



**HAL**  
open science

## Anchoring the institutional dimension of speech acts in agents' attitudes: a logical approach

Emiliano Lorini, Dominique Longin, Benoit Gaudou

### ► To cite this version:

Emiliano Lorini, Dominique Longin, Benoit Gaudou. Anchoring the institutional dimension of speech acts in agents' attitudes: a logical approach. IEEE International Conference on Research, Innovation and Vision for the Futur 2009, Jul 2009, Da Nang City, Vietnam. pp.65-72. hal-03672518

**HAL Id: hal-03672518**

**<https://hal.science/hal-03672518>**

Submitted on 19 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Anchoring the institutional dimension of speech acts in agents’ attitudes: a logical approach

Emiliano Lorini\*, Dominique Longin\* and Benoit Gaudou†‡

\* Université de Toulouse, CNRS, Institut de Recherche en Informatique de Toulouse (IRIT),  
118 Route de Narbonne, F-31062, Toulouse, France

Emails: lorini@irit.fr, longin@irit.fr

† UMI 209 UMMISCO, Institut de Recherche pour le développement (IRD), Bondy, F-93143, France

‡ MSI, Institut de la Francophonie pour l’Informatique (IFI), Ha Noi, Viet Nam

Email: benoit.gaudou@alumni.enseiht.fr

**Abstract**—The focus of this contribution is to define speech acts as institutional actions where modeled institutions are social or informal institutions. With social or informal institution, we mean an institution which is grounded on the acceptances of its members. In the first part of the paper we present a logic of acceptance, goal and action. Then, we specify how agents can create and maintain normative and institutional facts on the basis of their acceptances *qua* members of a certain institution. In particular we focus on obligations and social commitments and provide an original reductionist characterization of these concepts, anchoring them in agents’ attitudes. Finally, we propose a formal characterization of the speech act *promise*. In this last part, we first define the constitutive rule which creates the relation between an utterance (as a physical action) and the speech act *promise* (as an institutional action). Then, we specify the deontic dimension of promise. Finally, we establish the relationship between promise and social commitment.

**Index Terms**—Modal logic, institutions, acceptance, speech act, commitment, multi-agent systems.

## I. INTRODUCTION

Saussure [1] distinguishes language (*la langue*) from speech (*la parole*). The “meaning” of the former is defined by institutions (social aspect), by a set of conventions outside the linguistic activity (static aspect). The latter corresponds to the use of the language (dynamical aspect) by individuals, and it is the intention of these individuals that assigns a “value” to the act of utterance (individual aspect). Metaphorically, language is a sort of score while speech is the execution of this score by musicians.

The Austin’s concept of “performative utterance” [2] is the first step to bring together language and speech because: first, performative utterances are regulated by social conventions that determine not only the “meaning” of the utterance, but also the “value” of its enunciation<sup>1</sup>; second, it is practically impossible to attribute a meaning to a performative utterance independently of the value of its enunciation.

Through the concept of illocutionary act, speech act theory [2], [3] generalizes the concept of performative utterance. In such a theory, the value of an enunciation cannot be no longer considered as a consequence of a preliminary meaning; but

contrary to the linguistic nihilism of the wittgensteinian trend and their slogan “Meaning is use” (See [3, Section 6.4]), speech act theory does not deny that the use of language is founded on a prior knowledge: it claims that the meaning of an utterance integrates some shared rules that fix the effects of its enunciation on the speech situation.

“It is essential to see [that] (...) language is the fundamental social institution.” [4, Section 5]. Language has a constitutive role in every social institution, it is a necessary condition to the existence of such institutions. The difference between a crude fact and an institutional fact is that the latter has some status associated to some function, and that there exists a collective representation of this status and of this function [4] : the “language (...) provides the means of this representation”.

Language is thus the carrier associating a concept to its meanings into a particular institution. This mapping is made through constitutive rules by (in)formal institutions. Contrary to normative rules that govern existing behaviors, constitutive rules do not have a purely normative role but create or define new forms of behavior whose the existence depends of these rules (see [3, Section 2.5]). For instance, politeness rules govern relationships between individuals although these relationships exist independently of politeness. Thus, politeness is a set of normative rules.

Subsequently, to speak a language, it is to adopt a form of intentional behavior (by performing speech acts) governed by constitutive rules related to a particular institution. These rules create or define new forms of behavior such as promises, orders, requests, assertions, *etc.* (see [3, Section 2.5]). Such speech acts are institutional facts defined by a set of (constitutive) rules of the form “ $X$  count as  $Y$  in the situation  $S$ ”. (See [3, Section 2.7] for more details.) For example, in the context of *ordinary communication* [5], a promise is defined and created on the basis of a constitutive rule (shared by the speaker and the hearer) of the form: a speaker’s enunciation of the utterance: “I am going to perform action  $a$ !” *counts as* a promise to the hearer to perform action  $a$ , under the condition that the hearer wants the speaker to perform action  $a$ . Subsequently, to speak a language, it is to adopt a form of

<sup>1</sup>By “enunciation”, we mean “the performance of an utterance”.

intentional behavior (by performing a speech act) governed by such a kind of constitutive rules.

The focus of this contribution is to define speech acts as institutional actions where modeled institutions are social or informal institutions. With social or informal institution, we mean an institution which is grounded on the acceptances of its members. Differently from formal (legal) institutions, informal institutions are *rule-governed social practices* in which no member with ‘special’ powers to create and eliminate institutional facts is introduced (see Section IV).<sup>2</sup> After a short introduction about the concept of acceptance (Section II), we present a modal logic of acceptance, goal and action (Section III). On the basis of the notion of acceptance, we specify how a group of agents can create normative and institutional facts which hold only in an attitude-dependent way (Section IV). In Section V we define the promissive speech act.

## II. THE CONCEPT OF ACCEPTANCE

Whereas beliefs have been studied for decades, acceptances have only been examined since [7] and [8] while studying the nature of argument premises or reformulating Moore’s paradox [8]. 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” [9] without being necessarily connected to the actual truth of the proposition. Another difference between belief and acceptance is that beliefs are context-independent, whilst acceptance depends on context [10]. 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 acceptance is studied in private contexts, here we continue the work initiated in [11] by exploring the role of acceptance 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 individual and collective attitudes (individual and collective acceptances) that can be formally captured. In the context of Clue, for instance, agents accept that something has happened (see Example 2) *qua* players of Clue. The state of acceptance *qua* member of an institution is the kind of acceptance one is committed to when one is “functioning as member of the institution” [12].

## III. THE LOGICAL FRAMEWORK

### A. Syntax

The syntactic primitives of our logic of acceptance, actions and goals are the following:

- a finite set of  $n > 0$  agents  $AGT = \{1, 2, \dots, n\}$ ;

- a nonempty finite set of *atomic actions*  $AT = \{a, b, \dots\}$ ;
- a finite set of atomic formulas  $ATM = \{p, q, \dots\}$ ;
- a finite set of labels denoting institutional contexts  $INST = \{inst_1, inst_2, \dots, inst_m\}$ .

Moreover, we note  $2^{AGT^*} = 2^{AGT} \setminus \{\emptyset\}$  the set of all non empty subsets of agents,  $\Delta = \{C:x | C \in 2^{AGT^*}, x \in INST\}$  the set of all couples of non empty subsets of agents and institutional contexts.

The language  $\mathcal{LANG}$  of the logic  $\mathcal{L}$  is defined as the smallest superset of  $ATM$  such that: if  $\varphi, \psi \in \mathcal{LANG}$ ,  $i, j \in AGT$ ,  $\alpha \in ACT$  and  $C:x \in \Delta$  then  $\neg\varphi$ ,  $\varphi \vee \psi$ ,  $[i:\alpha:j]\varphi$ ,  $Goal_i\varphi$  and  $Accept_{C:x}\varphi \in \mathcal{LANG}$ , where the set  $ACT$  is the smallest superset of  $AT$  such that: if  $i, j \in AGT$ ,  $a \in AT$  then  $inf(i:a:j) \in ACT$  (*informative actions*). The classical boolean connectives  $\wedge$ ,  $\rightarrow$ ,  $\leftrightarrow$ ,  $\top$  (tautology) and  $\perp$  (contradiction) are defined from  $\vee$  and  $\neg$  in the usual manner.

The set  $ACT$  includes all informative actions of type “informing that agent  $i$  is going to perform (atomic) action  $a$  for agent  $j$ ”, where, with informative action, we mean the performance of an utterance, the act of making an enunciation. Since  $AT$  and  $AGT$  are finite sets, it follows that  $ACT$  as well is finite.

Operators of the form  $Accept_{C:x}$  have been introduced in [11] where a logic of acceptance has been proposed. These operators allow to express those facts that the agents in a group  $C$  accept while identifying themselves with a certain institution  $x$ .

Formula  $Accept_{C:x}\varphi$  has to be read “the agents in  $C$  accept that  $\varphi$  while functioning as members of the institution  $x$ ”.

EXAMPLE 1.  $Accept_{C:Greenpeace}protectEarth$  is read “the agents in  $C$  accept that the mission of Greenpeace is to protect the Earth while functioning as activists of Greenpeace”.

For  $C:x \in \Delta$ ,  $Accept_{C:x}\perp$  has to be read “agents in  $C$  are not functioning as members of the institution  $x$ ” because we assume that functioning as a member of an institution is, at least in this minimal sense, a rational activity; conversely,  $\neg Accept_{C:x}\perp$  has to be read “agents in  $C$  are functioning as members of the institution  $x$ ”;  $\neg Accept_{C:x}\perp \wedge Accept_{C:x}\varphi$  stands for “agents in  $C$  are functioning as members of the institution  $x$  and they accept that  $\varphi$  while functioning as members of  $x$ ” or simply “agents in  $C$  accept that  $\varphi$  *qua* members of the institution  $x$ ” (*i.e.* group acceptance).

The dynamic operators of the form  $[i:\alpha:j]$  are just a generalization of standard operators of dynamic logic [13] where both the author (initiator) and the addressee of a certain action  $\alpha$  are specified. Formula  $[i:\alpha:j]\varphi$  has to be read “after agent  $i$  does action  $\alpha$  for agent  $j$ , it is the case that  $\varphi$ ”. Operators of the form  $Goal_i$  are standard operators for agents’ goals [14].  $Goal_i\varphi$  has to be read “agent  $i$  has the goal that  $\varphi$  holds” (or “ $i$  wants  $\varphi$  to be true”).

<sup>2</sup>For the distinction between formal and informal institutions, see also [6].

The following abbreviations are given for any  $i, j \in AGT$ ,  $a \in AT$  and  $\alpha \in ACT$ :

$$\begin{aligned} \langle i:\alpha:j \rangle \varphi &\stackrel{\text{def}}{=} \neg [i:\alpha:j] \neg \varphi \\ Int_{i,j}(\alpha) &\stackrel{\text{def}}{=} Goal_i \langle i:\alpha:j \rangle \top \\ Inf_{i,j}(a) &\stackrel{\text{def}}{=} \langle i:inf(i:a:j):j \rangle \top \end{aligned}$$

Formula  $\langle i:\alpha:j \rangle \varphi$  is meant to stand for ‘‘agent  $i$  performs action  $a$  and  $\varphi$  is true afterward’’.  $Int_{i,j}(\alpha)$ : agent  $i$  intends to perform action  $\alpha$  for  $j$ .  $Inf_{i,j}(a)$ : agent  $i$  informs agent  $j$  that he is going to perform the (atomic) action  $a$  for him.

## B. Semantics

We use a possible worlds semantics. A model of the logic  $\mathcal{L}$  is a tuple  $\mathcal{M} = \langle W, \mathcal{A}, \mathcal{R}, \mathcal{G}, \mathcal{V} \rangle$  where:

- $W$  is a set of possible worlds;
- $\mathcal{A} : \Delta \rightarrow (W \rightarrow 2^W)$  associates each  $C:x \in \Delta$  and possible world  $w$  with the set  $\mathcal{A}_{C:x}(w)$  of possible worlds accepted by the group  $C$  in  $w$ , where the agents in  $C$  are functioning as members of the institution  $x$ ;
- $\mathcal{R} : AGT \times ACT \times AGT \rightarrow (W \rightarrow 2^W)$  associates each two agents  $i, j \in AGT$ , action  $\alpha \in ACT$  and possible world  $w$  with the set  $\mathcal{R}_{i:\alpha:j}(w)$  of possible worlds that are reachable from  $w$  through the occurrence of action  $\alpha$  performed by  $i$  for agent  $j$ ;
- $\mathcal{G} : AGT \rightarrow (W \rightarrow 2^W)$  associates each agent  $i \in AGT$  and possible world  $w$  with the set  $\mathcal{G}_i(w)$  of worlds that are compatible with agent  $i$ 's goals in  $w$ ;
- $\mathcal{V} : W \rightarrow 2^{ATM}$  is a truth assignment which associates each world  $w$  with the set  $\mathcal{V}(w)$  of atomic propositions true in  $w$ .

To the standard truth conditions for atomic formulas, negation and disjunction we add :

- $\mathcal{M}, w \models Accept_{C:x}\varphi$  iff for all  $w' \in W$ , if  $w' \in \mathcal{A}_{C:x}(w)$  then  $\mathcal{M}, w' \models \varphi$ ;
- $\mathcal{M}, w \models [i:\alpha:j]\varphi$  iff for all  $w' \in W$ , if  $w' \in \mathcal{R}_{i:\alpha:j}(w)$  then  $\mathcal{M}, w' \models \varphi$ ;
- $\mathcal{M}, w \models Goal_i\varphi$  iff for all  $w' \in W$ , if  $w' \in \mathcal{G}_i(w)$  then  $\mathcal{M}, w' \models \varphi$ .

## C. Axiomatization

Every operator of type  $Accept_{C:x}$  and  $[i:\alpha:j]$  is supposed to be a normal modal operator satisfying standard axioms and rules of inference of system  $K$ . Every operator  $Goal_i$  is supposed to be a normal modal operator satisfying standard axioms and rules of inference of system  $KD$ .<sup>3</sup> Axiom  $D$  for goals corresponds to the following constraint of seriality over  $\mathcal{L}$  models. For every  $i \in AGT$  and  $w \in W$ :

$$\mathcal{G}_i(w) \neq \emptyset \quad \mathbf{S1}$$

<sup>3</sup>Axiom  $D$  for  $Goal_i$  is:  $\neg(Goal_i\varphi \wedge Goal_i\neg\varphi)$ .

The rest of the section contains other axioms for acceptance, action and intention and corresponding semantic constraints over  $\mathcal{L}$  models.

We suppose that given a set of agents  $C$ , all  $B \subseteq C$  have access to all the facts that are (not) accepted by agents in  $C$  while functioning as members of the institution  $x$ . In particular, we suppose that: if agents in  $C$  (do not) accept that  $\varphi$  while functioning as members of the institution  $x$  then for every subset  $B$  of  $C$  and institution  $y$  while functioning as members of the institution  $y$ , agents in  $B$  accept that agents in  $C$  (do not) accept that  $\varphi$  while functioning as members of the institution  $x$ . Such properties are captured by the following two axiom schemas. For every  $C:x, B:y \in \Delta$ , if  $B \subseteq C$  then:

$$\begin{aligned} Accept_{C:x}\varphi &\rightarrow Accept_{B:y}Accept_{C:x}\varphi & \mathbf{4}_{Accept} \\ \neg Accept_{C:x}\varphi &\rightarrow Accept_{B:y}\neg Accept_{C:x}\varphi & \mathbf{5}_{Accept} \end{aligned}$$

Axioms  $\mathbf{4}_{Accept}$  and  $\mathbf{5}_{Accept}$  together correspond to the following semantic constraint over  $\mathcal{L}$  models. For every  $w \in W$  and  $C:x, B:y \in \Delta$ , if  $B \subseteq C$  then:

$$\text{if } w' \in \mathcal{A}_{B:y}(w) \text{ then } \mathcal{A}_{C:x}(w') = \mathcal{A}_{C:x}(w) \quad \mathbf{S2}$$

We also suppose that if agents in  $C$  accept that  $\varphi$  qua members of the institution  $x$  then, for every subset  $B$  of  $C$ , it holds that agents in  $B$  accept  $\varphi$  qua members of the institution  $x$ . Formally, for every  $C:x, B:x \in \Delta$ , if  $B \subseteq C$  then:

$$\begin{aligned} (\neg Accept_{C:x}\perp \wedge Accept_{C:x}\varphi) &\rightarrow \\ (\neg Accept_{B:x}\perp \wedge Accept_{B:x}\varphi) & \quad \mathbf{Inc}_{Accept} \end{aligned}$$

**EXAMPLE 2.** Imagine three agents  $i, j, k$  that, qua Clue players, accept that someone called Mrs. Red, has been killed:  $\neg Accept_{\{i,j,k\}:Clue}\perp \wedge Accept_{\{i,j,k\}:Clue}killedMrsRed$ . This implies that also the two agents  $i, j$  qua Clue players accept that someone called Mrs. Red has been killed:

$$\neg Accept_{\{i,j\}:Clue}\perp \wedge Accept_{\{i,j\}:Clue}killedMrsRed.$$

Axiom  $\mathbf{Inc}_{Accept}$  corresponds to the following semantic constraint over  $\mathcal{L}$  models. For every  $w \in W$ ,  $C:x, B:x \in \Delta$ , if  $B \subseteq C$ :

$$\begin{aligned} \text{if } \mathcal{A}_{C:x}(w) \neq \emptyset \\ \text{then } \mathcal{A}_{B:x}(w) \neq \emptyset \text{ and } \mathcal{A}_{B:x}(w) \subseteq \mathcal{A}_{C:x}(w) \end{aligned} \quad \mathbf{S3}$$

We suppose the following additional constraint over  $\mathcal{L}$  models. For every  $w \in W$ ,  $i, j, i', j' \in AGT$  and  $\alpha, \beta \in ACT$ :

$$\text{if } w' \in \mathcal{R}_{i:\alpha:j}(w) \text{ and } w'' \in \mathcal{R}_{i':\beta:j'}(w) \text{ then } w' = w'' \quad \mathbf{S4}$$

The property  $\mathbf{S4}$  says that all actions occurring in a world  $w$  lead to the same world. Thus, all actions occur in parallel and they do not have non-deterministic effects. This explains why we have phrased  $\langle i:\alpha:j \rangle \varphi$  ‘‘ $i$  does  $\alpha$  for  $j$  and  $\varphi$  holds afterward’’ rather than ‘‘it is possible that  $i$  does  $\alpha$  for  $j$  and  $\varphi$  holds afterward’’. Constraint  $\mathbf{S4}$  corresponds to the following axiom of our logic. For every  $i, j, i', j' \in AGT$  and  $\alpha, \beta \in ACT$ :

$$\langle i:\alpha:j \rangle \varphi \rightarrow [i':\beta:j'] \varphi \quad \mathbf{Det}$$

We also suppose that the world is never static in our framework, that is, for every world  $w$  there exists at least two agents  $i, j$  and action  $\alpha$  such that  $i$  performs  $\alpha$  for  $j$  at  $w$ . Formally, given a  $\mathcal{L}$  model  $M$ , for every  $w \in W$  we have that:

$$\exists i, j \in AGT, \exists \alpha \in ACT, \exists w' \in W \text{ s.t. } w' \in \mathcal{R}_{i:\alpha:j}(w) \quad \mathbf{S5}$$

Property **S5** of  $\mathcal{L}$  models corresponds to the following axiom of our logic.

$$\bigvee_{i,j \in AGT, \alpha \in ACT} \langle i:\alpha:j \rangle \top \quad \mathbf{Active}$$

Axiom **Active** ensures that for every world  $w$  there is a *next* world of  $w$  which is reachable from  $w$  by the occurrence of some action performed by some agent for another agent. This is the reason why the operator  $X$  for *next* of LTL (linear temporal logic) can be defined as follows:<sup>4</sup>

$$\mathbf{X}\varphi \stackrel{def}{=} \bigvee_{i,j \in AGT, \alpha \in ACT} \langle i:\alpha:j \rangle \varphi$$

where  $\mathbf{X}\varphi$  is meant to stand for “ $\varphi$  will be true in the next state”.

The following axiom relates intentions with actions. For every  $i, j \in AGT$  and  $\alpha \in ACT$ :

$$\langle i:\alpha:j \rangle \top \rightarrow \mathit{Int}_{i,j}(\alpha) \quad \mathbf{IntAct}$$

According to Axiom **IntAct**, an agent  $i$  performs action  $\alpha$  for agent  $j$  only if he has the intention to do  $\alpha$  for  $j$ . In this sense we suppose that an agent’s *doing* is by definition intentional. **IntAct** corresponds to the following semantic constraint over  $\mathcal{L}$  models. For every  $i, j \in AGT$ ,  $\alpha \in ACT$  and  $w \in W$ :

$$\begin{aligned} &\text{if } \exists v' \text{ such that } v' \in \mathcal{R}_{i:\alpha:j}(w) \text{ then} \\ &\forall w' \in \mathcal{G}_i(w), \exists w'' \text{ such that } w'' \in \mathcal{R}_{i:\alpha:j}(w') \end{aligned} \quad \mathbf{S6}$$

We call  $\mathcal{L}$  the logic axiomatized by the principles presented above and we write  $\vdash_{\mathcal{L}} \varphi$  iff formula  $\varphi$  is a theorem of  $\mathcal{L}$ . Moreover, we write  $\models_{\mathcal{L}} \varphi$  iff formula  $\varphi$  is *valid* in all  $\mathcal{L}$  models, *i.e.*  $\mathcal{M}, w \models \varphi$  for every  $\mathcal{L}$  model  $\mathcal{M}$  and world  $w$  in  $\mathcal{M}$ . Finally, we say that a formula  $\varphi$  is *satisfiable* if there exists an  $\mathcal{L}$  model  $\mathcal{M}$  and a world  $w$  in  $\mathcal{M}$  such that  $\mathcal{M}, w \models \varphi$ .

#### IV. INSTITUTIONAL CONCEPTS

Normative and institutional facts are a class of facts that are typical of institutional contexts. According to [15], such facts have the peculiar feature of being dependent on the agents’ attitudes in a way that we are in the position to specify in detail in the logic  $\mathcal{L}$ . More precisely it has been noted that these facts are characterized at least by two features [16], [15], [17]: *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 3.** *If the agents qua group members accept 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 the logic  $\mathcal{L}$  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.

##### A. Truth in an institutional context

We formalize the notion of truth w.r.t. a certain institutional context with the operator  $\llbracket x \rrbracket$ . A formula  $\llbracket x \rrbracket \varphi$  is read “within the institutional context  $x$ , it is the case that  $\varphi$ ”. Here we suppose that “within the institutional context  $x$  it is the case that  $\varphi$ ” if and only if “for every set of agents  $C$ , the agents in  $C$  accept that  $\varphi$  while functioning as members of the institution  $x$ ”. We introduce the following formal definition.

*Definition 1:*

$$\llbracket x \rrbracket \varphi \stackrel{def}{=} \bigwedge_{C \in 2^{AGT^*}} \mathit{Accept}_{C:x} \varphi$$

It is straightforward to prove that every  $\llbracket x \rrbracket$  is a normal modal operator. Given the previous analysis, a fact is true w.r.t. an institutional context  $x$  if and only if such fact is accepted by all the agents while they function as members of  $x$  (hence the performativity and the reflexivity principles are maintained).

At this point, it might be objected that there are facts which are true in an institutional context but only “special” members of the institution are aware of them and can change them. For instance, there are laws in every country which are known and can be changed only by the specialists of the domain (lawyers, judges, members of the parliament, *etc.*). In order to resist to this objection recall that, at this stage, our model applies to informal institutions of a society, in particular to language. Relative to this restriction, the proposed assumption is justified because, for informal institutions, there is no special agent who has the power to create and eliminate institutional facts characterizing the institution itself (*i.e.* nobody has the power to change the rules of the speech act *promise* in the context of language).

Finally, the following abbreviation is given.

*Definition 2:*

$$\llbracket Univ \rrbracket \varphi \stackrel{def}{=} \bigwedge_{x \in INST} \llbracket x \rrbracket \varphi$$

$\llbracket Univ \rrbracket \varphi$  stands for “ $\varphi$  is universally accepted as true”.

<sup>4</sup>Note that  $X$  satisfies the standard property  $X\varphi \leftrightarrow \neg X\neg\varphi$ .

## B. Constitutive rules

From the concept of truth with respect to an institutional context a notion of *constitutive rule* of the form “ $\varphi$  counts as  $\psi$  in the institutional context  $x$ ” can be defined. We conceive a constitutive rule as a material implication of the form  $\varphi \rightarrow \psi$  in the scope of an operator  $\llbracket x \rrbracket$ . We suppose that a constitutive rule is intrinsically local, that is, a rule that is not universally valid while it is accepted by the members of a certain institution. More generally, for every  $x \in INST$  the following abbreviation (that stands for “ $\varphi$  counts as  $\psi$  in the institutional context  $x$ ”) is given.<sup>5</sup>

*Definition 3:*

$$\varphi \triangleright^x \psi \stackrel{def}{=} \llbracket x \rrbracket (\varphi \rightarrow \psi) \wedge \neg \llbracket Univ \rrbracket (\varphi \rightarrow \psi)$$

EXAMPLE 4. *In the context of gestural language there exists a constitutive rule according to which, the nodding gesture “counts as” an endorsement of what the speaker is suggesting, noted nodding<sup>gesture</sup>  $\triangleright$  yes. But in different contexts the same gesture may express exactly the opposite fact (viz.  $\neg \llbracket Univ \rrbracket$  (nodding  $\rightarrow$  yes) holds).*

## C. Obligations

Informal institutions such as language involve a deontic dimension that up to now we have ignored. In order to capture this core feature we extend the logic  $\mathcal{L}$  by introducing a *violation* atom  $V$  as in Anderson’s reduction of deontic logic to alethic logic [19]. By means to this new formal construct we can formally characterize the obligations which are valid in a certain institution by anchoring them in the acceptances of the members of that institution.

We say that “in the institutional context  $x$ , agent  $i$  has the obligation to perform action  $\alpha$  for agent  $j$  under the condition  $\varphi$ ” (noted  $Obl_x(i, \alpha, j, \varphi)$ ) if and only if “in the institutional context  $x$ , the fact that  $i$  does not perform action  $\alpha$  for  $j$  under the condition  $\varphi$  counts as a violation at the next step of time”.<sup>6</sup> Formally:

*Definition 4:*

$$Obl_x(i, \alpha, j, \varphi) \stackrel{def}{=} (\varphi \wedge [i:\alpha:j] \perp) \triangleright^x \mathbf{XV}$$

EXAMPLE 5. *Formula  $Obl_{EBay}(i, sendGoods, j, paid(j, i))$  expresses that, in the context of EBay, it is obligatory for an agent  $i$  to send certain goods to agent  $j$  under the condition that  $j$  has paid the goods to  $i$ .*

We here distinguish an obligation which is valid in a certain institution  $x$  (viz. institutional obligation) from the instantiations of this obligation in specific groups of agents which are

<sup>5</sup>Our notion of constitutive rule of the form  $\varphi \triangleright^x \psi$  is similar to the notion of *proper classificatory rule* given in [18].

<sup>6</sup>Formulas of type  $Obl_x(i, \alpha, j, \varphi)$  can be conceived as particular instances of so-called *regulative rules* in Searle’s sense [15]. On the distinction between *regulative rule* and *constitutive rule* see also [20].

members of this institution. We say that  $i$ ’s obligation to do action  $\alpha$  for  $j$  is instantiated in the group  $C$  of members of institution  $x$  (noted  $IObl_{C:x}(i, \alpha, j)$ ) if and only if the agents in  $C$ , *qua* members of the institution  $x$ , accept that if  $i$  does not perform action  $\alpha$  for  $j$  then  $i$  will incur a violation in the next state. Formally:

*Definition 5:*

$$IObl_{C:x}(i, \alpha, j) \stackrel{def}{=} \neg Accept_{C:x} \perp \wedge Accept_{C:x}([i:\alpha:j] \perp \rightarrow \mathbf{XV})$$

The following Theorem highlights the relationships between institutional obligations and instantiations of obligations in groups.

*Theorem 1:*

$$\vdash_{\mathcal{L}} (Obl_x(i, \alpha, j, \varphi) \wedge \neg Accept_{C:x} \perp \wedge Accept_{C:x} \varphi) \rightarrow IObl_{C:x}(i, \alpha, j)$$

According to Theorem 1, if the agents in  $C$ , *qua* members of institution  $x$ , accept that  $\varphi$  and, in  $x$ ,  $i$  has the obligation to perform action  $\alpha$  for agent  $j$  under the condition  $\varphi$  then,  $i$ ’s obligation to do action  $\alpha$  for  $j$  is instantiated in the group  $C$  of members of  $x$ .

EXAMPLE 6. *From the fact that in EBay, it is obligatory for an agent  $i$  to send certain goods to agent  $j$  under the condition that  $j$  has paid the goods to  $i$  (noted  $Obl_{EBay}(i, sendGoods, j, paid(j, i))$ ), and the fact that the group  $C$  of EBay surfers accept that  $j$  has paid certain goods to  $i$  (noted  $\neg Accept_{C:EBay} \perp \wedge Accept_{C:EBay} paid(j, i)$ ), we infer that  $i$ ’s obligation to send the goods to  $j$  is instantiated in the group  $C$  of EBay surfers (noted  $IObl_{C:EBay}(i, sendGoods, j)$ ).*

## D. Social commitment

Social commitment has a fundamental role in the interaction between agents in an institution. It fixes the relations between agents, bounding an agent toward another. Social commitment is thus a relational notion: it relies at least two agents, the agent who is committed (the *debtor* [21]) and the agent to whom the debtor is committed (the *creditor*). According to Castelfranchi [22], there are two crucial aspects of social commitment: a motivational aspect and a deontic aspect. If the debtor  $i$  is committed to the creditor  $j$  to perform action  $\alpha$  for him then, the creditor and the debtor must mutually know that the creditor is interested in the fact that the debtor performs  $\alpha$  (motivational aspect) and, the creditor must have specific rights on the debtor being entitled to ask to the debtor to perform action  $\alpha$  (deontic aspect).

In the present analysis we specify these two conditions for social commitment on the basis of the notions of acceptance and instantiated obligation. Differently from Castelfranchi, we make explicit the institutional dimension of social commitment by specifying the institution in which the commitment is

established. On the one hand, we characterize the deontic aspect of  $i$ 's commitment to  $j$  to perform action  $\alpha$  by the fact that  $i$  and  $j$  function as members of a certain institution  $x$  and, as members of  $x$ ,  $i$  and  $j$  accept that  $i$  is obliged to perform action  $\alpha$  for  $j$ . That is,  $i$ 's obligation to do action  $\alpha$  for  $j$  is instantiated in the group  $\{i, j\}$  of members of  $x$ . On the other hand, we characterize the motivational aspect of  $i$ 's commitment to  $j$  by the fact that  $i$  and  $j$ , *qua* members of the institution  $x$ , accept that  $j$  wants  $i$  to do  $\alpha$  for him. Formally:

*Definition 6:*

$$SC_x(i, \alpha, j) \stackrel{def}{=} IObl_{\{i, j\}:x}(i, \alpha, j) \wedge \neg Accept_{\{i, j\}:x} \perp \wedge Accept_{\{i, j\}:x} Goal_j \langle i: \alpha: j \rangle \top$$

$SC_x(i, \alpha, j)$  stands for “ $i$  is committed to  $j$  to do  $\alpha$  w.r.t. institution  $x$ .”

a) *Related works on social commitment:* We can note that Castelfranchi also requires in his definition of social commitment the condition of  $i$  and  $j$ 's mutual knowledge about  $i$ 's intention to perform  $\alpha$  for  $j$ . We do not include this condition here since in our view it is too strong because it logically implies that  $i$ 's intention actually holds (under the commonly accepted hypothesis of introspection on intentions). But as Castelfranchi remarks himself, the actual intention of  $i$  to perform the action  $\alpha$  is neither necessary nor sufficient for his social commitment to do  $\alpha$ : the entailment link between social commitment and individual intention is represented by the hypothesis that the agent is honest. Moreover we consider that, after a promise, the social commitment can persist even if the debtor drops the intention that he manifested.

In opposition to Singh [23] and works on ACLs [21], [24], [25], in this work social commitment is not taken primitive and is anchored in group attitudes. In particular, our notion of social commitment is grounded on the acceptance of the debtor and the creditor *qua* members of a certain institution.

Colombetti et al. [26], [21] have developed an ACL semantics based on a *primitive* notion of commitment and its dynamics (through its states change). For example, when an agent promises to another one to perform an action, he creates a *pending commitment* (coinciding to the social commitment in the commonly used sense), that becomes *fulfilled* (resp. *violated*) if the agent performs (resp. do not perform) the action. The commitment can also be *anceled* by the creditor. When an agent is requested to performed an action, an *unset* commitment is created, which can be accepted (resp. refused) by the agent (the commitment becomes *pending* (resp. *anceled*)).

An *unset* commitment of  $c$  toward  $d$  represents the particular social relation that results from the performance by agent  $d$  of a request, an order or another directive speech act. This kind of commitment is quite disputable in particular in the case of an order. When I am driving my car and a policeman motions me to stop, I have automatically the obligation to stop my car. I have no refusal opportunity, neither do I have

to accept to stop. This example can hardly be represented in the Colombetti's account. We cannot say that I have an *unset* commitment to stop my car and that I can accept or refuse this commitment. We neither can say that an *pending* commitment is automatically created, because indeed I am not committed but only obliged to stop. Our framework is able to handle such a fine example because it contains both notions of commitment and obligation.

In Colombetti et al.'s account, nothing is said about the consequences of the cancelation of a commitment. In our characterization, after the creditor utters that he does not want anymore the action to be performed by the debtor, the debtor's commitment is dropped. But the the debtor's obligation might remain instantiated. Indeed, formula  $\neg SC_x(i, \alpha, j) \wedge IObl_{\{i, j\}:x}(i, \alpha, j)$  is satisfiable in our logic.

## V. APPLICATION TO THE PROMISE

Speech acts can be viewed as institutional facts represented by constitutive rules. In this section, we focus on a particular kind of speech act: the promissive. We first define the constitutive rule which creates the relation between an utterance (as a physical action) and the speech act *promise* (as an institutional action). Then, we specify the deontic dimension of promise. Finally, we establish the relationship between promise and social commitment.

According to [3], in the context of ordinary communication (noted *OC*) the speech act *promise* is defined on the basis of a constitutive rule of the form: a speaker's enunciation of the utterance: “I am going to perform action  $a$ ” counts as a promise to the hearer to perform action  $a$ , under the condition that the hearer wants the speaker to perform action  $a$ . We here consider the constitutive rule defining the speech act *promise* with respect to the institutional context of ordinary communication (*OC*) as a global axiom [27]. For every  $i, j \in AGT$  and  $a \in AT$  we suppose that:

$$(Inf_{i, j}(a) \wedge Goal_j \langle i: a: j \rangle \top) \stackrel{OC}{\triangleright} \mathbf{X} Promise(i, a, j) \quad (1)$$

where  $Promise(i, a, j)$  is an atom denoting the *institutional fact* that  $i$  has promised to  $j$  to do action  $a$  for him. The previous constitutive rule says that: in the context of ordinary communication, for every agents  $i, j$  and atomic action  $a$ ,  $i$ 's act of informing  $j$  that is going to perform  $a$  for him, under the condition in which  $j$  wants  $i$  to do  $a$ , counts as a promise of  $i$  to  $j$  to perform  $a$  for him. The reason why we have  $\mathbf{X} Promise(i, a, j)$  instead of  $Promise(i, a, j)$  in the consequent of the previous counts-as assertion is that the atom  $Promise(i, a, j)$  represents the institutional effect of  $i$ 's act of informing  $j$  that is going to perform  $a$  for him. Thus,  $Promise(i, a, j)$  must necessarily hold after the occurrence of  $i$ 's act of informing.

Note that the previous constitutive rule defining the speech act *promise* in the context of ordinary communication can not

be generalized to all institutional contexts. For example, in the context of the card game of poker the enunciation of the utterance: “I am going to perform a certain action” does not necessarily count as a promise of the speaker’s to the hearer to perform the action in question. Indeed, in the context of poker players are allowed to bluff. More generally, the following formula should be acceptable for the context of poker:

$$\neg((\text{Inf}_{i,j}(a) \wedge \text{Goal}_j \langle i:a:j \rangle \top) \triangleright^{\text{Poker}} \mathbf{X} \text{Promise}(i,a,j))$$

By definition, institutional facts are connected to a deontic dimension. In particular, an institutional fact is intrinsically connected to certain normative facts expressed in terms of obligations and permissions. For example, “being of age” is an institutional fact in many countries to which a certain number of permissions and obligations are associated (*e.g.* in many countries if you are of age you have the permission to vote and the obligation to fulfill the military duties). Therefore, since  $\text{Promise}(i,a,j)$  is an institutional fact, it must be connected to certain normative facts. The following global axiom is given in order to establish such a connection with the normative level of *promise*. For every  $i, j \in \text{AGT}$  and  $a \in \text{AT}$  we suppose that:

$$\text{Obl}_{OC}(i, a, j, \text{Promise}(i, a, j)) \quad (2)$$

$\text{Obl}_{OC}(i, a, j, \text{Promise}(i, a, j))$  means that: in the context of ordinary communication, under the condition that  $i$  has promised to  $j$  to perform action  $a$  for him,  $i$  is obliged to perform action  $a$  for  $j$ .

*Theorem 2:*

$$\vdash_{\mathcal{L}} (\neg \text{Accept}_{\{i,j\}:OC} \perp \wedge \text{Accept}_{\{i,j\}:OC} \text{Goal}_j \langle i:a:j \rangle \top \wedge \text{Accept}_{\{i,j\}:OC} \text{Promise}(i, a, j)) \rightarrow \text{SC}_{OC}(i, a, j)$$

This theorem highlights the relationship between promise and social commitment in the context of ordinary communication: the fact that in the context of ordinary communication  $i$  and  $j$  accept that  $i$  has promised to agent  $j$  to do action  $a$  for him and that  $j$  wants  $i$  to do  $a$  for him entails the fact that  $i$  is committed to  $j$  to do  $a$ .

## VI. CONCLUSION

We have introduced a logic of acceptance, action and goal and provided an institution-based semantics of the commissive act of promise. As far as we know, the only approach which is similar to ours is Colombetti’s approach [28] where speech acts are also modeled in terms of constitutive rules of the form “ $X$  count as  $Y$ ” and of the institutional effects brought about their performances. However, in Colombetti’s approach, there is no connection between institutional level and agentive level of speech acts and the relationship between constitutive rules defining speech acts and agents’ attitudes is not investigated. In our approach this relationship is established by anchoring constitutive rules and obligations in agents’ acceptances.

In the future will extend our analysis to directive speech acts. To this end, we will need to model the hierarchy between roles in an institution which entitles an agents playing a certain role (*e.g.* the employer of a company) to make orders to other agents playing other roles (*e.g.* the employees).

## REFERENCES

- [1] F. de Saussure, *Course in General Linguistics*. La Salle, Illinois: Open Court Publishing, 1986.
- [2] J. L. Austin, *How To Do Things With Words*. Oxford University Press, 1962.
- [3] J. R. Searle, *Speech acts: An essay in the philosophy of language*. New York: Cambridge University Press, 1969.
- [4] —, “What is an institution?” *Journal of Institutional Economics*, vol. 1, no. 1, pp. 1–22, 2005.
- [5] A. J. I. Jones and X. Parent, “Conventional signalling acts and conversation,” in *Advances in Agent Communication*, F. Dignum, Ed. Springer, 2004, vol. 2922, pp. 1–17.
- [6] D. North, *Institutions, Institutional Change, and Economic Performance*. Cambridge: Cambridge University Press, 1990.
- [7] R. Stalnaker, *Inquiry*. Cambridge: MIT Press, 1984.
- [8] L. J. Cohen, *An essay on belief and acceptance*. New York, USA: Oxford University Press, 1992.
- [9] P. Hakli, “Group beliefs and the distinction between belief and acceptance,” *Cognitive Systems Research*, vol. 7, pp. 286–297, 2006.
- [10] M. E. Bratman, “Practical reasoning and acceptance in context,” *Mind*, vol. 101, no. 401, pp. 1–15, 1992.
- [11] B. Gaudou, D. Longin, E. Lorini, and L. Tummolini, “Anchoring Institutions in Agents’ Attitudes: Towards a Logical Framework for Autonomous MAS,” in *Proc. of AAMAS’08*, to appear.
- [12] R. Tuomela, *The Philosophy of Sociality*. Oxford: Oxford University Press, 2007.
- [13] D. Harel, D. Kozen, and J. Tiuryn, *Dynamic Logic*. Cambridge: MIT Press, 2000.
- [14] P. R. Cohen and H. J. Levesque, “Intention is choice with commitment,” *Artificial Intelligence Journal*, vol. 42, no. 2–3, pp. 213–261, 1990.
- [15] J. R. Searle, *The Construction of Social Reality*. New York: The Free Press, 1995.
- [16] E. Lagerspetz, *The opposite mirrors*. Dordrecht: Kluwer, 1995.
- [17] R. Tuomela, *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge: Cambridge University Press, 2002.
- [18] D. Grossi, J.-J. C. Meyer, and F. Dignum, “Classificatory aspects of counts-as: An analysis in modal logic,” *Journal of Logic and Computation*, vol. 16(5), pp. 613–643, 2006.
- [19] A. Anderson, “A reduction of deontic logic to alethic modal logic,” *Mind*, vol. 22, pp. 100–103, 1958.
- [20] G. Boella and L. Van der Torre, “Regulative and constitutive norms in normative multiagent systems,” in *Proc. of KR 2004*. AAAI Press, 2004, pp. 255–266.
- [21] N. Fornara, F. Viganò, and M. Colombetti, “Agent communication and artificial institutions,” *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, vol. 14, pp. 121–142, 2007.
- [22] C. Castelfranchi, “Commitment: from individual intentions to groups and organizations,” in *Proc. of ICMAS’95*. MIT press, 1995, pp. 528–535.
- [23] M. P. Singh, “A social semantics for agent communication languages,” in *Issues in Agent Communication*, ser. LNCS, F. Dignum and M. Greaves, Eds. Springer, 2000, vol. 1916, pp. 75–88.
- [24] A. U. Mallya and M. P. Singh, “An algebra for commitment protocols,” *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, vol. 14, pp. 143–163, 2007.
- [25] R. A. Flores, P. Pasquier, and B. Chaib-draa, “Conversational semantics with social commitments,” *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, vol. 14, pp. 165–186, 2007.
- [26] N. Fornara and M. Colombetti, “Operational specification of a commitment-based agent communication language,” in *Proceedings of the First International Conference on Autonomous Agents and Multi-agent Systems (AAMAS02)*. Bologna, Italy: ACM Press, 2002, pp. 536–542.



- [27] M. Fitting, *Proof Methods for Modal and Intuitionistic Logics*. Dordrecht: D. Reidel, 1983.
- [28] M. Colombetti and M. Verdicchio, "An analysis of agent speech acts as institutional actions," in *Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems (AAMAS02)*. ACM Press, 2002, pp. 1157–1164.