphi_2 \leftrightarrow \varphi_1 \stackrel{x}{\rhd} \varphi_3)$

(3) *From* $\vdash_{\mathcal{AL}} (\varphi_1 \leftrightarrow \varphi_3)$ *infer* $\vdash_{\mathcal{AL}} (\varphi_1 \stackrel{x}{\rhd} \varphi_2 \leftrightarrow \varphi_3 \stackrel{x}{\rhd} \varphi_2)$

(4) $\vdash_{\mathcal{AL}} (\varphi_1 \stackrel{x}{\rhd} \varphi_2 \wedge \varphi_1 \stackrel{x}{\rhd} \varphi_3) \rightarrow (\varphi_1 \stackrel{x}{\rhd} (\varphi_2 \wedge \varphi_3))$

(5) $\vdash_{\mathcal{AL}} (\varphi_1 \stackrel{x}{\rhd} \varphi_2 \wedge \varphi_3 \stackrel{x}{\rhd} \varphi_2) \rightarrow ((\varphi_1 \vee \varphi_3) \stackrel{x}{\rhd} \varphi_2)$

(6) $\vdash_{\mathcal{AL}} (\varphi_1 \stackrel{x}{\rhd} \varphi_2 \wedge (\varphi_1 \wedge \varphi_2) \stackrel{x}{\rhd} \varphi_3) \rightarrow (\varphi_1 \stackrel{x}{\rhd} \varphi_3)$

PROOF. We only provide a proof of Theorem 6 as an example. This theorem expresses a property of cumulative transitivity (cut). The other theorems and rules of inference can be proved straightforwardly by definition of $\varphi \stackrel{x}{\rhd} \psi$) and the axioms and rules of inference of $\mathcal{AL}$. $\varphi_1 \stackrel{x}{\rhd} \varphi_2 \wedge (\varphi_1 \wedge \varphi_2) \stackrel{x}{\rhd} \varphi_3$ implies $[x]\,(\varphi_1 \rightarrow \varphi_2)$ and $[x]\,((\varphi_1 \wedge \varphi_2) \rightarrow \varphi_3)$ which in turn imply $[x]\,(\varphi_1 \rightarrow \varphi_3)$ (by the fact that $[x]$ is normal. Moreover, $\varphi_1 \stackrel{x}{\rhd} \varphi_2 \wedge (\varphi_1 \wedge \varphi_2) \stackrel{x}{\rhd} \varphi_3$ implies $\neg\,[Univ]\,((\varphi_1 \wedge \varphi_2) \rightarrow \varphi_3)$ which is equivalent to $\neg\,[Univ]\,(\neg\varphi_1 \vee \neg\varphi_2 \vee \varphi_3)$. It is straightforward to prove that operator $[Univ]$ is also a normal modal operator (space restrictions prevent from giving the proof here). Therefore, $\neg\,[Univ]\,(\neg\varphi_1 \vee \neg\varphi_2 \vee \varphi_3)$ implies $\neg\,[Univ]\,(\neg\varphi_1 \vee \varphi_3)$ (by the fact that $[Univ]$ is normal) which in turn is equivalent to $\neg\,[Univ]\,(\varphi_1 \rightarrow \varphi_3)$.

Moreover, we can easily show that our concept of contextual conditional does not satisfy reflexivity, transitivity and weakening of the antecedent, that is, the following three formulas are not valid: $\varphi \stackrel{x}{\rhd} \varphi$, $(\varphi_1 \stackrel{x}{\rhd} \varphi_2 \wedge \varphi_2 \stackrel{x}{\rhd} \varphi_3) \rightarrow \varphi_1 \stackrel{x}{\rhd} \varphi_3$, and $\varphi_1 \stackrel{x}{\rhd} \varphi_2 \rightarrow (\varphi_1 \wedge \varphi_3) \stackrel{x}{\rhd} \varphi_2$. As discussed in section 6.2 our notion of contextual conditional is similar to the notion of *proper classificatory rule* given in [13]. [4]

5.3 Normative facts

While contextual conditionals are useful to understand the notion of institutional facts, they are not sufficient for a more precise characterization. In fact, as noted in [26], institutional facts are always connected to a deontic dimension that up to now is still missing.

In our perspective, a contextual conditional $\varphi \stackrel{x}{\rhd} \psi$ can be adopted to represent an institutional fact if and only if the term ψ in the contextual conditional is a fact to which a certain number of obligations and permissions are associated within the institutional context x. In this sense, ψ is an institutional fact with respect to the institutional context x.

EXAMPLE 7.*"Being eighteen years old counts as being of age" is a constitutive rule accepted by a set of agents* qua *citizens in Italy and "being of age" is an institutional fact with respect to this context. Moreover, to such an institutional fact a certain number of permissions and obligations are associated (*e.g. *in Italy if you are of age you have the permission to vote and the obligation to fulfill the military duties). In this sense the constitutive rule "being eighteen years old counts as being of age" connects the institutional fact "being of age" with the brute fact "being eighteen years old" which is a fact intrinsically connected to certain normative facts.*

In order to capture this core feature, our logic $\mathcal{AL}$ can be appropriately extended by introducing a *violation* atom V as in Anderson's reduction of deontic logic to alethic logic [2] and in dynamic deontic logic [22]. By means of this new formal construct we can specify normative facts (*i.e.* what it is obligatory and permitted) in a way that respect their being also a kind of attitude-depend fact holding relative to certain attitudes and in a specific institutional context. As far as obligations are concerned, we say that "φ is something obligatory within the institutional context x" (noted $O(\varphi, x)$) if and only if "$\neg\varphi \rightarrow V$ is a contextual conditional in the institutional context x" or, more specifically, "$\neg\varphi$ counts as a violation within the institutional context x". Formally:

$$O(\varphi, x) \stackrel{def}{=} \neg\varphi \stackrel{x}{\rhd} V.$$

As far as permission are concerned we say that "φ is something permitted within the institutional context x" (noted $P(\varphi, x)$) if and only if $\neg\varphi$ is not obligatory within the institutional context x. Formally:

$$P(\varphi, x) \stackrel{def}{=} \neg O(\neg\varphi, x).$$

[4] We refer to [13] for interesting arguments concerning why proper classificatory rules should not necessarily satisfy reflexivity, transitivity and weakening of the antecedent.

Formulas of type $O(\varphi, x)$ and $P(\varphi, x)$ can be conceived as particular instances of so-called *regulative rules*, that is, rules which specify the ideal behavior of agents in terms of permissions, obligations, and prohibitions. We refer to these rules as normative facts.[5]

EXAMPLE 8. *The formula $O(driveCar \rightarrow \neg RightSide, UK)$ is a normative fact in the UK within whose context it is obligatory to drive on the left side of the street (*i.e. *"driving a car on the right side of the street counts as violation in UK").*

Again, it is important to stress the fact that normative facts, by being represented with a contextual conditional are attitude-dependent facts and are intrinsically connected with the acceptance of all the agents *qua* group members in a specific institutional context.

5.4 Institutional facts and constitutive rules

We are now in the position to formalize what an institutional fact is. Let $2^{\mathcal{L}_{\mathcal{AL}}\star} = 2^{\mathcal{L}_{\mathcal{AL}}} \setminus \{\emptyset\}$ the set of non empty subsets of $\mathcal{L}_{\mathcal{AL}}$. From the previous construction $\varphi \overset{x}{\triangleright} \psi$ it is straightforward to come up with a formal characterization of the concept of such a fact. Formally, for every $x \in INST$ and $\Sigma_O, \Sigma_P \in 2^{\mathcal{L}_{\mathcal{AL}}\star}$:

$$InstFact_x^{\Sigma_O,\Sigma_P}(\varphi) \overset{def}{=} \bigwedge_{\sigma\in\Sigma_O} O(\varphi \rightarrow \sigma, x) \wedge \bigwedge_{\sigma'\in\Sigma_P} P(\varphi \wedge \sigma', x)$$

$InstFact_x^{\Sigma_O,\Sigma_P}(\varphi)$ stands for "φ is an institutional fact within the institutional context x characterized by the set of obligations Σ_O and the set of permissions Σ_P".

EXAMPLE 9. *The formula $InstFact_{Italy}^{\{military\},\{vote\}}(toBeOfAge)$ stands for "being of age is an institutional fact in the context of Italy and is characterized by the permission to vote in the political elections and the obligation to fulfill the military duties".*[6]

From the concept of institutional fact we can also formalize the concept of constitutive rule. To this aim, we must make explicit the fact that the term ψ in $\varphi \overset{x}{\triangleright} \psi$ is an institutional fact to which a set of obligations and a set of permissions are associated. Formally, for every $x \in INST$ and $\Sigma_O, \Sigma_P \in 2^{\mathcal{L}_{\mathcal{AL}}\star}$:

$$ConstRule_x^{\Sigma_O,\Sigma_P}(\varphi, \psi) \overset{def}{=} \varphi \overset{x}{\triangleright} \psi \wedge InstFact_x^{\Sigma_O,\Sigma_P}(\psi)$$

$ConstRule_x^{\Sigma_O,\Sigma_P}(\varphi, \psi)$ stands for "φ counts as ψ is a constitutive rule of institution x where ψ is an institutional fact within the institutional context x characterized by the set of obligations Σ_O and the set of permissions Σ_P".

EXAMPLE 10. *The formula $ConstRule_{Italy}^{\{military\},\{vote\}}(eighteen, toBeOfAge)$ stands for "being eighteen years old counts as being of age is a constitutive rule in the context of Italy and being of age is an institutional fact characterized by the permission to vote in the political elections and the obligation to fulfill the military duties". In this sense $ConstRule_{Italy}^{\{military\},\{vote\}}(eighteen, toBeOfAge)$ is a specific kind of contextual conditional in which the connection between the institutional fact $toBeOfAge$ and the brute fact $eighteen$ is established. A number of normative facts consisting in obligations and permissions pertain to the institutional fact $toBeOfAge$, namely $O(toBeOfAge \rightarrow military, Italy)$ and $P(toBeOfAge \wedge vote, Italy)$.*

[5]The distinction between *regulative rule* and *constitutive rule* has been emphasized by Searle [26] and then modelled in logic by several authors. For an example see [3].

[6]A more precise formulation of this example needs a representation of the right relation which is, however, beyond the scope of this article. See [20] for more details.

6. RELATED WORKS

6.1 Link between $\mathcal{AL}$ and the G logic

A logic of what is publicly grounded in a group has been introduced in [11]: $G_C\varphi$ means that "it is publicly grounded for group C that φ is true". When C is reduced to a singleton $\{i\}$, $G_{\{i\}}$ is identified with the belief *à la* Hintikka [16]. We can show that it can be viewed as an operator of group belief (in the Gilbert's sense [12] [7]). In this view, group belief is rational (**D**$_{G_C}$), public for every subgroup (**SR**$_+$ and **SR**$_-$) and it has been formed by the joint acceptance of all members (**WR** and **CG**). Its axiomatics is thus the following one:

(**D**$_{G_C}$) $G_C\varphi \rightarrow \neg G_C\neg\varphi$
(**SR**$_+$) $G_C\varphi \rightarrow G_{C'}G_C\varphi, C' \subseteq C$
(**SR**$_-$) $\neg G_C\varphi \rightarrow G_{C'}\neg G_C\varphi, C' \subseteq C$
(**WR**) $G_C\varphi \rightarrow G_C G_{C'}\varphi, C' \subseteq C$ and φ objective.[8]
(**CG**) $(\bigwedge_{i\in C} G_C G_i\varphi) \rightarrow G_C\varphi$

Notions of group belief and group acceptance seem to be very close. Thus the idea of expressing the G operator in $\mathcal{AL}$ appears intuitive because $\mathcal{AL}$ is more expressive, with the notion of context lacking in the G logic. We show in the sequel that $\mathcal{AL}$ can subsume the G's logic.

Contrary to our framework, G operator does not take into account various institutional contexts, what expresses that it considers (implicitly) only one. Thus formally we have: $G_C\varphi \equiv [C{:}x_C]\,\varphi$, where x_C is the only institution whereby C is concerned and where $x_{\{i\}} \equiv \lambda$.

We need both to examine and compare axiomatics. Axioms $\mathbf{4}_{[C:x],[B:y]}$ and $\mathbf{5}_{[C:x],[B:y]}$ are generalizations of the (**SR**$_+$) and (**SR**$_-$) for contexts x_C and x_B instead of x and y. They represent the public nature of both notions. Axiom $\mathbf{Inc}_{[C:x],[B:x]}$ cannot be expressed in the grounding logic[9]. An axiom such as: $G_C\varphi \rightarrow G_B\varphi$, would be too strong because we consider that belief of a subgroup is not related to the uppergroup beliefs (and in particular group belief is totally independent of individual group members beliefs).

Some axioms lack in $\mathcal{AL}$ to represent the G operators. In particular the axiom $\mathbf{D}_{[i:\lambda]}$ should be generalized to $[C{:}x_C]$ representing that agents in C are *de facto* functioning as group members in the context x_C. Moreover axioms (**WR**) and (**CG**) express that a group belief is established by a consensus of expressed opinion. They do not have a counterpart in the $\mathcal{AL}$ logic, because we are only concerned here by properties of acceptance (not by its formation). (**WR**) and (**CG**) could be translated directly.

These three additional axioms are due to the features of the particular context x_C: they represent the strong link existing between x_C and C. We can note that theorem 1 is also a theorem of the grounding logic. In the sequel, we explore interactions between G_C defined as $[C{:}x_C]$ and general acceptance $[C{:}x]$, which produces mixed theorems.

[7]The proof is based on common features shared by both notions. They are public, commonly believed when established and formed by consensus (*i.e.* by acceptance of every member).

[8]An objective formula is a formula that is not equivalent to a formula under the scope of G_i operator, for each member i of C. This restriction is due to the fact that, by asserting proposition φ, an agent expresses that he believes φ and thus this belief is automatically grounded for the group thank to public actions hypotheses.

[9]Except under his tautological and uninformative form where $B = C$.

As $\neg [C{:}x] \bot \rightarrow \neg [B{:}x] \bot$ (with $B \subseteq C$) is a theorem of $\mathcal{AL}$ (the proof can be easily built from $\mathbf{Inc}_{[C:x],[B:x]}$), we have: $\neg [C{:}x_C] \bot \rightarrow \neg [B{:}x_C] \bot$, with $B \subseteq C$, which means that every agent in subsets of C are also functioning as group members in the context x_C.

Moreover, as $\neg [C{:}x_C] \bot$ is valid, Axiom $\mathbf{Inc}_{[C:x],[B:x]}$ is reduced to the following theorem: $[C{:}x_C]\, \varphi \rightarrow [B{:}x_C]\, \varphi$, with $B \subseteq C$. Thus if φ is a belief of the group C, every subgroup accept it in the context of x_C; there is a group acceptance on what is collectively believed. For example, if it is collectively believed by the activists that the aim of Greenpeace is to protect the Earth then in the context of Greenpeace every subgroup must accept it. This does not implies anything about subgroup and individual beliefs.

From previous theorem, we can also prove that: $[C{:}x_C]\, \varphi \rightarrow [C{:}x_C]\, [B{:}x_C]\, \varphi$, with $B \subseteq C$. This theorem extends the previous one: if φ is collectively believed, every subgroup accept it in the context x_C (by former theorem), but this acceptance is also collectively believed. This theorem is in fact quite close to axiom (**WR**) in the grounding logic.

6.2 Related works on normative systems

Because of interesting formal similarities and given the space restrictions, we will just compare $\mathcal{AL}$ with [13] in which a modal logic for the formalization of count-as assertions and the specification of normative systems has been proposed. This logic is based on a set of modal operators $[x]^*$ where the index x is in a set of indexes C_0.[10] An index x is supposed to denote a certain institutional context (or normative system). Operators $[x]^*$ are similar to our operators $[x]$ defined in Section 5.1. A formula $[x]^*\, \varphi$ approximately stands for "in the institutional context/normative system x it is the case that φ". An operator $[u]^*$ is also used for denoting facts which universally hold. The set $C = C_0 \cup \{u\}$ is given by adding index u to the set of indexes C_0. Differently from our logic where the contextual operator $[x]$ is built on the notion of group acceptance, in Grossi's logic the contextual operator $[x]^*$ is given as a primitive operator. Operators $[x]^*$ and $[u]^*$ are exploited in Grossi's logic to define contextual conditionals called *proper classificatory rules* noted by $\varphi \Rightarrow_i^{cl+} \psi$ which is an abbreviation of $[x]^*\, (\varphi \rightarrow \psi) \wedge \neg [u]^*\, (\varphi \rightarrow \psi)$ and is meant to stand for "φ counts as ψ in the normative system x". The construction $\varphi \Rightarrow_i^{cl+} \psi$ is similar to our $\varphi \overset{x}{\triangleright} \psi$.[11] Operator $[u]^*$ is $S5$ and the logic is supposed to satisfy the following additional principles. For any $x, y \in C$:

1. $[x]^*\, \varphi \rightarrow [y]^*\, [x]^*\, \varphi$
2. $\neg [x]^*\, \varphi \rightarrow [y]^*\, \neg [x]^*\, \varphi$
3. $[u]^*\, \varphi \rightarrow [x]^*\, \varphi$
4. $[u]^*\, \varphi \rightarrow \varphi$

According to the both principles 1. and 2., truth and falsehood in institutional contexts/normative systems are absolute because they remain invariant even if evaluated from another institutional context/normative system. This means that every normative system y has full access to all facts which are true in a different normative system x. These two principles are in our view criticizable because they rely on the very counter-intuitive assumption that all facts true in an institutional context are public to all other institutional contexts. But, what does it mean that a fact is known by an institution? Our aim here is to show that such an assumption can be disambiguated in our logical framework. The relevant question is: under what additional assumptions formulas $[x]\, \varphi \rightarrow [y]\, [x]\, \varphi$ and $\neg [x]\, \varphi \rightarrow [y]\, \neg [x]\, \varphi$ can be inferred in our logic? On the one hand, it is easy to prove that the principles given in Section 3.3 are not sufficient to infer such formulas. Indeed, formulas $[x]\, \varphi \wedge \neg [y]\, [x]\, \varphi$ and $\neg [x]\, \varphi \wedge \neg [y]\, \neg [x]\, \varphi$ are satisfiable in $\mathcal{AL}$. On the other hand, it is straightforward to show that: if Axioms $\mathbf{4}_{[C:x],[B:y]}$ and $\mathbf{5}_{[C:x],[B:y]}$ are weakened by supposing that they **also** hold for $B \nsubseteq C$, then formulas $[x]\, \varphi \rightarrow [y]\, [x]\, \varphi$ and $\neg [x]\, \varphi \rightarrow [y]\, \neg [x]\, \varphi$ can be inferred. This means that in our logic Grossi's properties can be derived under the assumption that, given two arbitrary sets of agents B and C, agents in B has access to all facts that agents in C accept (do not accept), while functioning as group members in a certain institutional context x. That is, given an arbitrary set of agents C, if agents in C accept that φ while functioning as a group members in the institutional context x then this fact is public in such a way that all other agents outside C accept that agents in C accept that φ while functioning as group members in the institutional context x.

Concerning the principle 4., it says that: if φ universally holds then φ is true. This principle is also criticizable in our opinion. For instance, during the 7th-6th century BC people believed that the earth was flat. But it has never been the case that earth was/is/will be flat.

More generally, if we suppose that: Axioms $\mathbf{4}_{[C:x],[B:y]}$ and $\mathbf{5}_{[C:x],[B:y]}$ studied in Section 3.3 are also valid for $B \nsubseteq C$; the T axiom is valid for $[Univ]$ operator (in a similar way of the previous principle 4.); and the following translations of Grossi's operators $[x]^*$ and $[u]^*$ into our logic $\mathcal{AL}$ are given

- $tr([x]^*\, \varphi) = [x]\, \varphi$
- $tr([u]^*\, \varphi) = [Univ]\, \varphi$,

we can prove that the translations into $\mathcal{AL}$ of all Grossi's axioms are $\mathcal{AL}$ theorems. This is shown by the following theorem.

THEOREM. *Suppose that: i) $[Univ]\, \varphi \rightarrow \varphi$ is valid, and that for every $C{:}x, B{:}y \in \Delta$, ii) $[C{:}x]\, \varphi \rightarrow [B{:}y]\, [C{:}x]\, \varphi$ and iii) $\neg [C{:}x]\, \varphi \rightarrow [B{:}y]\, \neg [C{:}x]\, \varphi$ are valid in $\mathcal{AL}$. Thus, the following properties can be inferred in $\mathcal{AL}$:*

- $[x]\, \varphi \rightarrow [y]\, [x]\, \varphi$
- $\neg [x]\, \varphi \rightarrow [y]\, \neg [x]\, \varphi$
- $[Univ]\, \varphi \rightarrow [x]\, \varphi$
- $[Univ]$ *satisfies all Axioms and rules of inference of the system* $S5$

PROOF. We only provide a proof of the last item of the theorem. The other items can be proved in a similar way. First of all $[Univ]$ is a normal modal operator by definition as a conjunction of normal modal operators $[x]$. We have property $\mathbf{T}_{[Univ]}$ by Hypothesis i). We only need to prove that $\mathbf{4}_{[Univ]}$ and $\mathbf{5}_{[Univ]}$ can be inferred from the hypotheses. From Hypothesis ii) we can deduce that $[C{:}x]\, \varphi \rightarrow \bigwedge_{B:y \in \Delta} [B{:}y]\, [C{:}x]\, \varphi$ which is equivalent (by definition of $[Univ]\, \varphi$) to $[C{:}x]\, \varphi \rightarrow [Univ]\, [C{:}x]\, \varphi$, which entails $\bigwedge_{C:x \in \Delta} [C{:}x]\, \varphi \rightarrow \bigwedge_{C:x \in \Delta} [Univ]\, [C{:}x]\, \varphi$, which is equivalent to $[Univ]\, \varphi \rightarrow [Univ]\, [Univ]\, \varphi$ (*i.e.* $\mathbf{4}_{[Univ]}$). $\mathbf{5}_{[Univ]}$ (*i.e.* $\neg [Univ]\, \varphi \rightarrow [Univ]\, \neg [Univ]\, \varphi$) can be inferred from Hypothesis iii) in a similar way.

[10] Here we use the notation $[x]^*$ in order to distinguish Grossi's operators from our operators $[x]$.

[11] The author distinguishes *proper classificatory rules* from mere *classificatory rules* and *constitutive rules*. Differently from *classificatory rules*, *proper classificatory rules* are rules which would not hold without the normative system/institution stating them. In [14] a further distinction between *classificatory rules* and *constitutive rules* is given.

7. CONCLUSION

Let's take stock. We have started the paper by raising the challenge of *autonomy* at the level of MASs, so that they will be able to bind themselves in ways that further the achievement of collective goods in dynamic and uncertain environments as human societies do.

As a first step to meet this challenge, we have proposed the $\mathcal{AL}$ logic in which the agents' attitudes *qua* group members can be analyzed. Given the properties of a demystified notion of group acceptance in an institutional context, we have provided an analysis of the kind of attitude-dependent facts typical of institutions. In particular, we have introduced a notion of obligation and permission with respect to an institutional context (*i.e.* so-called normative facts). Then, we have defined institutional facts. In our perspective an institutional fact within the institutional context x is a fact to which a number of obligations and permissions are (contextually) associated. Finally, we have formalized the concept of constitutive rule, that is, a rule which is responsible for the connection between an institutional fact and a brute physical fact. In our view, a constitutive rule is a rule of type "φ counts as ψ in the institutional context x" where ψ denotes an institutional fact within the institutional context x. While such rules are usually defined from the external perspective of a normative system or institution, we have, once again, anchored these rules in the agents' attitudes.

Although the present model is focused on the neglected layer of informal institutions, it still lacks sufficient expressiveness to represent the phenomenon of "institutionalized power" [17] which is, of course, crucial also within this kind of institutions. In order to cope with limitation, in future work, we will expand $\mathcal{AL}$ with Propositional Dynamic Logic (PDL) in order to be able to talk about actions within our language. Moreover, a first kind of dynamics will be studied in which agents, *qua* group members in specific institutional contexts, will be able to create new institutional facts. Given the way we have modeled such facts, the agents will update and revise their own deontic commitments accordingly.

This extension will further give the opportunity for a foundation of artificial legal institutions and for their connections with informal ones. In fact the "basic norm" [18], *i.e.* the basic informal institution that provides the validity of legal systems, will be represented on the model of the other informal institutions. Representing the "basic norm" is in fact the crucial step for making it possible for a MAS to create and maintain by itself a legal system that is acknowledged as valid from the agents themselves.

The long term project is then to provide a three-layered model in which legal institutions, social institutions, and the socio-cognitive relations between the agents dynamically interact in order to enable institutional change and adaptation.

8. ACKNOWLEDGMENTS

Emiliano Lorini is financed by the project ForTrust (Social Trust Analysis and Formalization) funded by the french Agence Nationale de la Recherche (ANR). Luca Tummolini is supported by the project SOCCOP (The Social and Mental Dynamics of Cooperation), funded by the ESF.

9. REFERENCES

[1] T. Ågotnes, W. van der Hoek, J. Rodriguez-Aguilar, C. Sierra, and M. Wooldridge. On the logic of normative systems. In M. Veloso, editor, *Proceedings of IJCAI2007*, California, 2007. AAAI Press.

[2] A. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 22:100–103, 1958.

[3] G. Boella and L. Van der Torre. Regulative and constitutive norms in normative multiagent systems. In D. Dubois, A. Christopher, A. Welty, and M. Williams, editors, *Proceedings of KR2004*, pages 255–266. AAAI Press, 2004.

[4] M. E. Bratman. Practical reasoning and acceptance in context. *Mind*, 101(401):1–15, 1992.

[5] C. Castelfranchi. Formalising the informal? Dynamic social order, bottom-up social control, and spontaneous normative relations. *Journal of Applied Logic*, 1(1-2):47–92, 2003.

[6] L. J. Cohen. *An essay on belief and acceptance*. Owford University Press, New York, USA, 1992.

[7] R. Conte and C. Castelfranchi. *Cognitive and social action*. London University College of London Press, London, 1995.

[8] R. Conte, C. Castelfranchi, and F. Dignum. Autonomous norm acceptance. In J. Müller, M. P. Singh, and A. S. Rao, editors, *Proceedings of ATAL-98*, volume 1555 of *LNAI*, pages 99–112. Springer-Verlag, 1999.

[9] V. Dignum and F. Dignum. Modelling agent societies: Coordination frameworks and institutions. In P. Brazdil and A. Jorge, editors, *LNAI 2258*, Berlin, 2001. Springer-Verlag.

[10] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, 1995.

[11] B. Gaudou, A. Herzig, and D. Longin. Grounding and the expression of belief . In *Proceedings of KR 2006*, pages 211–229. AAAI Press, 2006.

[12] M. Gilbert. *On Social Facts*. Routledge, 1989.

[13] D. Grossi. Classificatory aspects of counts-as: An analysis in modal logic. *Journal of Logic and Computation*, 16(5):613–643, 2006.

[14] D. Grossi, J.-J. C. Meyer, and F. Dignum. Counts-as: Classification or constitution? An answer using modal logic. In *Proceedings of DEON'06*, 2006.

[15] P. Hakli. Group beliefs and the distinction between belief and acceptance. *Cognitive Systems Research*, 7:286–297, 2006.

[16] J. Hintikka. *Knowledge and Belief*. Cornell Uni. Press, 1962.

[17] A. Jones and M. J. Sergot. A formal characterization institutionalised power. *J. of the IGPL*, 4:429–445, 1996.

[18] H. Kelsen. *Pure Theory of Law*. UC Berkeley press, 1967.

[19] E. Lagerspetz. *The opposite mirrors*. Kluwer, 1995.

[20] D. Makinson. On the formal representation of rights relations. *J. of Philosophical Logic*, 15(4):403–425, 1986.

[21] C. Mantzavinos, D. North, and S. Shariq. Learning, institutions, and economic performance. *Perspectives on Politics*, 2:75–84, 2004.

[22] J. J. Meyer. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29(1):109–136, 1988.

[23] P. Noriega and C. Sierra. Electronic institutions: Future trends and challenges. In *LNAI 2446*, 2002. Springer.

[24] D. North. *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press, Cambridge, 1990.

[25] P. Pettit. Deliberative democracy and the discursive dilemma. *Philosophical Issues*, 11:268–99, 2001.

[26] J. Searle. *The Construction of Social Reality*. The Free Press, New York, 1995.

[27] R. Stalnaker. *Inquiry*. MIT Press, Cambridge, 1984.

[28] R. Tuomela. *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge Univ. Press, 2002.

[29] R. Tuomela. *The Philosophy of Sociality*. Oxford University Press, Oxford, 2007.