

ITC

**Institute for
Communication
Technologies**

www.itc.com.unisi.ch

Faculty of
Communication
Sciences

Università della
Svizzera
italiana

via Buffi 13
CH-6900
Lugano

Technical report

No. 4

Artificial Institutions: A Model of Institutional Reality for Open Multiagent Systems

Nicoletta Fornara, Francesco Viganò, Mario Verdicchio, and Marco Colombetti

Institute for Communication Technologies
Università della Svizzera Italiana, 2006.

Artificial Institutions: A Model of Institutional Reality for Open Multiagent Systems ^{*}

Nicoletta Fornara¹, Francesco Viganò¹, Mario Verdicchio², and Marco Colombetti^{1,3}

¹ Università della Svizzera italiana, via G. Buffi 13, 6900 Lugano, Switzerland
{nicoletta.fornara, francesco.vigano, marco.colombetti}@lu.unisi.ch,

² Università degli Studi di Bergamo, via Salvecchio 19, 24100 Bergamo, Italy
mario.verdicchio@unibg.it

³ Politecnico di Milano, piazza Leonardo Da Vinci 32, Milano, Italy
marco.colombetti@polimi.it

Abstract. Software agents' ability to interact within different open systems, designed by different groups, presupposes an agreement on an unambiguous definition of a set of concepts, used to describe the context of the interaction and the communication language the agents can use. Agents' interactions ought to allow for reliable expectations on the possible evolution of the system; however, in open systems interacting agents may not conform to predefined specifications. A possible solution is to define interaction environments including a normative component, with suitable rules to regulate the behaviour of agents.

To tackle this problem, we propose an application-independent model of artificial institutions that can be used to define open multiagent systems. With respect to other approaches to artificial (or electronic) institutions, which mainly focus on the definition of the normative component of open systems, our proposal has a wider scope, in that we model the social context of the interaction, define the semantics of an Agent Communication Language to operate on such a context, and give an operational definition of the norms that are necessary to constrain the agents' actions. In particular, we define the semantics of a library of communicative acts in terms of operations on agents' social reality, more specifically on commitments, and regard norms as event-driven rules that, when fired by events happening in the system, create or modify a set of commitments. An interesting aspect of our proposal is that both the definition of the ACL and the definition of norms are based on the same notion of commitment. Therefore an agent capable of reasoning on commitments can reason both on the semantics of communicative acts and on the normative system.

1 Introduction

Since the advent of the Internet, about ten years ago, the landscape of computer science has substantially changed. For the first time, it has become feasible to

^{*} Supported by Swiss National Science Foundation project 200021-100260, "An Open Interaction Framework for Communicative Agents"

develop open distributed systems, that is, software systems whose components can aggregate dynamically and are free to enter and leave an interaction at their will. In an open distributed system, the interacting agents are typically designed and implemented by different parties, and may represent conflicting interests, as it happens for example in e-commerce applications. This fact has two important consequences. The first is that interaction will not even be possible unless agents are designed to comply with well-defined standards. The second consequence is that interaction will not lead to coherent outcomes unless the agents' behaviour is suitably regulated. The first problem (i.e., the problem of standards) has been tackled by a number of organizations, including FIPA, the Foundation for Intelligent Physical Agents (<http://www.fipa.org>). The second problem (i.e., the problem of regulating interactions) is much more elusive, and has become an important object of research at least since Noriega's and Sierra's works on electronic institutions in the late nineteen-nineties [21, 20]. Since then, several authors (see for example [10, 1, 30, 31, 11, 3]) have contributed to the specification of electronic institutions and artificial institutions [14, 34].

An electronic institution is usually viewed as a set of norms, which the agents interacting within an open distributed systems ought to follow [20]. Interaction protocols themselves can be conceived as sets of norms (about what an agent may do and when), and are the most obvious component of an electronic institution [11]. As a whole, the function of an electronic institution is to guarantee that if its norms are followed by all agents, the interaction will produce a desirable outcome. Indeed, it is part of the very concept of an autonomous agent that norms may be violated; the system implementing an electronic institution ought to detect such violations and to manage the situation, for example by applying suitable sanctions. Given that norms are an essential component of social reality (see for example [25]), we can regard electronic institution as a means for imposing a well-defined structure to the social reality within which agents interact. However, norms are just one component of social reality; other components, which also seem to us to be important for the specification of open distributed systems, have been largely neglected by most proposals concerning electronic institutions. For this reason we would like to put forward an extension of the concept of an electronic institution, that we call "artificial institution".

Social reality is that part of the world that exists only because it is collectively recognized to exist by a group of agents. Norms are an obvious example of something that exists only because it is accepted by the members of some community; other important components are: (i) a socially defined ontology, and (ii), linguistic conventions. Let us briefly analyse these components.

Institutions not only regulate, but also create components of social reality. For example, the institution of ownership does not just regulate a pre-existing piece of reality, but creates the very concept of owning something. In other words, an institution introduces a new ontology, including a set of entities with their properties and relationships. A very important class of entities created by an institution is the set of institutional actions. For example, actions like

selling and buying, renting and hiring, lending and borrowing (and, in fact, also stealing) are only meaningful within a suitable institution of ownership.

Contrary to natural actions, like eating or walking around, institutional actions cannot be executed only by exploiting physical abilities. Rather, they require suitable conventions that allow agents to perform institutional actions by producing certain forms of behaviour, typically linguistic behaviour. The need of specifying linguistic conventions for open distributed systems has been recognised at least since the definition of KQLM [12]. Since then, Agent Communication Languages (ACLs) have been an important subject of research. Early models of multiagent systems tended to regard ACLs as local to every specific system (or even to every pair of roles in the system, as suggested by the Aalaadin model [17]). Later research in agent communication, however, highlighted the importance of defining a universal ACL [7, 29]. Moreover, some approaches to the semantics of ACLs show that the very definition of a communication language may be based on elements of social reality, like the concept of commitment [28, 4, 5, 14]: if this approach is correct, ACL messages can be understood only within a suitable system of artificial institutions.

In general, producing an instance of the correct conventional behaviour is necessary but not sufficient to perform an institutional action. For example, the president of society can open the annual meeting by saying “The meeting is open”, while a generic member of the society cannot do the same. The issue, here, is power or, as we prefer to call it, authorization: institutional actions are successful only if they are performed by an authorized agent, and authorizations are typically associated to the roles played by the agent in the institution. Even authorized actions, however, cannot always be performed freely. For example, the owner of a car is authorized to sell it, but may be prohibited to do so by his or her spouse (in this example two institutions, ownership and family, interact). In an institution, a set of norms typically regulate the execution of authorized action.

This brief analysis shows that institutions are made up by a number of components: an ontology (including the definition of institutional actions), a set of authorizations to perform institutional actions (typically associated to roles), a set of linguistic conventions for the execution of institutional actions, and a system of norms regulating the agents’ interaction. All these components are so strictly interconnected that they have to be dealt with within a single conceptual framework. The metamodel of artificial institutions that we propose in this paper is an attempt to define such a framework. Our presentation is structured as follows. In Section 2 we introduce our metamodel of artificial institutions, defining their fundamental concepts and relationships. In Section 3 an example of the specification of an artificial institution, the Basic Institution, is presented. Using the notion of commitment defined by the Basic Institution, in Section 4 we present our operational definition of norms. Finally, in Section 5 we conclude with a brief discussion of related work and open questions.

2 A Metamodel of Artificial Institutions

Artificial institutions are *models* of institutional reality which specify a class of interaction domains by exploiting a set of common concepts and notations. For this reason we introduce a *metamodel* of artificial institutions to describe an abstract syntax and a semantics for each construct that characterizes artificial institutions. In Figure 1 the abstract syntax is presented by means of a notation inspired by the UML metamodel [23], showing the fundamental concepts and the relationships existing between them.

Even if we have adopted a notation inspired by UML, we are not prescribing that artificial institutions and agents acting within them should be implemented by object oriented technologies. In fact, artificial institutions are intended to specify institutional reality for open multiagent systems, which are characterized by the fact that heterogeneous agents realized with different technologies and by different organizations should interact. Therefore, it is important that the specification does not commit to a specific implementation language, platform, or internal architecture. For this reason, we do not exploit most of the features of object oriented technologies, like methods or polymorphism, and to avoid confusion we will not talk of classes. Therefore, we do not adopt the meta-metamodel of UML⁴, which introduces, among others, the concept of MetaClass and MetaOperations.

We use UML as a starting point, reusing its well known graphical notation and the Object Constraint Language (OCL [22]) as our language to express constraints. An advantage of this approach is that it employs concepts that are close to the intuition and knowledge of practitioners. We believe the metamodel we have developed can be easily understood by software engineers who design and implement open multiagent systems.

Like the metamodel of UML, our metamodel of artificial institution has a declarative semantics and suppresses implementation details. Therefore, we abstract away from methods and implementation issues and provide an abstract syntax to specify agent interaction systems.

In this paper we do not propose a specific architecture for the management of institutional states in real systems, because we think that different architectures may be chosen and adapted according to the needs of real systems. Anyway, agents using the interfaces specified through an artificial institution model should not even notice the existence of such design and implementation choices.

In the following sections we shall introduce a notation for the components of artificial institutions and describe their meaning in natural language, whereas the metamodel reported in Figure 1 provides a set of well-formedness rules for a model of an artificial institution. In Section 3 we will exemplify our approach by defining a model of a specific institution, the Basic Institution.

2.1 The Core Ontology

⁴ For reason of conciseness we do not discuss our meta-metamodel and the relations existing between it and the meta-metamodel defined by UML.

communicative acts are regarded as institutional actions, because they create commitments or modify their attributes.

More precisely, an institutional action is characterized by:

- an *action name*;
- a possibly empty set of *parameters*;
- a possibly empty set of *ontological preconditions*, which specify the values that certain institutional attributes must have for the action to be meaningful (for example, opening an auction is meaningful only if the auction is not already open);
- a nonempty set of *postconditions*, which specify the values of certain institutional attributes after a successful performance of the action.

Preconditions and postconditions of institutional actions are expressed through OCLExpressions [22], which can refer only to the parameters of the institutional action or institutional attributes defined by the core ontology, reflecting the fact that the only effect of an institutional action is to modify institutional attributes.

Because institutional actions change institutional attributes, which exist only thanks to common agreement, it follows that institutional actions have an intrinsic social nature. Therefore, agents cannot perform such actions by exploiting causal links occurring in the natural world, as it would be done to open a door or to move a physical object. Furthermore, a crucial condition for the actual performance of institutional actions is that they must be *public*, that is, made known to the relevant agents by means of some action that can be directly executed by an artificial agent. In the human world such instrumental actions vary from certain bodily movements (raising one’s arm to vote), to the use of specific physical tools (waving a white flag to surrender), to the use of language (saying “the auction is open” to open an auction). We assume that in a multiagent system all institutional actions are performed by means of a single type of instrumental action, namely exchanging a message.

Roles A core ontology also defines a set of relationships between agents and entities, and we regard such relationships as *roles*; thus roles are always defined relative to entities. For example, an agent can be the *owner* of a commodity. Clearly, such relationships are supported by the common agreement of agents and therefore roles have the same nature and characteristics of institutional attributes. Therefore, roles are also institutional facts and they can be changed only by performing institutional actions.

As the metamodel reported in Figure 1 shows, a role is related to authorizations and Event-Condition-Action rules (ECA rules, see Section 2.3) which are also used to represent norms (see Section 4). Therefore, roles can also be seen as way to relate agents to their authorizations and norms, abstracting from the concrete agents that will be actually involved in an interaction. In general, a sound specification of an artificial institution cannot define a role with an empty set of authorizations or ECA rules.

Events A core ontology also describes a set of events which are relevant for the definition and activation of ECA rules (see Section 2.3). UML models four kinds of events [23] : *signals*, *calls*, *passing of time* and *change in state*. Unfortunately, the notation proposed in UML for modeling events is bound to features of State Machine and Statechart Diagrams, while we need a way to describe events in general.

In our framework, an event type can have attributes, providing information about the state transition that caused it, and it is possible to model hierarchies of event types. Moreover we assume that every event has a *time* attribute reporting the time at which an event has occurred. In our formalization we have singled out three main categories of events:

- a *TimeEvent*, which occurs when the system reaches a certain instant of time;
- a *ChangeEvent*, which happens when an institutional entity changes in some way. This kind of event type can be further specialized:
 - an *AttributeChange*, which is registered when an attribute has changed its value.
 - a *RelationChange*, which happens when a new relation is *created* or an existing one between two institutional entities is *dropped*.
 - a *GeneralizationChange*, which occurs when an entity modifies its type in a given taxonomy.
- an *ActionEvent*, that happens when an agent perform an action. In particular, an interesting type of this kind of events is *ExchMsg*, which represents the act of exchanging a message.

The definition of event types allows us to define *event templates*, that is, event types that have some restriction on certain attributes and describe a set of possible event occurrences. Event templates are used in the *on* section of ECA rules to specify what kind of domain dependent events activates a rule.

2.2 Authorizations and Conventions

Given that institutional actions modify institutional attributes, agents cannot directly perform such actions. Instead, we assume that all institutional actions are performed by exchanging messages, thanks to *constitutive rules* [25] which bind the exchange of a message to the performance of an institutional action.

To model the connection between instrumental actions and institutional actions we introduce the notion of a *convention*, that is, an agreement about what type of message is bound to a given type of institutional action. In our model the definition of a convention has the following generic form:

$$ExchMsg(message\ type, sender, receivers, iaction(parameters)) =_{conv} iaction(parameters)$$

In [14] we have defined some useful conventions for the performance of institutional actions (in particular communicative acts). For example, by definition

[24], the content of a declaration describes precisely the institutional changes that it brings about. Therefore, we take messages of type *declare* as the fundamental means to perform institutional actions and the convention is defined as follows:

$$\text{ExchMsg}(\text{declare}, \text{sender}, \text{receivers}, \text{content}) =_{\text{conv}} \text{iaction}(\text{parameters})$$

By itself, a convention is not sufficient to guarantee the successful performance of an institutional action by the exchange of the appropriate message: indeed, some additional *contextual conditions*² must be satisfied about the agent that sends the message, about the agents that receive the message, and about the state of the system in relation to the content of the message.

Conditions on the sender of the message. The sender of the message must be authorized to perform an institutional action. In the specification of an interaction system authorizations are expressed in term of roles; for example, only the auctioneer can open an auction by sending a suitable message to the participants. Moreover an authorization can be given only if certain conditions about the state of the system, expressed by suitable Boolean expressions, are satisfied. For example, it may be established that an auction is validly opened only if there are at least two participants. Therefore, we abstractly define the authorization to perform a specific institutional action (with given parameters) associating it to a role defined in the context of a specific entity as follows:

$$\text{Auth}(\text{entity.role}, \text{iaction}(\text{param}), \text{conditions})$$

It is worth highlighting that authorizations are a necessary condition for the successful performance of an institutional action and that designers should specify a set of authorizations whenever they introduce a new institutional action. As we will see in Section 4, this fact allows us to assume that every *authorized* institutional actions that is not prohibited is *permitted*.

Conditions about the receivers of the message. Secondly, messages realizing institutional actions should be received by all agents that are affected by the performance of the act; for example all the participants of an auction must receive the message that opens it.

Conditions about the state of the system. All the preconditions of the institutional action associated to the performance of the exchange of the message must be satisfied; for instance, an auction cannot be closed if it has not been opened yet.

² Following Searle [25], the construction of social reality in the human world is possible thanks to *constitutive rules* of the form *X counts as Y in C*. Similarly, in an artificial system constitutive rules bind the performance of a message exchange to the corresponding institutional action only if certain conditions about the context are satisfied.

2.3 Event-Condition-Action Rules

The last component of our metamodel is represented by Event-Condition-Action rules, which allow us to model the fact that institutional reality may change as a consequence of the occurrence of certain events. For example, a competition may start at a fixed time instant and be considered closed at another time without an agent having declared it. ECA rules are also useful to model how the performance of an institutional action executed within an institution may affect another institution (see for example [34]).

Inspired by Active Database models, we define an ECA rule as composed by three elements:

1. an *event template*, which represents the class of events which may activate the rule.
2. a *condition*, expressed through a Boolean OCLEExpression, which can refer to the variable e which contains a description of the event that has activated the rule.
3. a nonempty set of *institutional actions* which are executed by the rules.

The semantics of ECA rules is given as usual: when an event matching the description given by the event template occurs in the system, the variable e is filled with the event instance and the condition is evaluated; if the condition is verified, the set of institutional actions are executed and their effects are brought about in the system. We assume that the system is authorized by default to perform every institutional action. In our model, ECA rules are specified according to the following notation:

```

on  $e$ : event-template
if condition then
  do institutionalAction(parameters)+

```

As we will see in Section 4, we define norms as ECA rules that perform institutional actions creating, canceling or modifying commitments. Therefore, before introducing norms, we need to define the Basic Institutions, that is, the artificial institution that defines what social commitments are.

3 The Basic Institution

The *Basic Institution* is the institution that defines a fundamental concept: the notion of social *commitment*. The importance of commitment is due to the fact that, from our perspective, it is essential to express the meaning of most types of communicative acts (see [13, 14]) and because it is used to define the semantics of norms as will be shown in Section 4. Therefore we assume that the Basic Institution has to be used, together with other special institutions, in the specification of every open interaction framework.

In this section we report a fragment of the specification of an artificial institution, the Basic Institution, using the concepts and the meta-model introduced

so far (for the complete specification see [14]). Other institutions, that we call *special institutions*, can then be defined to model the aspects of institutional reality typical of certain application domains. For instance, for electronic commerce applications it will be necessary to model the institutions of ownership, money, business transactions, auctions, and so on. A detailed specification of the English auction, of the Dutch Auction, and of the Auction House as special institutions can be found in [14, 34].

The Basic Institution defines the ontology of commitment, the institutional actions necessary to operate on it, and a set of authorizations for the performance of these institutional actions. In general, institutions also define sets of norms to regulate the behaviour of agents, but in our current view, the Basic Institution does not specify norms.

3.1 Commitments and Temporal Propositions

In this paper we give only a short description of our model of *commitment* and of *temporal proposition*, that are used to represent the content of commitments (a complete treatment of these two concepts can be found in [14, 5]). A commitment is characterized by the following attributes: a *debtor*, a *creditor*, a *content*, and a *state*. Figure 2 represents the class diagram of the ontology of the Basic Institution. In the rest of the paper we will refer to commitments using the following notation:

$$Comm(state, debtor, creditor, content)$$

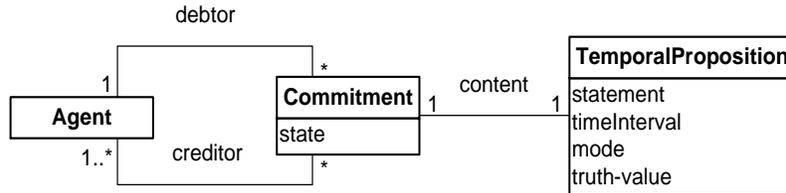


Fig. 2. Class diagram of the ontology of the Basic Institution

The content of commitments is expressed using temporal propositions. A temporal proposition is characterized by a *statement* about a state of affairs or about an action. The statement is referred to a *time interval* with two possible different *modes*: *exist* (\exists) or *for all* (\forall). The *truth-value* of a temporal proposition is initially *undefined* (\perp). It becomes *true*, thanks to a “notifier”, if the mode is *for all* and the statement is true for every instant of the associated time interval or if the mode is *exist* and the statement is true for some instant in the associated time interval, otherwise it becomes *false*. Temporal propositions are represented with the following notation:

$$TP(statement, [t_{start}, t_{end}], mode, truth-value)$$

A commitment undergoes the life cycle described in [13] by reacting either to institutional actions performed by agents or to domain-dependent events, which modify the truth value of the temporal proposition in its content. If its temporal proposition becomes true, the commitment becomes *fulfilled*; if it becomes false the commitment becomes *violated*. For instance the commitment of agent a_1 to agent a_2 to open the auction i *within* 10 minutes from *now* (or *between now* and $now + 10$) is represented as:

$$Comm(pending, a_1, a_2, TP(open(a_1, auction_i), [now, now+10m], \exists, \perp))$$

In our framework every agent is authorized to create a commitment by performing the *makeCommitment* institutional action, whose successful performance creates an *unset* commitment. The debtor of an *unset* commitment may refuse it by executing *setCancel*, or it may undertake the proposed commitment by executing *setPending*. We represent a refused commitment by means of the *cancelled* state, whereas an accepted commitment is depicted with the *pending* state. The creditor of a *pending* or *unset* commitment can always set it to *cancelled*. Here, due to space limitation, we report only the definition of one institutional action and of one authorization. The institutional action *makePendingComm*, used in Section 4, creates a *pending* commitment and its execution coincides with the sequential performance of *makeCommitment* and *setPending*:

name : **makePendingComm**(*debtor, creditor, content*)
pre : *not Comm.allInstances()* \rightarrow *exists(c|c.debtor = debtor*
 and c.creditor = creditor and c.content = content)
post : *Comm.allInstances()* \rightarrow *exists(c|c.state=pending and*
 c.debtor=debtor and c.creditor=creditor and
 c.content = content)

Every registered agent *RegAgt* is authorized to create a commitment having itself as *debtor*:

$$Auth(RegAgt, \mathbf{makePendingComm}(debtor, creditor, content), \\ RegAgt = debtor).$$

4 Norms

In the literature, it is possible to find two different approaches to the specification of norms for open systems. The more formal approach uses deontic logic to express the semantics of norms and is more suitable when the goal is to verify the consistency of a set of norms [8]. The other approach focuses on the operational aspects of norms, and is more suitable when the goal is the actual implementation of norms in open systems [32, 16]. In this paper we study the formalization of norms from the *institutional and operational perspective* (opposite to the agent

perspective) [32], that is, our goal is to define norms that make the evolution of the state of an interaction system partially predictable without violating the autonomy of the interacting agents. In order to reach this purpose, norms describe, on the basis of the state of the system, the expected behaviour of the agents playing different roles, and make it possible to detect deviations from such a behaviour, that is, the violation of obligations and prohibitions defined in the system.

As already said in Section 2.3 in our model norms are a special type of ECA rules characterized by the fact that when they are fired by events happening in the system, they create or cancel commitments to perform or to not perform an action within a certain interval of time, thus affecting *each* agent that satisfies a suitable selection expression. Usually the collection of liable agents corresponds to the set of agents that play a given role in the institution.

From our point of view, commitments are not a specialization of norms as in [19] and norms are not themselves a special kind of commitments as in [2] and [27]. We perceive norms as rules that manipulate commitments of the agents engaged in an interaction. This because in the abstract formalization of a system it is important to model norms associated to roles rather than to individual agents, whereas during the actual evolution of the system it is fundamental to create commitments associated to individual agents and to detect their fulfillment or violation. Obviously such violations can be interpreted also as the violation of the corresponding norm.

When an agent fills a role in an institution, we assume that it accepts that norms will create commitments binding it to a pseudoagent representing the institution, which we call *institutional agent*. Such an agent allows us to keep trace of the commitments created by an instance of an institution. This also means that commitments created by norms of an institution can be cancelled only by other norms defined by the same institution, because only the creditor of a pending commitment is authorized to set it to cancelled [14]. The general structure of a norm can be described as follows:

```

on e : event-template
if condition then
  foreach agent in selection-expression
  do commitment-Ops

```

where *agent* is an identifier varying on the set of agents that satisfy the selection expression; *selection-expression* is a list of agent identifiers or a *role* defined in a certain artificial institution; *commitment-operations* is a sequence of commitment operations defined using BNF notation as follows:

```

commitment-Ops := comm-Op | comm-Op, commitment-Ops
comm-Op := makePendComm(agent, instAgent, content) |
           cancelComm(agent, instAgent, content)
content := TP(action, [t, t],  $\exists$ ) | TP( $\neg$ action, [t, t],  $\exists$ )
t := now | event.time | instant | now+number | event.time+number

```

where *action* is an institutional action as described in Section 2; *now* is the time of the system when the temporal proposition is created and *instant* is an instant of time.

In [14] we give a formalization of the English Auction Institution. Here we report only an example of a norm that creates an obligation for the auctioneer to open the auction when the *start.time* has elapsed, if at least two agents have been registered as participants:

```

on e: TimeEvent(a.startTime)
if UnsetEnglishAuction.participant.sizeOf() >= 2 then
foreach agent in UnsetEnglishAuction.auctioneer
  do makePendingComm(agent, institutionalAgent,
    TP(openAuction(UnsetEnglishAuction.id), [now, now + δ], ∃))

```

where δ is the time allowed to the agent to fulfill its obligation.

4.1 Normative Positions

Using our metamodel of artificial institutions and the definition of norms it is possible to represent fundamental normative positions between agents [26]:

- First of all, an institutional action has to be authorized, that is, it has to be accepted by the agents taking part to the interaction (or their designers) that the exchange of a message counts as the performance of an institutional action. If an agent performs an instrumental action bound by a convention to an institutional action a_i , but is not authorized to perform a_i , neither the “counts as” relation will hold, nor the effects of a_i will take place.
- When an action is authorized and is not prohibited, it is *permitted*. Usually in the agent literature authorization is not distinguished from permission or the former encompasses the latter. Coherently with the concept of institutionalized power [18], we distinguish between the notions of authorization and permission. The main difference between authorization and permission resides in the effects of the action. Whereas the former represents a necessary condition for the execution of institutional action, permission reflects the need to regulate the performance of authorized actions, but cannot prevent the effects deriving from the performance of a forbidden act.
- An action is *prohibited* when a norm has created a commitment not to perform it. If an agent is prohibited to perform action a_i but performs it anyway, the effects of the action take place and the commitment to not perform the action is violated. At this point it is up to the system to impose sanctions to the misbehaving agent, and possibly recover the system to a safe state.
- *Obligations* are represented by commitments, created by norms, to perform an action of a given type. It may happen that agents do not perform obligatory actions; in such a case the commitment becomes violated and the system does not evolve to the expected new state. The violation of an obligation may therefore be punished, and in critical cases the system may be recovered to a safe state by performing a suitable set of actions.

The treatment of sanctions, either direct (fine, expulsion) or indirect (trust, reputation), and the study of plans to recover the system from unsafe states are a wide and interesting research problem we plan to tackle in the future.

5 Conclusions and Future Works

In this paper we have described in short our operational model of Artificial Institutions that can be used for the specification of open interaction systems. In particular, we have focused on the definition of a standard syntax and an intuitive semantics of the concepts that an agent, or its designer, has to take into account and to reason on, in order to be able to interact properly with different systems.

Existing proposals for the definition of electronic institutions like the one proposed in the EIDE platform [10, 9], the one presented by Cliffe and Padget [3], or the OMNI framework [33] are mainly focused on the specification of the normative component. Unlike them, our model includes also the concepts for defining the institutional reality of the system. Moreover, differently from [10, 9], where the semantics of the communicative acts is given in terms of predefined state transitions, in our model it is possible, using commitments, to define an application independent semantics of a Communicative Act Library, as presented in [14], with a syntax compatible with FIPA-ACL [15].

As far as the formalization of normative aspects is concerned, likewise the other mentioned proposals, we distinguish among permission, obligation, and prohibition. We also distinguish between normative aspects and facts dealing with institutional power [18], which separate meaningful/empowered actions from meaningless actions that do not have an effect. In our opinion, this last type of fact, like in [1], should be considered as part of the declarative apparatus of an Electronic Institution, and not of the normative one. Indeed we think that norms regulate agent interaction by indicating, among all meaningful actions at a certain stage, those that are obliged or forbidden. Moreover an interesting aspect of our proposal is its uniformity; in fact, our operational formalization of norms uses the same concepts used for the definition of the semantics of ACL, that is, social commitments.

Several research questions are still open, and will be tackled in future works. We will investigate the development of methods for discovering inconsistencies among the specification of one or more artificial institutions. In particular, we are interested in verifying during the specification phase whether certain norms may create obligations to perform unauthorized actions, or under what conditions two norms may generate conflicting commitments. Moreover we intend to devise an explicit representation of the sanctions and repair procedures connected to the violation of commitments.

References

1. A. Artikis, M. Sergot, and J. Pitt. Animated Specifications of Computational Societies. In C. Castelfranchi and W. L. Johnson, editor, *Proceedings of the 1st*

- International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, pages 535–542. ACM Press, 2002.
2. C. Castelfranchi. Commitments: From Individual Intentions to Groups and Organizations. In V. Lesser, editor, *Proceedings of the 1st International Conference on Multi-Agent Systems*, pages 528–535, San Francisco, USA, 1995. AAAI-Press and MIT Press.
 3. O. Cliffe, M. D. Vos, and J. Padget. Specifying and Analysing Agent-based Social Institutions using Answer Set Programming Agents. In D. Lindermann, S. Ossowski, J. Padget, and J. Vázquez-Salceda, editors, *Proc. of the AAMAS05 Workshop on Agents, Norms and Institutions for Regulated Multiagent Systems*, pages 107–122, 2005.
 4. M. Colombetti. A commitment-based approach to agent speech acts and conversations. In *Proceedings of the Workshop on Agent Languages and Communication Policies, 4th International Conference on Autonomous Agents (Agents 2000)*, pages 21–29, Barcelona, Spain, 2000.
 5. M. Colombetti, N. Fornara, and M. Verdicchio. A Social Approach to Communication in Multiagent Systems. In J. A. Leite, A. Omicini, L. Sterling, and P. Torroni, editors, *Declarative Agent Languages and Technologies*, volume 2990 of *LNAI*, pages 191–220. Springer, 2004.
 6. M. Colombetti and M. Verdicchio. An analysis of agent speech acts as institutional actions. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, pages 1157–1166, 2002.
 7. F. Dignum, editor. *Advances in Agent Communication, International Workshop on Agent Communication Languages, ACL 2003, Melbourne, Australia, July 14, 2003*, volume 2922 of *LNAI*. Springer, 2004.
 8. F. Dignum, J. Broersen, V. Dignum, and J. C. Meyer. Meeting the deadline: Why, When and How. In *In Proceedings of the 3rd International Workshop on Formal Approaches to Agent-Based Systems (FAABS)*, pages 30–40, 2004.
 9. M. Esteva, J. A. Rodríguez-Aguilar, B. Rosell, and J. L. Arcos. AMELI: An Agent-based Middleware for Electronic Institutions. In N. R. Jennings, C. Sierra, L. Sonenberg, and M. Tambe, editors, *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004)*, pages 236–243. ACM Press, 2004.
 10. M. Esteva, J. A. Rodríguez-Aguilar, C. Sierra, P. Garcia, and J. L. Arcos. On the formal specification of electronic institutions. In F. Dignum and C. Sierra, editors, *Agent Mediated Electronic Commerce, The European AgentLink Perspective*, volume 1991 of *LNAI*, pages 126–147. Springer, 2001.
 11. M. Esteva, J. A. Rodríguez-Aguilar, C. Sierra, and W. Vasconcelos. Verifying Norm Consistency in Electronic Institutions. In V. Dignum, D. Corkill, C. Jonker, and F. Dignum, editors, *Proceedings of the AAAI-04 Workshop on Agent Organizations: Theory and Practice*, pages 8–15, San Jose, California, USA, 2004.
 12. T. Finin, Y. Labrou, and J. Mayfield. KQML as an agent communication language. In J. M. Bradshaw, editor, *Software Agents*, chapter 14, pages 291–316. AAAI Press / The MIT Press, 1997.
 13. N. Fornara and M. Colombetti. A commitment-based approach to agent communication. *Applied Artificial Intelligence an International Journal*, 18(9-10):853–866, 2004.
 14. N. Fornara, F. Viganò, and M. Colombetti. Agent Communication and Artificial Institutions. *Journal of Autonomous Agents and Multi-Agent Systems*. to appear.

15. Foundation for Intelligent Physical Agents. FIPA Communicative Act Library Specification. <http://www.fipa.org>, 2002.
16. A. Garcia-Camino, P. Noriega, and J. A. Rodriguez-Aguilar. Implementing norms in electronic institutions. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2005)*, pages 667–673. ACM Press, 2005.
17. J. Ferber and O. Gutknecht. A meta-model for the analysis and design of organizations in multi-agent systems. In *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS98)*, pages 128–135, Paris, France, 1998.
18. A. Jones and M. J. Sergot. A formal characterisation of institutionalised power. *Journal of the IGPL*, 4(3):429–445, 1996.
19. F. López y López, M. Luck, and M. d’Inverno. Normative Agent Reasoning in Dynamic Societies. In N. R. Jennings, C. Sierra, L. Sonenberg, and M. Tambe, editors, *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004)*, pages 535–542. ACM Press, 2004.
20. P. Noriega. *Agent mediated auctions: The Fishmarket Metaphor*. PhD thesis, Universitat Autnoma de Barcelona, 1997.
21. P. Noriega and C. Sierra. Towards layered dialogical agents. In J. P. Mueller, M. Wooldridge, and N. Jennings, editors, *Proceedings ATAL 96. Intelligent Agents III. Agent Theories, Architectures, and Languages*, volume 1193 of *LNAI*. Springer-Verlag, 1996.
22. Object Management Group. UML 2.0 OCL Specification. <http://www.omg.org/>, 2005.
23. Object Management Group. Unified Modeling Language Specification. <http://www.omg.org/>, 2005.
24. J. R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, United Kingdom, 1969.
25. J. R. Searle. *The construction of social reality*. Free Press, New York, 1995.
26. M. Sergot. A computational theory of normative positions. *ACM Transactions on Computational Logic*, 2(4):522–581, 2001.
27. M. P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7:97–113, 1999.
28. M. P. Singh. A social semantics for agent communication languages. In *Proceedings of IJCAI-99 Workshop on Agent Communication Languages*, pages 75–88, 1999.
29. R. van Eijk, M. Huget, and F. Dignum, editors. *Agent Communication, International Workshop on Agent Communication, AC 2004, New York, NY, USA, July 19, 2004, Revised Selected and Invited Papers*, volume 3396 of *LNAI*. Springer, 2005.
30. W. Vasconcelos, J. Sabater, C. Sierra, and J. Querol. Skeleton-based agent development for electronic institutions. In C. Castelfranchi and W. L. Johnson, editor, *Proceedings of the 1st international joint conference on Autonomous Agents and Multiagent Systems (AAMAS ’02)*, pages 696–703. ACM Press, 2002.
31. J. Vázquez-Salceda. *The role of Norms and Electronic Institutions in Multi-Agent Systems applied to complex domains. The HARMONIA framework*. PhD thesis, Universitat Politècnica de Catalunya, 2003.
32. J. Vázquez-Salceda, H. Aldewereld, and F. Dignum. Norms in multiagent systems: some implementation guidelines. In *Proceedings of the European Conference on Multiagent Systems EUMAS04*, pages 737–748, Barcelona, Spain, 2004.
33. J. Vázquez-Salceda, V. Dignum, and F. Dignum. Organizing Multiagent Systems. Technical report, Institute of Information and Computing Sciences Utrecht University, 2004.

34. F. Viganò, N. Fornara, and M. Colombetti. An Event Driven Approach to Norms in Artificial Institutions. In D. Lindermann, S. Ossowski, J. Padget, and J. Vázquez-Salceda, editors, *Proceedings of the AAMAS05 Workshop on Agents, Norms and Institutions for Regulated Multiagent Systems, Utrecht, The Netherlands*, pages 149–162, 2005.

Technical report No. 1

Agent Communication and Institutional Reality

N. Fornara, F. Viganò, M. Colombetti

2004

Technical report No. 2

An Operational Approach to Norms in Artificial Institutions

F. Viganò, N. Fornara, and M. Colombetti

2005

Technical report No. 3

FIEVeL, a language for the specification and verification of artificial institutions

F. Viganò, 2006

ITC

**Institute for
Communication
Technologies**

www.itc.com.unisi.ch

Faculty of
Communication
Sciences

Università della
Svizzera
italiana

via Buffi 13
CH-6900
Lugano