The Role of Institutions in Multiagent Systems

Marco Colombetti[1,2], Nicoletta Fornara[2], and Mario Verdicchio[1]

[1]Politecnico di Milano, Dipartimento di Elettronica e Informazione
Piazza Leonardo da Vinci 32, 20133 Milano, Italia

[2]University of Lugano, Institute of Communication Technologies
Via Buffi 13, 6904 Lugano, Switzerland

Abstract. The development of open multiagent systems presupposes not only a standard Agent Communication Language and common conversation protocols, but also some kind of social framework within which agent interactions can be carried out. To account for such a social framework we propose a suitable construct, that we call an institution, imposing role-dependent authorizations and interaction rules on groups of agents. In particular, we introduce a Core Institution to set the conditions for the manipulation of social commitments, which we take as the fundamental notion for modeling communicative interactions in a multiagent system. Within the Core Institution we define a basic library of communicative acts. We then suggest that institutions can be employed to model conversational rules and special interaction frameworks, like those that are necessary to develop e-business applications.

1 Introduction

The development of genuinely open multiagent systems is one of the main goals in the field of autonomous agents. As part of this enterprise, many efforts have been and are devoted to the definition of a standard Agent Communication Language (ACL) and of standard conversation protocols. So far, two ACLs have been widely discussed in the literature: KQML [8] and FIPA ACL [9]. Defining a standard ACL is a demanding task, because it is not just a matter of choosing an appropriate syntax, an efficient coding system and a collection of suitable terminologies. As communication is a complex process involving individual and social aspects of the communicators' "life," an ACL must be embedded in a rich conceptual framework, able to account for all relevant aspects of agent interaction. To put it in different words, an ACL has to be co-designed with some form of agent society. We still do not know exactly what agent societies are going to be like. We do know, however, that we want agents to act as representatives of human individuals and organizations in a variety of scenarios, including information exchanges and e business transactions. Many computer scientists therefore assume, either explicitly or implicitly, that an agent society will reflect at least some aspects of human organization.

If this standpoint is correct, it follows that an ACL can only be developed within a conceptual framework that accounts for the basic elements of social structure – at least those

elements that are likely to be essential in the world of artificial agents. In this paper, we report about our ongoing work in this direction. In particular, in Section 2 we suggest that communicative acts should be regarded as institutional actions. In Section 3 we define commitment as the main institutional concept on which to base a treatment of communicative acts. In Section 4 we introduce the Core Institution (i.e., the system of social rules that allows agents to perform general purpose, application independent communicative acts), and then specify a library of Core Communicative Acts. In Section 5 we introduce special institutions, and show how these both regulate the use of Core Communicative Acts and allow for new types of communicative acts; we then propose a general structure for institutions and suggest how institutions can be actually represented. In Section 6 we briefly discuss some related work. Finally, Section 7 is devoted to some concluding remarks.

2 Communicative acts as institutional actions

We view communicative acts as a kind of institutional actions, that is, actions that are possible on the basis of a set of conventions and regulations, and whose effect is to bring about an institutional effect [4]. More precisely, a communicative act is an action realized as follows. An agent, the sender, sends a message to another agent, the receiver. By doing so, if certain contextual conditions hold, the sender performs a communicative act addressed to the receiver. The effect of such a performance is to bring about a number of institutional facts. The term ”institutional fact” is to be understood in opposition to ”natural fact”. A natural fact is a state of affairs that, in order to hold, does not presuppose any human convention: think for instance of the position in space of a physical object, or of the air temperature in a room. Of course, human conventions may be involved in the representation of a natural fact (think of the Fahrenheit and Celsius scales for temperature); however, the fact itself holds independently of such conventions. On the contrary, institutional facts can hold only thanks to systems of human conventions. As an example, consider the conventions and regulations concerning the institution of property: that an agent owns an object is not a natural, human independent fact. Our first working hypothesis is that all communicative acts can be regarded as actions by which agents manipulate institutional facts. More precisely, we suggest that a communicative act brings about an institutional fact when it is performed in a given context by an authorized agent. Authorizations, in turn, are in virtue of the role played by the agent in the relevant context. Given that an ACL is the tool by which agents perform communicative acts, designing an ACL presupposes an analysis of institutions in terms of such concepts as roles and authorizations. Companies, universities, marriage and property are examples of institutions within which human beings interact every day. Some, but not all, of these institutions will be reflected in agent societies. In e-business applications, for example, the institution of property obviously takes a central position. Our second working hypothesis is that there is one basic, primitive institution, which we call the Core Institution, which is always part of the context of interaction, and is exploited by all other institutions. The Core Institution regulates the most fundamental type of institutional facts, that is, (social) commitments.

3 Commitment

Recently, *commitment* has been identified by several authors as a fundamental notion for modeling agent interactions in multiagent systems. In several scientific papers [17, 18, 3, 21, 4, 11]

commitment has been adopted as the key concept for defining communicative acts. We conceive of commitments as "institutional objects," that is, as abstract objects that can be created and manipulated according to a set of conventional rules, on which there is collective agreement by a community of agents. We believe that commitment, also in the case of human interactions, is a primitive social concept. In particular, commitment cannot be defined in terms of agent mental states. The function of commitment is to stabilize social interactions, by making the behavior of agents predictable to their partners at least to some extent. The future behavior of autonomous, utility-driven agents representing different interests cannot be completely determined from the public, observable component of an interaction. However, commitments constrain future behavior in that the social management of commitments makes it rational for an agent to fulfill them, at least in normal situations. We take the notion of *conditional commitment* as the basic primitive notion of our treatment. A (conditional) commitment has five main components:

Debtor the agent that has the commitment.

Creditor the agent relative to which a commitment is made.

Condition a state of affairs that "activates" the commitment if it becomes true within a given *timeout*.

Content a state of affairs to which the debtor is committed relative to the creditor. It may have an associated *deadline* within which it ought to become true.

State A commitment can be in one of six possible states: *unset*, *cancelled*, *pending*, *active*, *fulfilled*, and *violated*. A commitment whose state is unset is also called a *precommitment*.

A commitment with state s, debtor x, creditor y, condition ϕ, and content ψ is described by the statement $C(s, x, y, \phi, \psi)$. A commitment can be created and manipulated by a set of *basic operations*. Its state evolves also as an effect of *environmental events*. A detailed description of the effects of operations and events on commitments is given in [11]. Here we sketch the essential aspects of our model.

Intuitively, the six different states of a commitment correspond to the following situations:

unset the commitment has been proposed (by its debtor, its creditor, or a third party) but it has not yet been accepted nor refused;

cancelled the commitment has been withdrawn by its creditor or refused by its debtor, or the timeout of the commitment has expired and the condition has not become true;

pending the commitment has been accepted but its condition has not yet become true;

active the commitment's condition has become true before the associated timeout;

fulfilled the content of an active commitment has become true before the associated deadline;

violated the deadline of an active commitment has expired and the content has not become true.

Commitments are created and manipulated through the following basic operations:

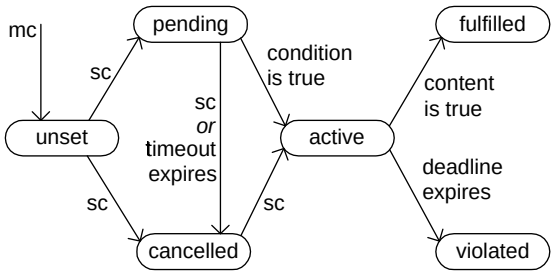

Figure 1: The lifecycle of commitments.

Make commitment The operation $mc(x, y, \phi, \psi)$ creates an unset commitment (with debtor x, creditor y, condition ϕ, and content ψ), and returns a reference to the representation of the commitment it creates.

Set commitment The operation $sc(c, s)$ sets to state s the state of the commitment referred to by c.

After a commitment is created, its state typically evolves, due to further operation executed on it and to external events. The lifecycle of a commitment is shown in Figure 1 (see [11] for further details).

Basic operations should not be viewed as actions to be directly performed by agents. Rather, they are low-level primitives used to implement commitment manipulation. Agents can operate on commitments only through a library of communicative acts, in particular through those acts that are defined within the Core Institution (see next section).

4 The Core Institution

4.1 Institutional Rules

The Core Institution sets the fundamental conditions for commitment manipulation. Such conditions can be expressed as a set of authorization rules:

Rule CI-1 any agent is authorized to create an unset commitment with arbitrary debtor, creditor, condition, and content.

Rule CI-2 the debtor of an unset commitment is authorized to set it to either *cancelled* or *pending*.

Rule CI-3 the creditor of an unset, pending or active commitment is authorized to set it to *cancelled*.

It is feasible to assume that more rules will have to be added when operations for commitment negotiations are considered.

4.2 The Core Communicative Act Library

It is now possible to define a library of communicative acts, the Core Communicative Act Library, that can be performed by using the basic commitment manipulation operations in

such a way that the authorization rules of the Core Institution are obeyed. Below we define a number of Core Communicative Acts. In our definitions we use the following metalinguistic conventions:

- a statement of the form $Do(x, \alpha)$ means that agent x performs action α;
- an expression of the form $\{op_1; ...; op_n\}$ means that the basic operations op_1,...,op_n are executed in sequence;
- the '$=_{def}$' sign means that performing the action represented on the left-hand side is the same as performing the action or executing the operations represented on the right-hand side;
- a statement of the form '$c \leftarrow op$' means that operation op returns a reference to a commitment which is assigned to variable c;
- a statement of the form '$c \leftarrow C(s, a, b, \phi, \psi)$' means that the reference to an existing commitment of the form $C(s, a, b, \phi, \psi)$ is assigned to variable c.

Inform

$Do(x, inform(y, \phi)) =_{def} \{c \leftarrow mc(x, y, true, \phi); sc(c, pending)\}$.

Performing an act of informing amounts to making the unconditional commitment that the content ϕ of the inform act is true, and setting it to *pending*. The former operation is authorized by Rule CI-1; the latter, by Rule CI-2.

Request

A (conditional) request is the request to perform an action α if a given condition ϕ obtains. It amounts to creating an unset commitment:
$Do(x, request(y, \phi, \alpha)) =_{def} \{mc(y, x, \phi, Do(y, \alpha))\}$.

QueryRef

This is the request of being informed about the value of a term τ, possibly a definite description:
$Do(x, queryRef(y, \tau)) =_{def} Do(x, request(y, const = \tau, inform(x, const = \tau)))$.

Here $const$ must be understood as a metavariable standing for an arbitrary constant. The definition says that the addressee is requested to inform the speaker that $const = \tau$, for some constant $const$, under the condition that actually $const = \tau$.

QueryIf

This is the request to be informed about the truth value of a statement:
$Do(x, queryIf(y, \phi)) =_{def} Do(x, queryRef(y, truthVal(\phi)))$.

A functional term of the form $truthVal(\phi)$, where ϕ is a statement, denotes the current truth value of ϕ.

Accept

Accepting an unset commitment means setting its state to *pending*:

$Do(x, accept(y, \phi, Do(x, \alpha))) =_{def} \{c \leftarrow C(unset, x, y, \phi, Do(x, \alpha)); sc(c, pending)\}$.
This action is made possible by Rule CI-2, and succeeds only if a commitment of the form $C(unset, y, x, \phi, Do(y, \alpha))$ currently exists.

Refuse

Refusing an unset commitment means setting its state to *cancelled*:
$Do(x, refuse(y, \phi, Do(x, \alpha))) =_{def} \{c \leftarrow C(unset, x, y, \phi, Do(x, \alpha)); sc(c, cancelled)\}$.
This action is made possible by Rule CI-3, and succeeds only if a commitment of the form $C(unset, x, y, \phi, Do(x, \alpha))$ currently exists.

5 Special Institution

The Core Institution is sufficient to account for those communicative interactions in which all commitments are made by the agents through the autonomous performance of communicative acts. In many interesting situations, however, agents interact in a social context that pre-imposes a number of further authorizations and duties. For example, in an electronic auction the auctioneer is obliged, just because it is the auctioneer, to accept the clients' bids and to assign the products according to the auction's rules. On the other hand, the auctioneer is also authorized to perform a number of auction-related actions (like opening and closing an auction, assigning a product to a client, and so on) from which the clients are precluded. Our hypothesis is that all this is possible just because auctions are a special kind of institution, with own rules that establish specific authorizations and duties for the agents playing different roles.

5.1 Conversations

The simplest kind of special institution is a *conversation*. In fact, the Core Institution alone is not sufficient to guarantee that the ordinary rules of conversations are followed. Consider for example an agent, a, making to agent b a request that b is not willing to accept. According to the rules of the Core Institution, agent b *may* explicitly refuse the request. However, no rule *binds* b to do so: b might as well keep silent, thus leaving a uncertain about the outcome of its request. In a human interaction, failing to react to a request would be considered as breaking the ordinary rules of conversation. We believe that a similar idea should apply to agents.

An Institution of Conversations might be defined as follows. A conversation is opened by an agent, when it addresses to one or more other agents the invitation to join it in a new conversation. The initiating agent, together with the agents that accept its proposal, become the *partners* of the new conversation. At any moment, a partner is authorized leave the conversation; by doing so, it implicitly cancels all the unset commitments that have been made in the conversation and of which the agent is either the debtor or the creditor. The conversation is implicitly closed when the last partner leaves it. Throughout a conversation, every partner may perform any communicative act authorized by the Core Institution, but is subject to the following additional rule:

Rule IC-1 every partner that is inserted as the debtor in an unset commitment, whose creditor is another partner, must explicitly accept or refuse the commitment.

This rule states a duty that all partners implicitly accept when they join a conversation. The treatment sketched above is just one of many possible definitions of a conversation. For example, one may want to allow any agent to join a conversation at any moment, even without an invitation; or give special authorizations to the initiating agent (in this case, the role of the *initiator* will be distinguished by the role of all other partners). In any case, the point is that any definition of conversation will introduce a special institution that can be exploited by agents in their interactions.

E-business environments, like electronic marketplaces, appear to be specific cases of special institutions. For example, an English auction may be viewed as a special kind of conversation, in which additional authorizations and duties are enforced. As developed so far, however, the notion of an institution is an abstract concept; we now need to specify how institutions can be formally represented in computational systems. An interesting proposal in this direction is put forward by Esteva and colleagues [7]. We think, however, that the notion of an electronic institution proposed in [7] is too complex and detailed to serve the purpose of a general building block of agent societies. In the rest of this section we present an alternative treatment, which we regard as a suitable component of an agent interaction framework.

5.2 The Structure of Institutions

We think of an institution as made up by four components: a set of *registration rules*, a set of *interaction rules*, a set of *authorizations*, and an *internal ontology*.

The registration rules define the procedure for agent registration. When an agent applies for registration to an institution and the procedure is carried out successfully, the agent is assigned a *role* in the institution. By the term *group* we refer to the collection of agents filling the roles of an institution at any particular moment.

After registration, an agent may interact with the other members of the group according to role-dependent interaction rules. These rules can be viewed as a set of restrictions (i.e., permissions and obligations) on the execution of the actions for which the agent has the necessary authorizations. For example, agent conversation protocols may be regarded as the interaction rules of an Institution of Conversations.

Which actions may be performed by an agent, depending on its role in the institution, is specified by the set of authorizations. For example, the authorizations of the Core Institution specify which communicative act can be performed in general by an agent.

Finally, the internal ontology accounts for the institutional facts and events that are typical of the institution, and therefore defines the new ”reality” that is introduced by the institution itself. The internal ontology thus provides for the conceptual framework necessary to define the actions that may be performed within the institution. For example, the internal ontology of the Core Institution is the ontology of commitment, which allows for the definition of communicative acts.

Given that we are presenting an abstract model, we make no assumption on how the various components of an institution are realized. In general, however, we expect that such rules will be implicitly enforced, rather than explicitly represented. Here is a possible way of doing so:

- Registration and interaction rules can be implemented by a collection of *protocols*. Following the protocols ensures that the institutional rules are obeyed.

- Authorizations can be implemented by a *library* of role dependent actions that can be performed by the members of the group. The library is designed in such a way that if an action is available to an agent playing a given role, the authorizations associated to that role are implicitly guaranteed.

A specific institution need not implement every component of the general structure. For example, the Core Institution can be described as follows:

Roles and registration rules There is only one role, which we call *speaker*, and no registration rules, because every agent is allowed to act as the speaker of a communicative act.

Interaction rules There are no interaction rules.

Authorization These are stated by Rules CI-1, CI-2 and CI-3 of Subsection 4.1. These authorizations may be implemented through a Communicative Act Library like the one sketched in Subsection 4.2.

Internal ontology The internal ontology is the ontology of commitment, and may be implemented in terms of a representation of commitments together with a set of basic operations on such a representation.

Another simple example of an institution is given by the Institution of Conversations, with a number of roles for the partners, and a library of conversation protocols that define the duties of the partners. Conversations are carried out by performing communicative acts. This means that an agent playing the role of a partner in a conversation also plays the role of the speaker in the Core Institution. This relation between the two roles is a case of role subsumption, a notion that is clarified in the next subsection.

5.3 Role Subsumption

Often, to play a role in an institution an agent has also to play some role in another institution. Relationships between roles (within a single institution) have been already analyzed in the literature, in particular by Esteva and colleagues [7], who introduced the term *role subsumption*. We define role subsumption to be a static relationship between roles of different institutions, where by "static" we mean that the relationship is part of the definition of the institution. If role r_1 of institution I_1 subsumes role r_2 of institution I_2, when an agent is registered in I_1 to play r_1 it is also registered in I_2 to play r_2. When a role in institution I_1 subsumes a role in institution I_2, the internal ontology and interaction rules of I_2 are imported into I_1. By doing so, a conflict may arise between local and imported authorizations and duties. We plan to investigate how such conflicts can be managed in the near future. Note that, in general, the role of speaker in the Core Institution is subsumed by all roles that require agents to interact via ACL messages.

6 Related Work

In the last few years, many researchers have become aware of the importance of the social dimension of agent interaction. As a first result, semantics of ACLs based on mental states

[1, 2, 14, 9] have been criticized, and alternative approaches have been explored. In particular some researchers [17, 4] have proposed to base the semantics of ACLs on social commitment, a concept that has been already exploited by Searle [15] in his speech act theory, then recognized as central by Winograd and Flores [20] and later discussed, among others, by Conte and Castelfranchi [5].

The approaches based on commitment have the notable advantage of providing a simple and neat treatment of concepts like negotiation, agreement and contract, which are going to play a fundamental role in agent interactions. The importance of these concepts has been explicitly recognized by FIPA in the Request for Information issued in October 2001 by the Technical Committee for Semantics [10]. Examples of how a commitment-based ACL can be used to negotiate contracts are presented, for example, by Dellarocas [6].

Within the context of commitment-based semantics, this paper mainly deals with the concept of institution. As we have already remarked, the idea of an electronic institution has been already put forward by Esteva and colleagues [7]. However, in our definition of an institution we have been mainly influenced by Searle's analysis of the "counts as" relationship [16] and by Jones and Sergot's model of institutionalized power [12]. In fact, what we call an "authorization" can be viewed as a form of institutionalized power.

Recently, Andrew Jones has proposed a comprehensive model of conventional signalling acts, to be used as a basis for the definition of an ACL [13]. Jones's model, developed within the European project ALFABIITE, provides for an alternative to mentalistic ACL semantics that is not based on the concept of commitment. We plan to analyze the relationships between our approach and Jones's proposal in the near future.

7 Conclusions

The research proposal outlined in this paper starts from recognizing that communication is an inherently social activity. Therefore, as we already said in the Introduction, an ACL can only be developed in the context of a sufficiently rich agent interaction framework, to be used as the basis of an agent society.

It is tempting to view our proposal, like all similar efforts, as an effort to endow agents with some form of "social intelligence." We believe, however, that agent research should not become a slave of its own metaphors. In particular, we would like to stress that our proposal makes no assumption whatever on the intelligence of individual agents. Indeed, we regard an agent society not as a community of intelligent artificial systems, but as an artifact by which human beings try to automate part of their rational social interactions.

As it often happens in the most advanced areas of Computer Science, research is trying to anticipate applications. The main risk associated with this style of work is that problems are tackled in such a way that solutions cannot be easily transferred to real applications. In our work, we try to avoid this pitfall. In particular, we believe that agent models ought to be formulated with a special focus on their operational aspects. We hope that in this way ideas can be easily transferred to the world of software practice in order to create effective prototypes in reasonable time.

References

[1] Cohen, P.R., and H.J. Levesque, Rational interaction as the basis for communication, in P.R. Cohen, J. Morgan and M.E. Pollack, eds., Intentions in communication, MIT Press, Cambridge, MA, (1990) 221–

256.

[2] Cohen, P.R., and H.J. Levesque, Communicative actions for artificial agents, Proceedings of the International Conference on Multi-Agent Systems, AAAI Press, San Francisco (1995).

[3] Colombetti, M., A commitment-based approach to agent speech acts and conversations, Proceedings of the Workshop on Agent Languages and Conversation Policies (Autonomous Agents 2000), Barcelona, (2000) 21–29.

[4] Colombetti, M., and M. Verdicchio, An analysis of agent speech acts as institutional actions, Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2002), Bologna (2002), to appear.

[5] Conte, R., and C. Castelfranchi, Cognitive and social action, UCL Press, London (1995).

[6] Dellarocas, C., Contractual agent societies: Negotiated shared context and social control in open multi-agent systems, Proceedings of the Workshop on Norms and Institutions in Multi-Agent Systems (Autonomous Agents 2000), Barcelona (2000).

[7] Esteva, M., J. A. Rodriguez, C. Sierra, P. Garcia, and J. L. Arcos, On the formal specifications of electronic institutions, in F. Dignum and C. Sierra, eds., Agent mediated electronic commerce (The European AgentLink Perspective), Springer, (2001) 126–147.

[8] Finin, T., Y. Labrou, and J. Mayfield, KQML as an agent communication language, in J. Bradshaw, ed., Sofware agents, MIT Press, Cambridge, MA (1995).

[9] FIPA, Agent Communication Language, FIPA 2000 Specification, Foundation for Intelligent Physical Agents, http://www.fipa.org (2000).

[10] FIPA TC Semantics, Request for information, Document f-out-00099, Foundation for Intelligent Physical Agents, http://www.fipa.org (2001).

[11] Fornara, N., and M. Colombetti, Operational specification of a commitment-based agent communication language, Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2002), Bologna (2002), to appear.

[12] Jones, A.J.I., and M.J. Sergot, A formal characterisation of institutionalised power, Journal of the Interest Group in Pure and Applied Logic, vol. 4, num. 3, (1996) 427–443.

[13] Jones, A.J.I., Conventional signalling acts, unpublished document presented at the 24th FIPA Meeting, Lausanne (2002).

[14] Labrou, Y., and T. Finin, Semantics and conversations for an agent communication language, Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI'97), Nagoya, Japan (1997).

[15] Searle, J.R., Speech acts, Cambridge University Press, Cambridge, UK (1969).

[16] Searle, J.R., The construction of social reality, The Free Press, New York (1995).

[17] Singh, M.P., Agent communication languages: Rethinking the principles, IEEE Computer, vol. 31, (1998) 40–47.

[18] Venkatraman, M., and M.P. Singh, Verifying compliance with commitment protocols, Autonomous Agents and Multi-Agent Systems, vol. 2, num. 3, (1999) 217–236.

[19] Verdicchio, M., and M. Colombetti, Commitments for agent-based supply chain management, ACM SIGecom Exchanges, vol. 3, num. 1, (2002) 13–23.

[20] Winograd, T., and F. Flores, Understanding computers and cognition: A new foundation for design, Ablex, Norwood, NJ (1986).

[21] Yolum, M.P., and M.P. Singh, Synthesizing finite state machines for communication protocols, Technical Report TR-2001-06, Department. of Computer Science, North Carolina State University (2001).