Specifying and Enforcing Norms in Artificial Institutions*

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

[1] Università della Svizzera italiana, via G. Buffi 13, 6900 Lugano, Switzerland
{nicoletta.fornara,marco.colombetti}@lu.unisi.ch
[2] Politecnico di Milano, piazza Leonardo Da Vinci 32, Milano, Italy
marco.colombetti@polimi.it

Abstract. In this paper we investigate two related aspects of the formalization of open interaction systems: how to specify norms, and how to enforce them by means of sanctions. The problem of specifying the sanctions associated with the violation of norms is crucial in an open system because, given that the compliance of autonomous agents to obligations and prohibitions cannot be taken for granted, norm enforcement is necessary to constrain the possible evolutions of the system, thus obtaining a degree of predictability that makes it rational for agents to interact with the system. In our model, we introduce a construct for the definition of norms in the design of artificial institutions, expressed in terms of roles and event times, which, when certain activating events take place, is transformed into commitments of the agents playing certain roles. Norms also specify different types of sanctions associated with their violation. In the paper, we analyze the concept of sanction in detail and propose a mechanism through which sanctions can be applied.

1 Introduction

In our previous works [10, 11, 26] we have presented a metamodel of artificial institutions called *OCeAN* (Ontology, CommitmEnts, Authorizations, Norms), which can be used to specify at a high level and in an unambiguous way *open interaction systems*, where heterogeneous and autonomous agents may interact.

In our view open interaction systems and artificial institutions, used to model them, are a technological extension of human reality, that is, they are an instrument by which human beings can enrich the type and the frequency of their interactions and overcome geographical distance. Potential users of this kind of systems are artificial agents, that can be more or less autonomous in making decisions on behalf of their owners, and human beings using an appropriate interface. For example, it is possible to devise an electronic auction where the artificial agents are autonomous in deciding the amount of their bids, or an interaction system for the organization of conferences in which human beings (like

* Supported by the Hasler Foundation project number 2204 title "Artificial institutions: specification of open distributed interaction systems".

M. Baldoni et al. (Eds.): DALT 2008, LNAI 5397, pp. 1–17, 2009.

the organizers, or the Program Committee members) act by means of artificial agents that have a very limited level of autonomy. In any case it is important to remark that in every type of system there is always a stage when the software agents have to interface with their human owners to perform certain actions in the real world. For these reasons artificial institutions have to reflect, with the necessary simplifications, crucial aspects of their human counterparts. Therefore in devising our model we draw inspiration from an analysis of social reality [22] and from human legal theory [14].

In this paper we concentrate mainly on the definition of the constructs necessary for the specification of the normative component of artificial institutions, that is, of obligations, permissions and prohibitions of the interacting agents. The normative component is fundamental because it can be used to specify the expected behavior of the interacting agents, for example by means of flexible protocols [27]. We shall extend our *OCeAN* metamodel by defining a construct for the specification of norms for open systems, whose semantics is expressed by means of *social commitments*, the same concept that we have used to specify the semantics of a library of communicative acts [9, 10]. Commitments, having a well defined life-cycle, will be used at run-time to detect and react to the violation of the corresponding norms.

An important feature of our proposal, with respect to other ones [1, 5, 12, 19, 24], is that, using the construct of a commitment, it gives a uniform solution to two crucial problems: the specification of the semantics of norms and the definition of the semantics of an Agent Communication Language. Therefore a software agent able to reason on one construct is able to reason on both communicative acts and norms.

Moreover we present an innovative and detailed analysis of problem of defining a mechanism for enforcing obligations and prohibitions by means of sanctions that, that is, a treatment of the actions that have to be performed when a violation occurs, in order to deter agents from misbehaving and to secure and recover the system from an undesirable state. We speak of "obligation and prohibition enforcement" instead of "norm enforcement", as done in other approaches, because our proposal can be used to enforce obligations and prohibitions that derive either from predefined norms or from the autonomous performance of communicative acts.

The problem of managing sanctions has been tackled in a few other works: for example, López y López et al. [19] propose to enforce norms using the "enforcement norms" that oblige agents entitled to do so to punish misbehaving agents but does not treat the actions that the misbehaving agents may have to perform to repair to its violation; Vázquez-Salceda et al. [24] present, in the OMNI framework, a method to enforce norms described at a different level of abstraction but do not investigate in detail the mechanism to manage santctions; whereas Grossi et al. in [13] develop a high-level analysis of the problem of enforcing norms. Other interesting proposals introduce norms to regulate the interaction in open systems but, even when the problem of enforcement is considered to be crucial, do not investigate with sufficient depth why an agent ought to comply with norms and what would happen if compliance does not occur. For instance, Esteva et al. [5, 12] propose ISLANDER, where a normative language with sanctions is defined but not discussed in detail, Boella et al. [3] model violations but do not analyze sanctions, and Artikis et al. [1] propose a model where the problem of norm enforcement using sanctions is mentioned but not fully investigated.

The paper is organized as follows: in Section 2 we briefly describe the part of metamodel for artificial institutions that we have presented in other works [10, 11, 26]. In Section 3 the reasons why in open interaction frameworks it makes sense to allow for the violation of obligations and prohibitions are discussed, and then in Section 4 a proposal on how to enforce obligations and prohibitions by means of sanctions is presented. In Section 5 our model of norms is described and the construct of commitment is extended, with respect to our previous works, by adding the treatment of sanctions. In Section 6 we briefly exemplify our proposal and finally in Section 7 we present conclusions.

2 The OCeAN Metamodel

In our previous works we have started to define the OCeAN metamodel [10, 11], that is, the set of concepts, briefly recalled in the sequel, that can be used to design artificial institutions. Examples of artificial institutions are the institution of language (that we call the *Basic Institution*, because we assume it will be used in the specification of every interaction system), the institution of English or Dutch auctions [10, 26], and the institution of organizations. In our view an open interaction system for autonomous agents can be specified using one or more artificial institutions. The state of the interaction system will then evolve on the basis of the events and actions that take place in the system, and whose effects are defined in the various institutions and on the basis of the life-cycle of the concepts defined in our model. (Investigating the relationships among the specification of different institutions is an interesting open problem [4], that we shall not tackle in this paper.) The concepts introduced by our metamodel are:

- The constructs necessary to define the *core ontology* of an institution, including: the notion of *entity*, used to define the concepts introduced by the institution (e.g., the notion of a run of an auction with its attributes introduced by the institution of auctions); the notion of *institutional action*, described by means of their preconditions and postconditions (e.g., the action of opening an auction, or declaring the current ask-price of an auction). The core ontology also defines the syntax of a list of *base-level actions*, like for instance the action of exchanging a message, whose function is to concretely execute institutional actions.
- Two fundamental concepts that are common to all artificial institutions: the notions of *role* and of *event*. In particular roles are used in the specification of authorizations and norms, while the happening of events is used to bring about the activation of a norm or to specify the initial or final instance of a time interval.

- The *counts-as relation* that is necessary for the concrete performance of institutional actions. In particular, such relation relies on a set of *conventions* that bind the exchange of a certain message, under a set of contextual conditions, to the execution of an institutional action. Contextual conditions include *authorizations* (called also *powers*) that specify what agents are authorized to perform certain institutional actions. Authorizations for the agent playing a given *role* to perform an institutional action *iaction* with a certain set of *parameters* if certain *conditions* are satisfied are represented with the following notation: *Auth(role,iaction(parameters), conditions).*
- The construct of *norm* analyzed, discussed, and defined in Section 5, used to impose obligations and prohibitions to perform certain actions on agents interacting with the system.

3 Regimentation vs. Enforcement

In our model, as it will be discussed in more detail in Section 5, an active obligation is expressed by means of *commitments* to perform an action of a given type within a specified interval of time; similarly, an active prohibition is expressed by a commitment not to perform an action of a given type; moreover, every action is permitted unless it is explicitly forbidden. Note that a commitment can be created not only by the activation of a norm, but also by the performance of a communicative act [10], for instance by a promise.

In this section we briefly discuss the reasons why in open interaction systems it makes sense, and sometimes it is also inevitable, to allow for commitment violations, that happen when a prohibited action is performed or when an obligatory action is not performed within a predefined interval of time. The question is, why should we give an agent the possibility to violate commitments? why not adopt what in the literature is called "regimentation" [14], as proposed in [13], by introducing a control mechanism that does not allow agents to violate commitments?

To answer this question, it is useful to distinguish between obligations and prohibitions. With respect to obligations, there is only one way to "regiment" the performance of an obliged action, that is, by making the system performing the obliged action instead of a misbehaving agent. But this solution is not always viable, especially when the agent has to set the values of some parameters of the action. For instance, the auctioneer of a Dutch Auction is repeatedly obliged to declare a new ask price, lower than the one previously declared, but can autonomously decide the value of the decrement; therefore it would be difficult for the system to perform the action on behalf of the auctioneer. In any case it has to be taken into account that, even if the regimentation of obligations violates the autonomy of self-interested interacting agents, sometimes it can be adopted to recover the system from an undesirable state.

With respect to the regimentation of prohibitions, it is useful to introduce a further distinction between *natural* (or physical) actions (like opening a door or physically delivering a product), whose effects take place thanks to physical laws, and *institutional* actions (like opening an auction or transferring the property of a product), whose effects take place thanks to the common agreement of the interacting agents (more precisely, of their designers). For our current purpose, the main difference between natural and institutional actions is that, under suitable conditions, the latter can be "voided", that is, their institutional effects can be nullified; on the contrary, this is not possible with natural actions. Consider for example the difference between destroying and selling an object: while in general a destroyed object cannot be brought back into existence, the transfer of ownership involved in selling it can always be nullified. The previous considerations imply that, in general, it is impossible to use regimentation to prevent the violation of a prohibition to perform a natural action. Concerning the prohibition of institutional actions, in our model it can be expressed using two different mechanisms: (i) through the lack of authorization: in fact, when an agent performs a base-level action bound by a convention to an institutional action a_i, but the agent is not authorized to perform a_i, neither the "counts-as" relation nor the effects of a_i take place; (ii) through a commitment not to perform such an action: in this case, if the action is authorized, its effects do take place but the corresponding commitment is violated. The solution to block the effects of certain actions by changing their authorizations during the life of the system is adopted for instance in AMELI (an infrastructure that mediates agent interactions by enforcing institutional rules) by means of *governors* [6], which filter the agents' actions letting only the allowed actions to be performed. However, this solution is not feasible when more than one institution contributes to the definition of an interaction system, as it happens for example when the Dutch Auction and the Auction-House institutions contribute to the specification of an interaction system as presented in [26] and briefly recalled in Section 6. In such cases, an action authorized by an institution cannot be voided by another institution, which can at most prohibit it.

It is moreover important to remark that in an open system, where heterogeneous agents interact exhibiting self-interested behavior based on a hidden utility function, it is impossible to predict at design phase all the interesting and fruitful behaviors that may emerge. To reach an optimal solution for all participants [28] it may be profitable to allow agents to violate their obligations and prohibitions.

We therefore conclude that regimenting an artificial system so that violations of commitments are completely avoided is often impossible and sometimes even detrimental, since it may preclude interesting evolutions of the system towards results that are impossible to foresee at design time. It is also true, however, that in order to make the evolution of the system at least partially predictable, misbehavior must be reduced to a minimum. But then, how is it possible to deter agents from violating commitments? An operational proposal to tackle this problem, based on the notion of sanction, is described in the following sections.

4 Sanctions

In this section we briefly discuss the crucial role played by *sanctions* in the specification of an open interaction system. In the Merriam-Webster On Line Dictionary[1] a sanction is defined as "the detriment, loss of reward, or coercive intervention annexed to a violation of a law as a means of enforcing the law". In an artificial system, even if the utility function of the misbehaving agent is not known, sanctions can be mainly devised to deter agents from misbehaving bringing about a loss for them in case of violation, under the assumption that the interacting heterogeneous agents are human beings or artificial agents able to reason on sanctions. Moreover sanctions can be devised to compensate the institution or other damaged agents for their loss due to the misbehavior of the agents; to contribute to the security of the system, for example by prohibiting misbehaving agents to interact any longer with the system; and to specify the acts that have to be performed to recover the system from an undesirable state [23].

When thinking about sanctions from an operational point of view, and in particular to the set of actions that have to be performed when a violation occurs, it is important to distinguish between two types of actions that differ mainly as far as their actors are concerned:

- One crucial type of action that deserves to be analyzed in detail, and that is not taken into account in other proposals [12, 19, 24], consists of the actions that the misbehaving agent itself has to perform when a violation occurs, and that are devised as a deterrent and/or a compensation for the violation. For instance, an unruly agent may have to pay a fine or compensate another agent for the damage. When trying to model this type of action it is important to take into account that it is also necessary to check that the compensating actions are performed and, if not, to sanction again the agent or, in some situations, to give it a new possibility to remedy the situation.
- Another type is characterized by the actions that certain agents are *authorized* to perform only against violations. In other existing proposals, for instance [19, 24], which do not highlight the notion of authorization (or power [15]), those actions are simply the actions that certain agents are obliged to perform against violations. From our point of view, instead, the obligation to sanction a violation should be distinguished from the authorization to do so. The reason why authorizations are crucial is obvious: sanctions can only be issued by agents playing certain specific roles in an institution. But an authorization does not always carry an obligation with it.

In some situations, and in particular when the sanction is crucial for the continuation of the interaction, one may want to express the obligation for authorized agents to react to violations by defining an appropriate new norm. For instance, in the organization of a conference if a referee does not meet the deadline for submitting a review, the organizers are not only authorized, but also obliged to reassign the paper to another referee. The norm that may be introduced to

[1] <http://www.m-w.com>

oblige the agents entitled to do so to manage the violation is similar to the "enforcement norm" proposed in [19]: it has to be activated by a violation and its content has to coincide with the sanctions of the violated obligation or prohibition. This norm may in turn be violated, and it is up to the designer of the system to decide when to stop the potentially infinite chain of violations and sanctions, leaving some violation unpunished.

Regarding this aspect, to make it reasonable for certain agents (or for their owner) to interact with an open system, it has to be possible to specify that certain violations will definitely be punished (assuming that there are not software failures). One approach is to specify that the actor of the actions performed as sanctions for those violations is the *interaction-system* itself, that therefore needs to be represented in our model as a "special agent". By "special" we mean that such an agent will not be able to take autonomous decisions, and will only be able to follow the system specifications that are stated before the interaction starts. We call this type of agents *heteronomous* (as opposite to autonomous). Given that the *interaction-system* can become, in an actual implementation, the actor of numerous actions performed as sanctions, it would be better to implement it in a distributed manner in order to avoid that it becomes a possible bottleneck.

Examples of reasonable sanctions that can be inflicted by means of norms in an open artificial system are the decrement of the trust or reputation level of the agent (similar to the reduction of the driving licence points that is nowadays applied in some countries), the revocation of the authorization to perform certain actions or a change of role (similar to confiscation of the driving licence) or, as a final action, the expulsion of the agent from the system. Another type of sanction typical of certain contracts (i.e., sets of correlated commitments created by performing certain communicative acts) is the authorization for an agent to break its part of the contract, without incurring a violation, if the counterpart has violated its own commitments.

5 Norms

Norms are taken as a specification of how a system ought to evolve. In an open system, they are necessary to impose obligations and prohibitions to the interacting agents, in order to make the system's evolution at least partially predictable [2, 20]. In particular, norms can be used to express interaction protocols as exemplified in [10, 26], where the English Auction and the Dutch Auction are specified by indicating what agents can do, cannot do, and have to do at each state of the interaction.

At design time, the main point is to guarantee that the system has certain crucial properties. This result can be achieved by formalizing obligations and prohibitions by means of logic and applying model checking techniques as studied in [17, 25]. At run time, and from the point of view of the interacting agents, norms can be used to reason about the relative utility of future actions [18]. Still at run time, but from the point of view of the open interaction system, norms can be used to check whether the agents' behavior is compliant with the specifications and able to suitably react to violations. Our model of norms is mainly suited for the last task.

Coherent with other approaches [1, 5, 12, 19, 24], in our view norms have to specify who is affected by them, who is the creditor, what are the actions that should or should not be performed, when a given norm is active, and what are the consequences of violating norms. For instance, a norm of a university may state that a professor has to be ready to give exams any day from the middle to the end of February, otherwise the dean is authorized to lower the professor's public reputation level.

In the definition of our model it is crucial to distinguish between the definition of a construct for the specification of norms in the design phase, that will be used by human designers, and the specification of how such a construct will evolve during the run-time phase to make it possible to detect and react to norm violations. In particular we assume that during the run-time of the system the interacting agents cannot create new norms, but can create new commitments, directed to specific agents, by performing suitable communicative acts, for example by making promises or by giving orders.

During the phase of specification of the set of norms of a certain artificial institution the designer does not know the actual set of agents that will interact with the system at a given time. In this phase it is therefore necessary to define norms based on the notion of role. Moreover, the time instant at which a norm becomes active is typically not known at design time, being related to the occurrence of certain events; for example, the agent playing the role of the auctioneer in an English auction is obliged to declare the current ask-price after receiving each bid by a participant. Therefore at design phase it is only possible to specify the type of event that, if it happens, will activate the norm.

During the system run time such a construct of norm, expressed in terms of roles and times of events, must be transformed into an unambiguous representation of the obligations and prohibitions that every agent has at every state of the interaction. To tackle this problem we propose to use Event-Condition-Action (ECA) rules to transform the norms given in the design phase into concrete commitments, whose operational semantics is given in our previous work [10] and will be extended in Section 5.2. The main advantage of using commitments to express active obligations and permissions is that the same construct is also used in our model of institutions to express the semantics of numerous communicative acts [10]. Interacting agents may therefore be designed to reason on just one construct to make them able to reason on all their obligations and prohibitions, derived both from norms and from the performance of communicative acts.

5.1 The Construct of Norm

First of all a norm is used to impose a certain behavior on certain agents in the system. Therefore a norm is applied to a set of agents, identified by means of the *debtors* attribute, on the basis of the roles they play in the system.

Another fundamental component of a norm is its *content*, which describes the actions that the debtors have to perform (if the norm expresses an obligation) or not to perform (if the norm expresses a prohibition) within a specified interval of time. In our model *temporal propositions*, which are defined by the Basic Institution (for a detailed treatment see [8]), are used to represent the content of commitments and, due to the strict connection between commitments and norms, are also used to represent the content of norms. A temporal proposition binds a *statement* about a state of affairs or about the performance of an action to a specific *interval of time* with a certain *mode* (that can be ∀ or ∃). Temporal propositions are represented with the following notation:

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

where the *truth-value* could be undefined ($\perp$), true or false. In particular when the *statement* represents the performance of an action and the *mode* is ∃, the norm is an obligation and the debtors of the norms have to perform the action within the interval of time. When the *statement* represents the non-performance of an action and the *mode* is ∀ the norm is a prohibition and the debtors of the norms should not perform the action within the interval of time. In particular t_{start} is always equal to the time of occurrence of the event that activates the norm. Regarding the verification of prohibitions, in order to be able to check that an action has not been performed during an interval of time it is necessary to rely on the closure assumption that if an action is not recorded as happened in the system, then it has not happened.

A norm becomes active when the *activation event* e_{start} happens. Activation can also depend on some Boolean *conditions*, that have to be true in order that the norm can become active; for instance an auctioneer may be obliged to open a run of an auction at time t_{start} if at least two participants are present.

An agent can reason whether to fulfil or not to fulfil a norm on the basis of the sanctions/reward (as discussed later) and of whom is the *creditor* of the norm, as proposed also in [16, 19]. For example, an agent with the role of auctioneer may decide to violate a norm imposed by the auction house if it is in conflict with another norm that regulates trade transactions in a certain country. Moreover the creditor of a norm is crucial because, given that it becomes the creditor of the commitments generated by the norm (as described in next section), is the only agent authorized to cancel such commitment [10]. In particular the cancelation of the commitment generated by the activation of a norm coincides with the operation of *exempting* an agent from obeying the norm in certain circumstances. Like for the *debtors* attribute, it is useful to express the creditor of declarative norms by means of their role. For instance, a norm may state that an employee is obliged to report to his director on the last day of each month; this norm will become active on the last day of each month and will be represented by means of a set of commitments, each having an actual employee as the debtor, and the employee's director as creditor.

Sometimes it may be useful to take the creditor of norms to be an *institutionalized agent*, that typically represents a human organization, like a university, a hospital, or a company, which can be regarded as the creditors of their bylaws.

In the human world, an institutionalized agent is an abstract entity that can perform actions only through a human being, who is its legal representative and has the right *mandate* [21]. On the contrary, in an artificial system it is always possible to create an agent that represents an organization but can directly execute actions. Therefore we prefer to view an institutionalized agent as a special role that can be assigned to one and only one agent having the appropriate authorizations, obligations, and prohibitions.

In order to enforce norms it is necessary to specify sanctions. More precisely, as discussed in the previous section, it is necessary to specify what actions have to be performed, when a violation occurs, by the debtors of a norm and by the agent(s) in charge of norm enforcement. These two types of actions, that we respectively call *a-sanctions* (active sanctions) and *p-sanctions* (passive sanctions) are sharply dissimilar, and thus require a different treatment. More specifically, to specify an *a-sanction* means to describe an action that the violator should perform in order to extinguish its violation; therefore, an *a-sanction* can be specified through a temporal proposition representing an action. On the contrary, to specify a *p-sanction* means to describe what actions the norm enforcer is authorized to perform in the face of a violation; therefore, a *p-sanction* can be specified by representing a suitable set of authorizations.

Regarding *a-sanctions*, it is necessary to consider that a violating agent may have more than one possibility to extinguish its violation. For example, an agent may have to pay a fine of x euro within one month, and failing to do so may have to pay a fine of $2 * x$ euro within two months. In principle we may regard the second sanction as a compensation for not paying the first fine in due time, but this approach would require an unnecessarily complex procedure of violation detection. Given that any Boolean combination of temporal propositions is still a temporal proposition, and that the truth-value of the resulting temporal proposition can be obtained from the truth-values of its components using an extended truth table to manage the indefinite truth-value [7], a more viable solution consists in specifying every possible action with a different temporal proposition, and combining them using the *OR* operator.

In summary, in our model the construct of norm is characterized by the following attributes having the specified domains:

debtors:	*role*;
creditor:	*role*;
content:	*temporal proposition*;
e_{start}:	*event-template*;
conditions:	*Boolean expression*;
a-sanctions:	*temporal proposition*;
p-sanctions:	*authorization*;

5.2 Commitments with Sanctions

In order to give an intuitive operational semantics to the construct of norms introduced so far, we now describe a mechanism to transform them, at run time, into *commitments* relative to specific agent and time interval. The transformation of norms defined at design time in commitments at run time is crucial because they are the mechanisms used to detect and react to violations. Moreover given that the activation event of norms may happen more than once in the life of the system, it is possible to distinguish between different activations and, in case, violations of the same norm. Given that our previous treatment of *commitment* [8, 10] does not cover sanctions, in this section we extend it to cover this aspect.

In our model a special institution, the Basic Institution, defines the construct of commitment, which is represented with the following notation:

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

The content of commitments is expressed using *temporal propositions* (briefly recalled in Section 5). The *state* of a commitment, as described in Figure 1, can change as an effect of the execution of institutional actions (solid lines) or of environmental events (dotted lines). Relevant events for the life cycle of commitments are due to the change of the truth-value of the commitment's content. If the content becomes true an event-driven routine (as discussed in detail in [26]) automatically changes the commitment's state to fulfilled, otherwise it becomes violated. In particular the *unset* state is used to represent commitments created by means of a request communicative act and that have not been already accepted by their debtor.

In our view an operational model of sanctions has to specify how to detect: (i), that a commitment has been violated (a mechanism already introduced in our model of commitment); (ii), that the debtor of the violated commitment performs the compensating actions; and (iii), that the agents entitled to enforce the norms have managed the violation by performing certain actions.

Regarding the necessity to check that the debtor performs the compensating actions, one solution may be to create a new commitment to perform those actions. But this solution brings in the problem of taking trace that the violation of a given commitment is extinguished by the fulfillment of another commitment. A simpler and more elegant solution consists in adding two new attributes, *a-sanctions* and *p-sanctions*, to commitments, and two new states, *extinguished* and *irrecoverable*, to their life-cycle. The value of the *a-sanctions* attribute is a temporal proposition describing the actions that the debtor of the commitment has to perform, within a given interval of time, to remedy the violation. If the actions indicated in the *a-sanctions* attribute are performed, the truth-value of the related temporal proposition becomes true and an event driven routine automatically changes the state of the violated commitment to *extinguished*, as reported in Figure 1. Analogously, if the debtor does not perform those actions, at the end of the specified time interval the truth-value of the temporal proposition becomes false and the state of the commitment becomes *irrecoverable*. Similarly to what we did for norms, the actions that certain agents are authorized to perform against the violation of the commitment are represented in the *p-sanctions* attribute. Note that whether such actions are or are not performed does not affect the life cycle of the commitment; this depends on the fact that

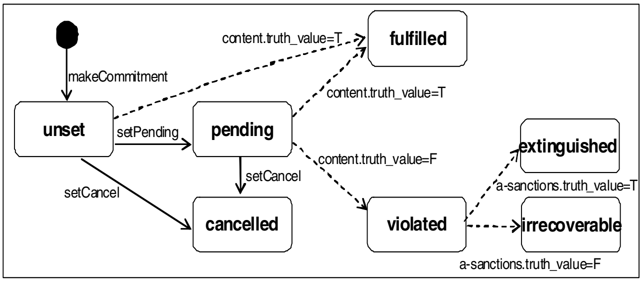

Fig. 1. The life-cycle of commitments

the agent that violated a commitment cannot be held responsible for a possible failure of other agents to actually carry out the actions they are authorized to perform.

Finally, for proper management of violation it may be necessary to trace the source of a commitment, either deriving it from the activation of a norm or from the performance of a communicative act. In order to represent this aspect we add to commitments an optional attribute called *source*. Our enriched notion of commitment is therefore represented with the following notation:

$Comm(state, debtor, creditor, content, a\text{-}sanctions, p\text{-}sanctions, source)$.

In our model we use *ECA-rules* (Event-Condition-Action rules), inspired by Active Database models, to specify that certain *actions* are executed when an event identified by an *event-templates* happens, provided that certain Boolean *conditions* are true. The semantics of ECA rules is given as usual: when an event matching 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 satisfied, the set of actions are executed and their effects are brought about in the system. The *interaction-system* agent (see Section 4) is the actor of the actions performed by means of ECA-rules, and has to have the necessary authorization in order to perform them. In our model, ECA rules are specified according to the following notation:

```
on e: event-template
if condition then
  do action(parameters)+
```

In particular the following two ECA-rules have to be present in every interaction system. One in necessary to transform at run time norms into commitments: when the activation event of the norm happens, the *makePendingComm* institutional action is performed and creates a pending commitment for each agent playing one of the roles specified in the *debtors* attribute of the norm:

```
on $e_{start}$
if $norm.conditions$ then
 do foreach $agent \mid agent.role$ in $norm.debtors$
   do $makePendingComm(agent, norm.creditor, norm.content$
      $norm.a\text{-}sanctions, norm.p\text{-}sanctions, norm\text{-}ref)$
```

The other is necessary to give the authorizations expressed in the *p-sanctions* attributes to the relevant agents when a commitment is violated:

```
on $e$: $AttributeChange(comm.state, violated)$
if $true$ then
 do foreach $auth$ in $comm.p\text{-}sanctions$
   do $createAuth(auth.role, auth.iaction)$
```

The *createAuth(role,iaction)* institutional action creates the authorization for the agents playing a certain role to perform a certain institutional action. We assume that the *interaction-system* (the actor of ECA-rules) is always authorized to create new authorizations. A similar ECA-rule has also to be defined to remove such authorizations once *iaction* has been performed.

In certain systems, to guarantee that the *interaction-system* actually performs the actions specified in the *p-sanctions* attribute, it is also possible to create the following ECA-rule that reacts to commitment violations performing those actions:

```
on $e$: $AttributeChange(comm.state, violated)$
if $true$ then
  do foreach $auth$ in $comm.p\text{-}sanctions$
       if $auth.role = interaction\text{-}system$
       do $auth.iaction(parameters)$
```

6 Example

An interesting example that highlights the importance of a clear distinction between permission and authorization, which becomes relevant when more than one institution is used to specify the interaction system, is the specification of the Dutch Auction as discussed in [26].

One of the norms of the Dutch Auction obliges the auctioneer to declare a new ask-price (within λ seconds) lowering the previous one by a certain amount κ, on condition that δ seconds have elapsed from the last declaration of the ask-price without any acceptance act from the participants. If the auctioneer violates this norm the interaction-system is authorized to declare the ask-price and to lower the auctioneer's public reputation level (obviously there is no need of an authorization to change a private reputation level), while the auctioneer has to pay a fine (within h seconds) to extinguish its violation. Such a norm can be expressed in the following way:

$debtors =$ $auctioneer;$
$creditor =$ $auction\text{-}house;$
$content =$ $TP(setAskPrice(DutchAuction.LastPrice\text{-}\kappa),$
$[time\text{-}of(e_{start}), time\text{-}of(e_{start}) + \lambda], \exists, \bot);$
$e_{start} =$ $TimeEvent(DutchAuction.timeLastPrice + \delta);$
$conditions =$ $DutchAuction.offer.value = null;$
$a\text{-}sanctions = TP(pay(ask\text{-}price, interaction\text{-}system),$
$[time\text{-}of(e), time\text{-}of(e) + h], \exists, \bot);$
$p\text{-}sanctions = Auth(interaction\text{-}system, setAskPrice(DutchAuction.LastPrice\text{-}\kappa)),$
$Auth(interaction\text{-}system, ChangeReputation(auctioneer, value));$

where variable e refers to the event that happens if the commitment generated at run-time by this norm is violated.

At the same time, the seller of a product can fix the minimum price (*minPrice*) at which the product can be sold, for example by means of an act of proposal [7]. The auction house, by means of its auctioneer, sells the product in a run of the Dutch Auction where the auctioneer is authorized to lower the price to a predetermined *reservation price*. The reservation price fixed by the auction house can be lower than *minPrice*, for example because in previous runs of the auction the product remained unsold. If the auctioneer actually sells the product at a price (*winnerPrice*) lower than *minPrice*, the sale is valid but the auction house violates its commitment with the seller of the product and will incur the corresponding sanctions; for example, it may have to refund the seller, while the seller is authorized to lower the reputation of the auction house. This situation can be modelled by the following commitment between the seller and the auction house:

$state =$ $pending;$
$debtor =$ $auction\text{-}house;$
$creditor =$ $seller;$
$content =$ $TP(not\ setCurPrice(p) \mid p < minPrice, [now, +\infty)], \forall, \bot);$
$a\text{-}sanctions = TP(pay(seller, minPrice\text{-}winnerPrice),$
$[time\text{-}of(e), time\text{-}of(e) + 15 days], \exists, \bot);$
$p\text{-}sanctions = Auth(seller, ChangeReputation(auction\text{-}house, value));$

where variable e refers to the event that happens if the commitment is violated.

7 Conclusions

In this paper we have discussed the importance of formalizing and enforcing obligations and prohibitions in the specification of open interaction frameworks. We have proposed a construct to define norms in the design of institutions expressed in terms of roles and event times. The operational semantics of norms is defined by the commitments they generate through ECA-rules.

The innovative aspects of our proposal are the definition of different types of sanctions and of the operational mechanisms for monitoring the behavior of the agents and reacting to commitment violations. In particular, an interesting feature of our proposal is that the construct of commitment is uniformly used to model the semantics of communicative acts and of norms. Differently from [19] our model of norms specifies the interval of time within which norms are active. Thanks to their transformation into commitments, it is possible to apply certain norms (whose activation event may happen many times) more than once in the life of the system. Another crucial aspect of our norms is that, differently from [19], they are activated by the occurrence of events and not simply if a certain state holds. Regarding the treatment of sanctions our model is more in-depth with respect to other proposals [13, 19, 24] because we distinguish the actions of the debtors from the actions of the other agents that are entitled to react to violations. In particular, regarding the actions of the debtors, we propose an effective solution for managing multiple sanctions, that is, multiple possibilities to compensate the violation (for example, paying an increasing amount of money), without entering in an infinite loop of checking violations and applying punishments. Regarding the sanctions applied by other agents, we discussed the reasons why a norm expresses what actions are authorized against violations and the reasons why some norms may be enforced by the interaction-system itself, which is treated as a special heteronomous agent.

Acknowledgements

We would like to thank Bertil Cottier, professor of Law at Università della Svizzera italiana, for helping us to improve our knowledge on the legal aspects of sanctions.

References

1. Artikis, A., Sergot, M., Pitt, J.: Animated Specifications of Computational Societies. In: Castelfranchi, C., Johnson, W.L. (eds.) Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002), pp. 535–542. ACM Press, New York (2002)
2. Barbuceanu, M., Gray, T., Mankovski, S.: Coordinating with obligations. In: Sycara, K.P., Wooldridge, M. (eds.) Proceedings of the 2nd International Conference on Autonomous Agents (Agents 1998), pp. 62–69. ACM Press, New York (1998)
3. Boella, G., van der Torre, L.: Contracts as legal institutions in organizations of autonomous agents. In: Dignum, V., Corkill, D., Jonker, C., Dignum, F. (eds.) Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004), pp. 948–955. IEEE Computer Society, Los Alamitos (2004)
4. Cliffe, O., Vos, M.D., Padget, J.: Specifying and Reasoning About Multiple Institutions. In: Noriega, P., Vázquez-Salceda, J., Boella, G., Boissier, O., Dignum, V., Fornara, N., Matson, E. (eds.) COIN 2006. LNCS (LNAI), vol. 4386, pp. 67–85. Springer, Heidelberg (2007)
5. Esteva, M., Padget, J., Sierra, C.: Formalizing a language for institutions and norms. In: Meyer, J.-J.C., Tambe, M. (eds.) ATAL 2001. LNCS, vol. 2333, pp. 348–366. Springer, Heidelberg (2002)

6. Esteva, M., Rodríguez-Aguilar, J.A., Rosell, B., Arcos, J.L.: AMELI: An Agent-based Middleware for Electronic Institutions. In: Jennings, N.R., Sierra, C., Sonenberg, L., Tambe, M. (eds.) Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004), pp. 236–243. ACM Press, New York (2004)
7. Fornara, N.: Interaction and Communication among Autonomous Agents in Multiagent Systems. PhD thesis, Faculty of Communication Sciences, University of Lugano, Switzerland (2003), `http://doc.rero.ch`
8. Fornara, N., Colombetti, M.: A commitment-based approach to agent communication. Applied Artificial Intelligence an International Journal 18(9-10), 853–866 (2004)
9. Fornara, N., Viganò, F., Colombetti, M.: Agent communication and institutional reality. In: van Eijk, R., Huget, M., Dignum, F. (eds.) AC 2004. LNCS, vol. 3396, pp. 1–17. Springer, Heidelberg (2005)
10. Fornara, N., Viganò, F., Colombetti, M.: Agent communication and artificial institutions. Autonomous Agents and Multi-Agent Systems 14(2), 121–142 (2007)
11. Fornara, N., Viganò, F., Verdicchio, M., Colombetti, M.: Artificial institutions: A model of institutional reality for open multiagent systems. Artificial Intelligence and Law 16(1), 89–105 (2008)
12. Garcia-Camino, A., Noriega, P., Rodriguez-Aguilar, J.A.: Implementing norms in electronic institutions. In: Proceedings of the 4th International Joint Conference on Autonomous agents and Multi-Agent Systems (AAMAS 2005), pp. 667–673. ACM Press, New York (2005)
13. Grossi, D., Aldewereld, H., Dignum, F.: Ubi lex, ibi poena: Designing norm enforcement in e-institutions. In: Noriega, P., Vázquez-Salceda, J., Boella, G., Boissier, O., Dignum, V., Fornara, N., Matson, E. (eds.) COIN 2006. LNCS (LNAI), vol. 4386, pp. 101–114. Springer, Heidelberg (2007)
14. Hart, H.L.A.: The Concept of Law. Clarendon Press, Oxford (1961)
15. Jones, A., Sergot, M.J.: A formal characterisation of institutionalised power. Journal of the IGPL 4(3), 429–445 (1996)
16. Kagal, L., Finin, T.: Modeling Conversation Policies using Permissions and Obligations. In: van Eijk, R., Huget, M., Dignum, F. (eds.) AC 2004. LNCS, vol. 3396, pp. 123–133. Springer, Heidelberg (2005)
17. Lomuscio, A., Sergot, M.: A formulation of violation, error recovery, and enforcement in the bit transmission problem. Journal of Applied Logic (Selected articles from DEON 2002 - London) 1(2), 93–116 (2002)
18. López y López, F., Luck, M., d'Inverno, M.: Normative Agent Reasoning in Dynamic Societies. In: Jennings, N.R., Sierra, C., Sonenberg, L., Tambe, M. (eds.) Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004), pp. 535–542. ACM Press, New York (2004)
19. López y López, F., Luck, M., d'Inverno, M.: A Normative Framework for Agent-Based Systems. In: Proceedings of the First International Symposium on Normative Multi-Agent Systems, Hatfield (2005)
20. Moses, Y., Tennenholtz, M.: Artificial social systems. Computers and AI 14(6), 533–562 (1995)
21. Pacheco, O., Carmo, J.: A Role Based Model for the Normative Specification of Organized Collective Agency and Agents Interaction. Autonomous Agents and Multi-Agent Systems 6(2), 145–184 (2003)
22. Searle, J.R.: The construction of social reality. Free Press, New York (1995)

23. Vázquez-Salceda, J., Aldewereld, H., Dignum, F.: Implementing Norms in Multiagent Systems. In: Lindemann, G., Denzinger, J., Timm, I.J., Unland, R. (eds.) MATES 2004. LNCS (LNAI), vol. 3187, pp. 313–327. Springer, Heidelberg (2004)
24. Vázquez-Salceda, J., Dignum, V., Dignum, F.: Organizing multiagent systems. Autonomous Agents and Multi-Agent Systems 11(3), 307–360 (2005)
25. Viganò, F.: A Framework for Model Checking Institutions. In: Edelkamp, S., Lomuscio, A. (eds.) MoChArt IV. LNCS, vol. 4428, pp. 129–145. Springer, Heidelberg (2007)
26. Viganò, F., Fornara, N., Colombetti, M.: An Event Driven Approach to Norms in Artificial Institutions. In: Boissier, O., Padget, J., Dignum, V., Lindemann, G., Matson, E., Ossowski, S., Simao Sichman, J., Vázquez-Salceda, J. (eds.) ANIREM 2005 and OOOP 2005. LNCS (LNAI), vol. 3913, pp. 142–154. Springer, Heidelberg (2006)
27. Yolum, P., Singh, M.: Reasoning about commitment in the event calculus: An approach for specifying and executing protocols. Annals of Mathematics and Artificial Intelligence 42, 227–253 (2004)
28. Zambonelli, F., Jennings, N.R., Wooldridge, M.: Developing multiagent systems: The Gaia methodology. ACM Transactions on Software Engineering and Methodology (TOSEM) 12(3), 317–370 (2003)