Formalising the institutional interpretation of actions in an extended BDI logics

Carole Adam[1], Vincent Louis[1], and Robert Demolombe[2]

[1] Orange Labs, Lannion, France
[2] IRIT, Toulouse, France

Abstract. More and more logical frameworks provide social and normative concepts for multi-agent systems. Yet only a few of them combine these social aspects with the mental attitudes of individual agents, and when they do so they are restricted to the semantics of speech acts. In this paper we propose an original framework allowing to express both the institutional and intentional semantics of all kinds of actions, be they communicative or material.

1 Introduction

More and more logical frameworks allow agents to reason about social and institutional concepts [16, 3, 17], but often independently of their mental attitudes [12, 20]; only a few frameworks combine mental attitudes and social commitments [2, 14]. Moreover most of them are only dedicated to the semantics of communicative acts [14, 10], whereas as far as we know, no work provides an institutional semantics for material actions. The aim of this paper thus is to provide a logical framework combining the intentional and institutional dimensions of the performance of all kinds of actions, material or communicative. This paper thus has a double interest.

We would like to highlight that we consider here a very large notion of institution, since by institution we mean a set of rules and facts that are adopted by a group of agents (the members of the institution). This allows to account for as various institutional contexts as the law of a country, a contract between two parties in a business relationship, a social structure, the rules of a game; these can be formal legal institutions as well as informal ones.

The structure of the paper is the following. We begin (Section 2) with a state of the art of existing logical frameworks providing social notions. We then identify their limitations and propose our own generic logical framework (Section 3) and use it to define a model of the institutional interpretation of actions (Section 4). We then illustrate the expressivity of this model (Section 5) by formalising two actions: a communicative one, the Declare speech act; and a material one in the application field of B2B exchanges[3]. Finally (Section 6) we conclude about our future prospects.

[3] This example is extracted from our application of multi-agent mediation platform for automatic B2B exchanges, that will also be presented as a demonstration at ESAW.

2 State of the art

Approaches to the formalisation of the artificial institutions that influence agents' reasoning and behaviour divide in several main trends. One trend studies the notion of commitment [5] representing what an agent is publicly committed to [20, 12]. Another trend is based on the notion of acceptance [15] representing what a group of agents accepts as true while functioning in a given institutional context [14]. In this section we give some details about one approach in each of these trends.

2.1 A social commitments-based approach: Fornara and Colombetti

We detail here the logical framework of Fornara and Colombetti [12] that provides a formalisation of Castelfranchi's notion of commitment [5].

The formalisation of commitments Fornara and Colombetti consider a social commitment as an object described by an internal structure, a life cycle, and a set of methods to manipulate it. A commitment is represented by a predicate $C_{id}(state, debtor, creditor, content|conditions[, timeout])$:

- *id* is the unique identifier of the commitment;
- *state* is the current state of the commitment during its life cycle, among those detailed above;
- *debtor* is the debtor of the commitment, that is the committed agent;
- *creditor* is the creditor of the commitment, that is the agent to whom the debtor is committed;
- *content* is a temporal proposition representing the content of the commitment, that is the state or action to which the debtor is committed;
- *conditions* is a temporal proposition (whose time interval must precede the content's one) representing the condition to validate in order to get a conditional commitment to become active;
- *timeout* is optional since only unset commitments use it. An unset commitment that is not accepted, refused, or directly fulfilled before the end of this timeout is cancelled.

The content and conditions of a commitment are represented by a "temporal proposition", making explicit the interval of time during which the content is true. For example such properties allow to express the fact that a property must be true during a given interval, or that an action must be performed between two given instants.

Fornara and Colombetti provide two basic methods allowing to manipulate a commitment: creation (in a pending state), and state change. A commitment can also be modified by an event in the environment that affects the value of its condition or content. The authors give some rules relative to these environment events and build a finite state machine to describe the life cycle of a commitment [12, fig.1 p.4].

2.2 A group acceptance-based approach: Lorini *et al.*

Notion of acceptance Lorini *et al.* [14] build on Hakli's work [15] to define acceptance as "a decision to treat p as true in one's utterances and actions", independently of the agent's belief about p or of the actual truth of p. Actually there are several important differences between acceptance and belief: acceptances are voluntary, resulting from a decision that can be influenced by pragmatic reasons but not by evidence, qualitative and dependent on context; on the contrary beliefs are involuntary, only aiming at truth and thus influenced by evidence, quantitative, and independent of context.

For the authors institutional contexts are "rule-governed social practices on the background of which the agents reason". They actually consider informal institutions where no agent has special powers, and where rules are decided and accepted by all agents.

Acceptance Logic Lorini *et al.* define a logical language called Acceptance Logic that is a propositional language extended with a set of agents, a set of actions and a set of institutional contexts. They define an acceptance operator denoted $[C : x]\varphi$ meaning that "the agents in the non-empty group C accept that φ while functioning as group members in the institutional context x". Acceptance is supposed to be rational so $[C : x]\bot$ means that agents in C do not function as a group in the institutional context x. The authors introduce a particular institutional context denoted λ and called "private context". They can thus identify individual beliefs with a particular acceptance concerning one single agent functioning in this private context: $[\{i\} : \lambda]\varphi$ is read "i believes that φ". This acceptance operator is normal and is defined with a standard possible worlds semantics.

Axioms and properties The authors then study some properties of acceptances and their relations with other modalities. They assume the following hypothesis by considering corresponding axioms:

- the acceptances of a group are accessible to all subgroups of this group: that is, if a group C accepts φ in context x (*i.e.* $[C : x]\varphi$), all subgroups B of C accept in any context y that this acceptance is true (*i.e.* $[B : y][C : x]\varphi$). In particular (if $B = \{i\}$ and $y = \lambda$) we have that all members of group C are aware of acceptances of group C. Another particular case (when $B = C = \{i\}$ and $x = y = \lambda$) is the introspection axioms for individual beliefs;
- the acceptances of a group are shared by all its subgroups: that is, if a group C accepts φ in context x then all its subgroups B also accept φ in this same context x;
- Individual beliefs are rational, that is agents cannot believe contradictions. On the contrary acceptances are not (the fact that a group accepts a contradiction in an institutional context actually means that it is not functioning as a follower of this context).

Therefore they can deduce the following properties:

- group acceptances are public: there is an equivalence between the acceptance of φ by agents in C functioning as a group in institutional context x (denoted $[C : x]\varphi$) and the common belief in group C that this acceptance holds. The common belief in C that φ means that "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";
- group acceptances can diverge from the individual beliefs of the members of the group: one can privately believe something but publicly accept the opposite while functioning as a member of the group.

Institutional facts The authors then use their notion of acceptance to define non-primitive institutional operators. Thus their institutional notions depend on the agents' mental attitudes, and are not imposed by an external institution.

In particular they identify institutional facts (facts that are true in a given institutional context x, denoted $[x]\varphi$) with an acceptance by all groups of agents while functioning as a group in this institutional context. This implies that all agents are always aware of all institutional facts, but this is compatible with their limitation to informal institutions.

They also define "contextual conditionals" (similar to count as relationships [17]) denoted $\varphi \rhd_x \psi$ as a material implication locally recognised as an institutional fact: this material implication must be true in the context x but must not be true in all institutional contexts.

Finally they also have standard deontic operators to represent obligations and permissions.

2.3 Critics and transition

These frameworks are interesting, though we can identify several limitations. Fornara and Colombetti's notion of commitment does not have a parameter making explicit the institution in which this commitment is valid; actually this commitment may be valid in what Lorini *et al.* call the informal institutional context of "ordinary communication", but it has no legal value. A second important restriction of their framework is that it is limited to social aspects and does not formalise the agents' mental attitudes.

Lorini *et al.* address this problem since their notion of group acceptance is a social and public notion, while anchored in the agents' mental attitudes. However they do not have dynamic operators, preventing them from formalising the notion of power as well as the institutional dimension of actions. Moreover, even if the institutional context is explicit in their operators, they are limited to the case of informal institutions, where institutional facts are equated with the agents' acceptances. This is a restrictive hypothesis that is not valid in other kinds of institutions like legal ones, so our approach is more generic.

In order to make the importance of institutional context clearer, we can illustrate it on several examples showing that the institutional interpretation of actions depends on the institutional context. For example the action of nodding head is interpreted in the context of French gestural language as meaning "yes", while in the context of Bulgarian gestural language it is interpreted as meaning "no". Another example is that of a salesman that makes an offer to a client with a very low price; in the institutional context of trading, the salesman's offer was permitted (since he has the role of seller) and engages his company to sell the product at the proposed price; at the same time this action was forbidden in the rules of his company because one must not sell at loss, so he is liable for some sanction. Thus the interpretation of the performance of an action depends on the rules of the particular institutional context, so it is important to make it explicit.

In this paper we intend to provide a logical framework extended with institutional notions in order to be able to formalise the institutional interpretation of actions. This will allow to characterise the particular features of actions (both material and communicative) in an institutional context: permission conditions, normative effects... We want to unify the formalisations of both intentional and institutional dimensions of actions in our framework, which thus generalises the approaches presented above.

3 Our logical framework

Our logical framework is based on a BDI logic, a formalism classically used to represent the reasoning of autonomous agents [21, 23]. In order to formalise institutional notions one can extend BDI logics with deontic operators like obligation [9, 3] or with institutional operators like count as or institutional power [17]. We have developed such an extended logical framework in previous work [7] and we will use this language here.

This framework allows to combine in the same formalism both private (mental attitudes) and public (obligations and commitments) notions. It is expressive enough to represent both the agents' commitments and their mental attitudes, while a commitment-based approach alone like Fornara and Colombetti's one cannot express these mental attitudes. Therefore we will be able to express both the intentional and institutional semantics of actions.

3.1 Modalities

Epistemic modalities Our logical framework is a first-order predicate logic completed by some modalities. $B_i p$ means that agent i believes that p. This operator has a standard $KD45$ axiomatic. $I_i p$ means that agent i intends that p.

Dynamic modalities This framework also comes with dynamic modalities. $done(i, \alpha, \varphi) = \neg before_{i:\alpha} \neg \varphi$ means that agent i has just performed procedure α, and φ was true before. $happens(i, \alpha, \varphi) = \neg after_{i:\alpha} \neg \varphi$ means that agent i is about to perform procedure α and φ will be true just after. $before_{i:\alpha}$ and $after_{i:\alpha}$ are normal modal operators defined in standard tense logic in linear time version [4].

Deontic modalities This classical BDI framework is extended with deontic modalities, in particular an impersonal obligation to be. $O\varphi$ reads "it is obligatory that φ", and its axiomatic is that of the Standard Deontic Logic [16], that is KD. *Obligations to do* can be expressed as obligations to be in a state where the obliged action has been performed. Obligations are impersonal since no agent is explicitly responsible for their fulfilment, but such an agent can implicitly appear in their content. For instance $Odone(i, \alpha, \top)$ means that it is obligatory (for no one in particular) to be in a state where i has just performed action α; this can be understood as "i has the obligation to perform action α". Permissions and interdictions are defined from obligations in a standard way: $P\varphi = \neg O \neg p$ means that it is permitted that φ, and $F\varphi = O\neg\varphi$ means that it is forbidden that φ.

Institutional modalities Finally this framework also provides some specific operators to formalise institutional concepts. These operators have a parameter s specifying the institution within which they are valid. Here we consider an institution as a set of facts and rules that a group of agents (the "members" of this institution) adopt. This is a general view that can account for as various institutional contexts (be they formal or informal) as the law of a country, a contract between two parties in a business relationship, a social structure, the rules of a game...

An institutional fact is a fact that is recognised to be true in the context of a given institution, but that can make no sense in itself. For example the fact that two people are married, or that one is authorised to drive a truck, is only valid in the law of a country. It is not a physically observable fact, but something written in the registry of this institution. We represent these institutional facts with the operator $D_s\varphi$ meaning that for institution s, it is officially established that φ holds. This operator represents facts that are stored in the registry of the institution.

Remark: our operator of obligation $O\varphi$ means (semantically) that φ is true in all ideal worlds. However there is no absolute ideal world, but only particular ideal worlds relative to a given institutional context (even if some generic obligations can hold in most of institutions). Therefore we will never use the obligation operator alone, but always encapsulate it in an institutional fact ($D_s O\varphi$) in order to make explicit the particular institution in which this obligation holds.

Institutional facts can be deduced from observable facts thanks to the rules of the institution. For example the presentation of an invoice by a provider to his client counts as an obligation for the client to pay for it. The existence of

the invoice is physically observable, while the obligation is only valid in an institutional context. We represent these normative consequences with the operator $p \Rightarrow_s q$, meaning that according to the norms holding in institution s, p entails q. This operator is known in the literature as *count as*, and has been formalised by Sergot and Jones [17].

A particular case of normative consequence concerns the consequences of the performance of an official procedure. Actually some agents can have the power when performing a given procedure under some conditions to create new institutional facts. We represent these institutional powers as an abbreviation[4]: $power(i, s, cond, proc, n) = (cond \wedge done(i, proc, \top)) \Rightarrow_s n$. Intuitively this means that i has the power in institution s, by performing procedure *proc* in the context where condition *cond* holds, to see to it that n becomes officially true in institution s.

3.2 Axiomatics

D_s is a normal modal operator with standard (KD) axiomatics (for more details on the formal properties of normal modal logics, see [6, chap. 4]).

The $\Rightarrow_s$ connector satisfies the substitutivity rule for equivalent formulas, plus the following axioms:

$$((p \Rightarrow_s q) \wedge (p \Rightarrow_s q')) \rightarrow (p \Rightarrow_s (q \wedge q')) \quad \text{(CC)}$$

$$((p \Rightarrow_s q) \wedge (p' \Rightarrow_s q)) \rightarrow ((p \vee p') \Rightarrow_s q) \quad \text{(CA)}$$

$$(p \Rightarrow_s q) \rightarrow ((q \Rightarrow_s r) \rightarrow (p \Rightarrow_s r)) \quad \text{(S)}$$

It is linked to D_s by the following mix axioms:

$$(p \Rightarrow_s q) \rightarrow D_s(p \rightarrow q) \quad \text{(SD)}$$

$$(p \Rightarrow_s q) \rightarrow (p \rightarrow D_s p) \quad \text{(SC)}$$

From (SD) and axiom (K) for normal modal operators we can easily prove the following theorem:

$$(p \Rightarrow_s q) \rightarrow (D_s p \rightarrow D_s q) \quad (Th1)$$

The reciprocal of this implication is false, so that *count as* cannot be trivially reduced to a material implication between institutional facts.

3.3 Commitments, ratified mental attitudes, and obligations

We would like to highlight here how this framework allows to express the public notion of ratified belief, similar to the social notions of commitments or acceptance, and discuss the difference between commitments and obligations. We

[4] We only impose that condition *cond* is true after the procedure because we make the hypothesis that the procedure does not modify the truth value of *cond*, so if it is true after, it was also true before and during the performance of the procedure (see axiom $PP_{\alpha,\varphi}$ in section 4.3).

define an agent's ratified belief as a belief that is acknowledged by (and recorded in) the institution. The abbreviation $D_s B_i \varphi$ reads "it is official in institution s that agent i believes φ", and can be understood as "agent i is committed to belief φ in institution s". This notion of recorded belief is similar to Gaudou *et al.* 's notion of grounding [13]. But we also define in the same way an agent's ratified intention as an intention that is recorded in the institution. The abbreviation $D_s I_i \varphi$ reads "it is official in institution s that agent i intends to see to it that φ", and can be understood as i's commitment to the corresponding course of action. Since the agent should conform to these ratified mental attitudes in his subsequent behaviour, this notion is similar to Lorini *et al.* 's notion of acceptance or to Fornara and Colombetti's commitments.

It is also important to notice that commitments (or ratified beliefs and intentions) are fundamentally different from (and independent of) obligations. Obligations are imposed by the institution independently of the agent's will, while an agent can only commit voluntarily, *i.e.* intentionally; indeed commitments result from the agent's actions to promise something (that a proposition is true, or to perform a given action), and we assume that agents' actions all are intentional (unlike events). Please notice that this does not mean that commitments cannot be violated, since one can revise his intentions, or fail to achieve them. There can exist some links between obligations and commitments, but they depend on the institution and on the agent. On the one hand an institution can impose rules allowing to deduce obligations from the agents' commitments or ratified mental attitudes. Such a rule could be that agents must behave in accordance with their commitments: for example in a B2B contract, if a provider promises that the price of a given item is a given amount, then he is obliged to sell this item at this price. On the other hand each agent can reason about his obligations and decide to commit himself to respect them (or to ratify his intention to respect them) or not: for example an obeying agent would systematically commit to fulfill his obligations. Besides, institutions usually specify sanctions associated with the violation of obligations, while the consequences of the violation of commitments may not be specified in advance, and may depend on their creditor. For example if the law of your country forbids you to steal, it also specifies an associated sanction, like some prison term; on the contrary if you promise a friend to loan him your car, and you finally do not, there is no predefined sanction for this, but your friend will probably get angry at you, and may prepare some revenge.

3.4 Examples of formalisation of some institutional rules

In this paragraph we give some examples of institutional rules in order to illustrate the functioning of our operators. These are not generic axioms but specific rules given as examples. We consider a particular institution that is an interchange contract between two businesses in the context of B2B exchanges. In this institution noted $b2b$ we consider that we have the following rules:

- an agent who has ordered an item must pay for it when delivered:
  $done(j, deliver(j, i, item, price), \top) \Rightarrow_{b2b} O done(i, pay(i, j, price, item), \top)$;

- the agents involved in the exchange must know the rules of the contract (they are obliged to know about their obligations): $D_{b2b}O\varphi \rightarrow D_{b2b}OB_iD_{b2b}O\varphi$. This rule cannot be assumed in all institutions, for example it may be difficult to know all the laws of a country, except for lawyers;
- agents should not contradict themselves : $D_{b2b}B_i\varphi \rightarrow D_{b2b}OB_i\varphi$. Remark: one cannot check directly the agent's belief, but the agent's behaviour and speech acts should be in accordance with this obliged belief. As a consequence, if the agent adopts a behaviour that reveals an opposite belief (for example asserting the contrary of the proposition he is obliged to believe) he is violating this obligation. Besides please notice that this strong link between commitments and obligations cannot be assumed in all institutions.

4 Model of institutional interpretation of actions

In previous work [1] we provided a first preliminary formalisation of the concept of institutional interpretation of actions, focused on its application. In this paper we focus on the theoretical aspects and provide an improved model of this concept, in particular with a finer account of institutional preconditions.

4.1 Informal definition

Actions, be they communicative or material, have both an intentional dimension (why the agent wants to perform this action, which goal he is aiming at, under which conditions this action is feasible) and an institutional dimension (how this action is interpreted in the context of an institution, does it create new obligations). The intentional dimension of communicative actions is formalised in agent communication language semantics like FIPA [11]. The intentional dimension of material actions is studied by planning theories reasoning on their preconditions and effects. On the contrary the institutional dimension, which depends on each specific institution, has been little studied. Only little work provide an institutional semantics for some speech acts [12, 8, 18], and as far as we know there exist no work providing an institutional characterisation of material actions.

In this paper we intend to combine the intentional and institutional dimensions of actions, be they communicative or material, in a single framework. Thereby we have extracted the representative features of the institutional interpretation of an action.

4.2 Features of the institutional interpretation of an action

The same action can be interpreted differently depending on the institutional framework. Thus the features of the institutional interpretation of an action depend on the institution. For example in the context of B2B exchanges, the reception of a purchase order generally obliges the provider to confirm the order before delivering it; however in some particular contracts, notably when the provider and client are used to work together, the sending of the purchase order

can implicitly count as a client's commitment to pay for it, allowing the provider to directly dispatch the items without confirmation. The obligations created by the action of sending a purchase order thus differ from one institution to another (here an institution is a contract of *interchange* between a client and a provider).

Makinson [19] distinguishes between the legal power, the physical power to perform the procedure of this legal power, and the permission to perform this procedure. He assumes that "a capacity or power to perform a legal act is not quite the same thing as a permission to do so" [19, p.408]. He gives the example of a priest who has the power of marrying people, but can use this power while he is not allowed to do so, for instance for couples not being of the corresponding religion; in this case, this still counts as a valid marriage but the priest risks sanctions from his church. Following Makinson's work, that grounds on jurisprudential theory, we characterize institutional actions with two different conditions: a permission precondition, and a power precondition, that are detailed below.

The first feature of the institutional interpretation of an action is its **permission precondition**, that gives the permission to perform the action (different from its material feasibility precondition). For example to take some object in a shop, a material feasibility precondition is to be able to carry the object; however the permission precondition is to have paid this object: it is forbidden by the institution (here, the law of the country) to take the object without paying for it, and paying for it is enough to get the permission to take it. So the permission precondition is necessary and sufficient to have the permission to perform the action in the considered institutional context.

The second feature of the institutional interpretation of an action is its **institutional effect**, that is the institutional facts that the execution of this action can create in the institution. This institutional effect thus differs from the physical or rational effect of an action. For example a mayor has the power to pronounce a wedding, thereby creating a new institutional fact (that the people are now married). At the same time this action has the physical effect (or rational effect in Sadek's rational interaction theory [22], since this is a communicative action) of making the witnesses believe that the two people are now married.

This institutional effect is associated with an additional condition, the **power precondition**, that must be validated for that effect to be deduced. Actually this is similar to an institutional power whose procedure is this action, whose effect is the institutional effect of the action, and whose condition is the power precondition of the action. In fact there can be several pairs associating a power condition and an institutional effect, if the considered action triggers several powers. For example in a B2B context, we can consider the following rules for a contract of interchange between the two parties: when a client orders an item, if he refers to a valid catalogue then the provider is obliged to process his order; moreover if he has ordered items for a total amount that overtakes some threshold, he is obliged to pay before delivery. Thus there are two obligations (one for the client, one for the provider) that can be deduced from the performance of

the action to send a purchase order, depending on the preconditions that are valid.

Finally, sanctions can be associated with the forbidden performance of this action. For example the action of taking an object in a shop without having payed for it makes its author liable for prosecution and sanctions, like fines or prison sentences, depending on the value of the stolen object. These sanctions are specified in the law in effect in the country, that constitutes a particular institution.

4.3 Formal model of the institutional interpretation of actions

An action α is characterised by the following attributes determining its institutional interpretation:

- an institution s in which these features are valid;
- its precondition of permission: a formula φ that is necessary and sufficient to allow the performance of α;
- a sanction in case of forbidden performance of the action: an institutional fact χ that becomes true after a forbidden performance of α;
- a list of pairs of power precondition and associated explicit institutional effect: pairs of formulas $< \psi_i, \omega_i >$.

In the following subsections we detail these features and the axioms representing their role.

Permission precondition, implicit effect and sanctions The permission precondition is the condition that is necessary and sufficient in order to be permitted to perform the action. For example one has the permission to drive a car if and only if he has a valid driving license.

This equivalence between the permission precondition and the permission to perform the action is expressed by the following axiom, valid for any agent i and action α, where φ is the permission precondition of α:

$$\varphi \leftrightarrow D_s Phappens(i, \alpha, \top) \qquad (PPA_{\alpha,\varphi})$$

We also assume that by performing the action α, agent i commits himself to respecting its permission precondition. We call this commitment the implicit effect of action α in institution s:

$$done(i, \alpha, \top) \Rightarrow_s B_i \varphi \qquad (IIE_{\alpha,\varphi})$$

Finally the performance of action α without validating its permission precondition counts as the sanction specified in the attributes of this action. We note χ the proposition representing this sanction and thus have the following theorem:

$$done(i, \alpha, \neg\varphi) \Rightarrow_s \chi \qquad (UES_{\alpha,\varphi,\chi})$$

Actually, this may seem quite restrictive since sanctions should follow from actions of other agents like judges, but this is a simplified view.

Persistence hypothesis We assume that the performance of the action does not modify the truth value of its permission precondition, so that we have the two following equivalences:

$$done(i, \alpha, \varphi) \leftrightarrow (done(i, \alpha, \top) \wedge \varphi) \qquad (PP_{\alpha,\varphi})$$
$$done(i, \alpha, \neg\varphi) \leftrightarrow (done(i, \alpha, \top) \wedge \neg\varphi) \qquad (PP_{\alpha,\neg\varphi})$$

Power preconditions and explicit effects Each power precondition is not necessary for the action to be permitted, but it is necessary for the action to have a given institutional effect. The institutional effect is the effect of the action in the institution, in terms of new institutional facts. The created institutional facts can be normative facts (obligations, permissions...), or simple institutional facts like "the auctions are opened" or "i and j are married". Actually these simple facts may entail some new normative facts depending on the rules of the institution, for example married people have some new rights and new obligations associated with their status.

The institutional interpretation of an action is thus characterised by a set of pairs $< \psi_i, \omega_i >$ associating a power precondition ψ_i and an institutional effect ω_i. Their interpretation is similar to the notion of institutional power, and actually each pair $< \psi_i, \omega_i >$ is characterised by the following axiom:

$$power(i, s, \psi_i, \alpha, \omega_i) \qquad (EIE_{\alpha,\psi_i,\omega_i})$$

Remark We can notice that each action is also characterised by one special power, specifying the effect of its forbidden performance. Indeed from axiom $(UES_{\alpha,\varphi,\chi})$ we have: $done(i, \alpha, \neg\varphi) \Rightarrow_s \chi$. The persistence hypothesis on the negation of precondition $(PP_{\alpha,\neg\varphi})$, combined with the substitutivity rule for equivalent formulas that is valid for $\Rightarrow_s$, allows to deduce that: $done(i, \alpha, \top) \wedge \neg\varphi \Rightarrow_s \chi$. This is our definition of power, so we finally have:

$$power(i, s, \neg\varphi, \alpha, \chi)$$

This is actually similar to the rule $(EIE_{\alpha,\neg\varphi,\chi})$ for the specific pair $< \neg\varphi, \chi >$ associating the negation of the permission precondition with the sanction specified for forbidden performance. This means that each agent has the power to establish the sanction if he performs an action while he was not allowed to.

Theorem This axiom also allows to deduce an interesting theorem. Let's consider a pair $< \psi, \omega >$ characterising a power associated to α. From axiom (SC) and theorem $(Th1)$ (page 7) and the definition of *power*, we deduce $done(i, \alpha, \top) \wedge \psi \rightarrow D_s\omega$. Since $after_{i:\alpha}$ is a normal modal operator, we have the necessitation rule so we deduce the following theorem:

$$after_{i:\alpha}(\psi \rightarrow D_s\omega) \qquad \text{(Th2)}$$

4.4 Conclusion about our logical framework

Our formal framework is expressive enough to express the intentional dimension of actions (thanks to mental attitudes modalities like beliefs and intentions). Moreover we have added a characterisation of the institutional dimension of actions (both material and communicative) in terms of institutional facts and powers. This makes it a generic formalism for the interpretation of actions.

5 Illustration: formalisation of some actions

In order to illustrate the expressiveness of our model of the institutional dimension of actions, we give in this section the institutional characterisation of two actions: a material action from a practical application, and a speech act with institutional effects.

5.1 A material action: send a purchase order

We are involved in a project at Orange Labs aiming at developing a multi-agent mediation platform for automatic B2B exchanges [1]. The aim of this paper is not to give details about this practical application but to illustrate how we use our model of the institutional interpretation of actions in this project. Thus the following example is voluntarily simplified to illustrate the features of the model, and may not reflect exactly the real features of the action in actual B2B contracts of interchanges.

Let's take the example of the action of sending a purchase order: $\alpha = sendOrder(c, p, id)$ reads "client c sends a purchase order identified by id to provider p". We will give its institutional interpretation in institution $b2b$ representing a contract of interchange between the two involved businesses, the provider p and the client c.

To be permitted to send a purchase order to the provider, the client must have his catalogue. So the permission precondition is $\varphi = haveCatalogue(c, p)$. We can assume that a sanction is associated to the violation of this condition, for example the obligation to pay some damages: $\chi = O\,done(c, pay(c, p, 100), \top)$. We thus have the following institutional rules:

- $haveCatalogue(c, p) \leftrightarrow D_{b2b} P\,done(c, sendOrder(c, p, id), \top)$ reads "it is permitted that c sends an order to p if and only if he has p's catalogue";
- $done(c, sendOrder(c, p, id), \top) \Rightarrow_{b2b} B_c haveCatalogue(c, p)$ reads "if c has sent order id to provider p, then he is officially committed in institution $b2b$ to have p's catalogue" (and should behave accordingly, for example he cannot pretend to ignore the price of an item, or order it at a wrong price);
- $done(c, sendOrder(c, p, id), \neg haveCatalogue(c, p)) \Rightarrow_{b2b} O\,done(c, pay(c, p, 100), \top)$ reads "if c has sent order id to p while he did not have p's catalogue, it becomes official in $b2b$ that he is obliged to pay \$100 of damages to p".

Now by sending a purchase order to the provider, the client has the power to oblige him to process his order, on condition that the purchase order is correctly filled. This action thus has as an institutional effect $\omega = O done(p, processOrder(p, c, id), \top)$, associated with the power precondition $\psi = isCorrect(id)$. This corresponds to the following institutional rule:

$$power(c, b2b, isCorrect(id), sendOrder(c, p, id), O done(p, processOrder(p, c, id), \top))$$

We do not consider other institutional effects here, so the previous rules fully characterise the institutional interpretation of action $sendOrder(c, p, id)$ in institution $b2b$.

5.2 A communicative action: declare

In this subsection we illustrate our model by providing an institutional semantics for the declarative speech act $Declare(i, j, s, cond, n)$ reading "i declares to j in the setting of institution s that given condition $cond$, the fact n is now established".

This institutional dimension is clearly separated from its intentional dimension, that could be characterised in a FIPA compatible way by the following features, similar to those of the FIPA Inform act with $D_s n$ as its propositional content:

- Feasibility precondition FP $= \neg B_i D_s n$
- Rational effect RE $= B_j D_s n$

Please notice that the expected public effect $D_s n$ is actually not part of the rational effect; it is an institutional fact and it is thus rather part of the institutional effect, that is detailed below.

Regarding the institutional dimension, the permission precondition of a declaration is to be empowered to declare the concerned fact. So we have the permission precondition

$$PP(Declare(i, j, s, cond, n)) = power(i, s, cond, Declare(i, j, s, cond, n), n)$$

Such a power may be deduced from the agent's particular role in the institution, and depends on the content of the declaration. For example a reverend is empowered to declare weddings, while an auctioneer is empowered to declare the opening of an auction sale. This explicit empowerment is necessary to prevent any agent from declaring new institutional facts just by declaring that they are true with any trivial condition that is not attested by the institution as really entailing the new fact.

No general sanction is attached to this permission precondition, since it depends on the institution, the content of the declare, the role of the speaker... so possible sanctions would be determined by specific rules of the institution. For example a false mayor who declares a wedding while he is not empowered to do so could be pursued, while a child declaring that it is recreation time while

it is not his role but the professor's one would only risk to get some additional homework or to be deprived of recreation.

The implicit effect of a declaration is that the speaker is committed to having this power, *i.e.* this belief is ratified by the institution, that is to say recorded in the register of institutional facts.

The only explicit effect of a declaration is to establish the declared fact, and the additional fact that the hearer is aware of this new fact, but this is constrained by the condition of the power. So the list of pairs characterising the explicit effects of a declaration is the following:

$$EE(Declare(i, j, s, cond, n)) = \{< cond, n \wedge B_j D_s n >\}$$

6 Conclusion

We have provided here a unified approach for the formalisation of both intentional and institutional dimensions of communicative and material actions. This work is original since there exist few work unifying the formalisation of the institutional and intentional aspects of actions, and as far as we know all of them are restricted to communicative actions.

This formal framework has been implemented in JADE agents used to design a multi-agent mediation platform for automatic B2B exchanges [1, in French]. We have only presented a small illustration example in this paper that was focused on the theoretical aspects, but we hope to be able to demonstrate our prototype at the workshop.

In future works we intend to use this generic framework to extend the FIPA semantics with an institutional semantics for its speech acts. Therefore we will use the features that we have identified here as characterising the institutional interpretation of actions, and apply them to the particular case of communicative actions.

References

1. C. Adam, F. Bourge, V. Louis, and S. Picant. Un modèle d'actions institutionnelles pour un système multi-agents appliqué à la médiation d'échanges B2B. In *JFSMA*, 2008.
2. G. Boella, R. Damiano, J. Hulstijn, and L. van der Torre. Role-based semantics for agent communication: embedding of the 'mental attitudes' and 'social commitments' semantics. In *AAMAS'06*, 2006.
3. J. Broersen, M. Dastani, and L. van der Torre. Beliefs, obligations, intentions, and desires as components in an agent architecture. *International Journal of Intelligent Systems*, 20(9):893–919, 2005.
4. J.P. Burgess. *Handbook of philosophical logic*, volume 7, chapter Basic tense logic, pages 1–42. Kluwer Academic Publishers, 2nd edition, 2002.
5. C. Castelfranchi. Commitments : From individual intentions to groups and organizations. In *ICMAS-95*, pages 41–48, San Francisco, 1995.
6. B.F. Chellas. *Modal Logic: an introduction.* Cambridge University Press, 1980.

7. R. Demolombe and V. Louis. Norms, institutional power and roles: towards a logical framework. In *16th International Symposium on Methodologies for Intelligent Systems (ISMIS'06)*, volume LNAI 4203, pages 514–523. Springer, 2006.
8. R. Demolombe and V. Louis. Speech acts with institutional effects in agent societies. In *DEON'06*, 2006.
9. F. Dignum, D. Morley, E.A. Sonenberg, and L. Cavedon. Towards socially sophisticated bdi agents. In *ICMAS'2000*, pages 111–118, 2000.
10. F. Dignum and H. Weigand. Communication and deontic logic. In R. Wieringa and R. Feenstra, editors, *Information Systems, correctness and reusability*, pages 242–260, Singapore, 1995. World Scientific.
11. FIPA. The foundation for intelligent physical agents. http://www.fipa.org.
12. N. Fornara and M. Colombetti. A commitment-based approach to agent communication. *Applied Artificial Intelligence*, 18(9-10):853–866, 2004.
13. B. Gaudou, D. Longin, and A. Herzig. A logical framework for grounding-based dialogue analysis. In W. van der Hoek, A. Lomuscio, E. de Vink, and M. Wooldridge, editors, *LCMAS'05*, 2005.
14. B. Gaudou, D. Longin, E. Lorini, and L. Tummolini. Anchoring institutions in agents' attitudes: towards a loçgical framework for autonomous multi-agent systems. In *AAMAS'08*, 2008.
15. P. Hakli. Group beliefs and the distinction between belief and acceptance. *Cognitive Systems Research*, 7:286–297, 2006.
16. A. Jones and J. Carmo. *Handbook of philosophical logic*, volume 8, chapter Deontic Logic and Contrary-to-duties, pages 265–343. Kluwer Academic Publishers, 2nd edition, 2002.
17. A.J. Jones and M. Sergot. A formal characterisation of institutionalised power. *Journal of the interest group in pure and applied logics*, 4(3), 1996.
18. E. Lorini, D. Longin, and B. Gaudou. The institutional dimension of speech acts: a logical approach based on the concept of acceptance. Research report IRIT/RR-2008-9–FR, IRIT, 2008.
19. D. Makinson. On the formal representation of rights relations: remarks on the work of stig kanger and lars lindahl. *The Journal of Philosophical Logic*, 15:403–425, 1986.
20. P. Pasquier, R. Flores, and B. Chaib-draa. The enforcement of flexible social commitments. In F. Zambonelli M.-P. Gleizes, A. Omicini, editor, *ESAW'05*, volume LNAI 3451, pages 139–152. Springer-Verlag, 2005.
21. A.S. Rao and M.P. Georgeff. Modeling rational agents within a BDI-architecture. In *KR'91*, 1991.
22. D. Sadek. A study in the logic of intention. In *KR'92*, 1992.
23. M. Wooldridge. *Reasoning about rational agents*. MIT Press, 2000.