Specifying and Enforcing Norms in Artificial Institutions

Nicoletta Fornara [a] Marco Colombetti [ab]

[a] *Università della Svizzera italiana, via G. Buffi 13, 6900 Lugano, Switzerland*
[b] *Politecnico di Milano, piazza Leonardo Da Vinci 32, Milano, Italy*

Abstract

In this paper we investigate two important and 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 to 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, norms are specified declaratively. When certain events take place, norms become active and generate pending commitments for the agents playing certain roles. Norms also specify the sanctions associated to 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 [9, 26, 10] 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 taking 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 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 has 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 operational specification of the normative component of artificial institutions. We will develop our *OCeAN* metamodel by dealing with the problems of giving a declarative specification of norms for open systems and of devising efficient and complete computational mechanisms for managing norms. In particular we aim at automating the detection of, and reaction to, the violations of norms. An important feature of our framework ,with respect to other proposals [1, 4, 12, 19, 24] is that it gives a uniform solution to two crucial problems: the specification of norms and the definition of the semantics of an Agent Communication Language: indeed, our model of norms relies on the notion of commitment [7], that has been previously introduced to express the meaning of a library of communicative acts [8]. We analyze in details the

problem of defining a mechanism for enforcing obligations and prohibitions by means of sanctions, that is, a treatment of the actions 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", like the others do, 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 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; Vázquez-Salceda et al. [24] present, in the OMNI framework, a method to enforce norms described at different level of abstraction; and 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.[4, 12] propose ISLANDER, where a normative language with sanctions is defined but not discussed in details, Boella et al. [3] model violations but does 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 our metamodel for artificial institutions. In Section 3 the reasons why in open interaction framework 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 our previous construct of commitment is extended by adding the treatment of sanctions. In Section 6 we exemplify our proposal and finally in Section 7 we draw some conclusions.

2 The OCeAN model

Our metamodel of artificial institutions consists mainly of the following components:

- The constructs necessary to define the *core ontology* of an institution, including: the notion of an *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 an *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 and that are used in the definition of other constructs: the notions of a *role* and of an *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.

- A *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* that specify what agents are authorized to perform an institutional actions.

- The construct of *norm*, used to impose obligations and prohibitions to perform certain actions on agents interacting with the system. In our model, as it will be described in Section 5, we have *declarative* norms that, when their activating event happens, are transformed into their operational counterpart, that is, a *commitment*.

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 [9], 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 happens 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 adopting what in the literature is called "regimentation" [14], as proposed in [13], by introducing a control mechanism that does not allow to violate commitments?

To answer this question, it is useful to distinguish between *natural* (or physical) actions (like opening a door or physically delivering a product), whose effects take place thanks to nonconventional 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).

Regarding physical actions, it is important to remark that they cannot be regimented since, after they have been performed, they cannot be considered "void", that is, their effects cannot be annulled. Therefore it is impossible to use regimentation to prevent the violation of a prohibition to perform a given physical action.

Concerning institutional actions, the choice to allow for commitment violations or to impose regimentation is different in the case of obligations or prohibitions.

Prohibitions can be expressed using two different mechanisms: (i) through the absence 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 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* [5], 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 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 annulled by another institution, which at most can prohibit it.

As regard as obligations, there is only one way to "regiment" the performance of an obliged action, that is, by making the system performing the obliged action in place 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 an 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.

Finally it is 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 [27] 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 devised: (i) 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; (ii) to compensate the institution or other damaged agents for their loss due to the misbehavior of the agents; (iii) to contribute to the security of the system, for example by prohibiting misbehaving agents to interact any longer with the system; (iv) 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 among two types of actions that differ mainly as far as their actors are concerned.

One crucial type of actions that deserves to be analyzed in detail, and that is not taken into account in other proposals [19, 24, 12], consists of the actions that the misbehaving agent itself has to perform against a violation, 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 actions 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.

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 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 introduced to 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 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). Note that the 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.

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

Example 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 final action, the expulsion 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 in a violation, if the counterpart has violated its own commitments.

5 Norms

In an open system, norms are necessary to impose obligations and prohibitions to the interacting agents, in order to make the systems evolution at least partially predictable [2, 20]. In particular, norms can be used to express interaction protocols as exemplified in [9, 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. In this section we propose a development of the model of norms that we have presented in our previous works [9, 26, 10], which clearly separates the *declarative* form of norms from their *operational* counterpart, that is, *commitment*, and from the procedure to transform the former into the second.

Norms are taken as a specification of how a system ought to evolve. 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 on 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 to suitably react to violations. Our model of norms is mainly suited for the last task.

Coherently with other approaches [4, 1, 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, and what are the consequences of violating them. 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 will lower the professors level of trust.

From the point of view of the specification of a system, and in particular of its set of norms, it is crucial to abstract away from the actual set of agents that are interacting with the system at a given time, a result that can be achieved by using the notion of role in the definition of norms. 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. Finally, norms must produce an unambiguous representation of the obligations and prohibitions that every agent has at every state of the interaction. For these reasons we propose a declarative description of norms expressed in terms of roles and time of events, which at run time can generate commitments relative to specific agents and time intervals. The main advantage of using commitments to express active obligations and permissions is that the same construct used to represent the activation of declarative norms is also used in our model of institutions to express the semantics of numerous communicative acts [9]. 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 Declarative norms

First of all a norm is used to impose a certain behavior to 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 [7]), 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 $\forall$ or $\exists$). 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 $\exists$, 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 $\forall$ the norm is a prohibition and the debtors of the norms should not perform the action within the interval of time. The time interval of the content is strictly connected to norms activation and deactivation events, that are described later on. In particular t_{start} is always equal to the time of occurrence of the event that activates the norm, and t_{end} is equal to the time of occurrence of the event that deactivates 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 and becomes inactive when the *deactivation event* e_{end} takes place. 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 weather to fulfil or not to fulfill a norm on the basis of the sanctions (as discussed later) and of who 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. The creditor of a declarative norm, 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 [9]. In particular the operation of cancelling 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 employees director as the 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 organizations 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 *d-sanctions* (debtors sanctions) and *e-sanctions* (enforcers sanctions) are sharply dissimilar, and thus require a different treatment. More specifically, to specify a *d-sanction* means to describe an action that the violator should perform in order to extinguish its violation; therefore, a *d-sanction* can be specified through a temporal proposition representing an action. On the contrary, to specify an *e-sanction* means to describe what actions the norm enforcer is authorized to perform in front of a violation; therefore, an *e-sanction* can be specified by representing a suitable set of authorizations.

Regarding *d-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 [6], a more viable solution consists in specifying every possible actions with a different temporal proposition, and combining them using the OR operator.

In summary, in our model norms are characterized by the following attributes having the specified domains:

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

5.2 Commitments with Sanctions

In order to give an intuitive operational semantics to the declarative representation of norms introduced so far, we now describe an operational mechanism to transform them into their operational counterpart, that is, into *commitments* relative to specific agent and time interval. The transformation of declarative norms in commitments is crucial in the actual evolution of the system 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* [7, 9] 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 5.1). The *state* of a commitment can change as an effect of the execution of institutional actions or of environmental events. Relevant events for the life cycle of commitments are due to the change of the truth-value of the commitments content: if the content becomes true the commitment becomes fulfilled, otherwise it becomes violated as described in Figure 1.

In our view an operational model of sanctions has to specify how to detect that a commitment has been violated, that the debtor of the violated commitment performs the compensating actions and that the agents entitled to enforce the norms have managed the violation performing certain actions.

In our model, when the content of a commitment becomes false an event-driven routine (that as discussed in [26] can be implemented applying the *observer pattern* [11]) automatically changes the commitments state to violated. 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. A simpler and more elegant solution consists in adding two new attributes, *d-sanctions* and *e-sanctions*, to commitments, and two new states, *extinguished* and *irrecoverable*, to their life-cycle. The value of the *d-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 *d-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*. The actions that

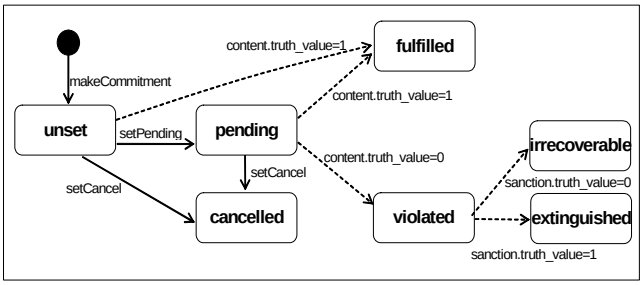

Figure 1: The life-cycle of commitments.

the agents entitled to do so are authorized to perform against the violation of the commitment are represented in the *e-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 the agent that violated a commitment cannot be held responsible of a possible failure of other agents to actually carry out the actions they are authorized to perform.

Finally, for a proper management of violation it may be necessary to trace the source of a commitment, either deriving from the activation of a norm or from the performance of a communicative acts. 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:

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

To transform our declarative norms into commitments we adopt *ECA-rules* (Event-Condition-Action rules). An ECA-rule executes certain *actions* when an event identified by an *event-templates* happens, provided that certain Boolean *conditions* are true; 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.

The following ECA-rule transforms norms into commitments: when the activation event (e_{start}) 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 $makePendingComm(agent, norm.creditor, norm.content,$
      $norm.d\text{-}sanctions, norm.e\text{-}sanctions, norm\text{-}ref)$

When a commitment is violated, another ECA-rule gives the authorizations expressed in the *e-sanctions* attributes to the relevant agents:

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

To guarantee that the *interaction-system* actually performs the actions specified in the *e-sanctions* attribute, it is possible to create an ECA-rule that reacts to commitments violation performing those actions:

on e: $AttributeChange(commitment.state, violated)$
if $true$ **then**
  do foreach $auth$ **in** $commitment.e\text{-}sanctions$
    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 of a certain amount κ, on condition that δ seconds are 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 not need of an authorization to change a private reputation level), while the auctioneer has to pay a fine to extinguish its violation. Such a norm can be expressed in the following way:

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

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 [6]. 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 until 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 resulted unsold. If the auctioneer actually sells the product at a price ($winnerPrice$) lower that $minPrice$, the sale is valid but the auction house violates its commitment with the seller of the product and will incur in 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=	$TP(not\ setCurPrice(p) \mid p < minPrice, [now, +\infty)], \exists, \bot)$
d-sanctions=	$TP(pay(seller, minPrice\text{-}winnerPrice),$ $[time\text{-}of(e), time\text{-}of(e)+15days], \exists, \bot)$
e-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 normative component characterized by declarative norms, expressed in terms of roles and event times. The operational semantics of the declarative 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; thus artificial agents able to reason on commitments can deal with both ACL semantics and the normative component of the interaction system.

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 [19, 24, 13] because we distinguish the actions of the debtors from the actions of the other agents 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 in improving our knowledge on legal aspects of sanctions.

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] M. Barbuceanu, T. Gray, and S. Mankovski. Coordinating with obligations. In K. P. Sycara and M. Wooldridge, editors, *Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98)*, pages 62–69, New York, 1998. ACM Press.

[3] G. Boella and L. van der Torre. Contracts as legal institutions in organizations of autonomous agents. In V. Dignum, D. Corkill, C. Jonker, and F. Dignum, editors, *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004)*, pages 948–955, Los Alamitos, CA, USA, 2004. IEEE Computer Society.

[4] M. Esteva, J. Padget, and C. Sierra. Formalizing a language for institutions and norms. In J. J. Meyer and M. Tambe, editors, *Intelligent Agents VIII : 8th International Workshop, ATAL 2001 Seattle, WA, USA, August 1-3, 2001 Revised Papers*, volume 2333 of *LNCS*. Springer, 2002.

[5] 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.

[6] N. Fornara. *Interaction and Communication among Autonomous Agents in Multiagent Systems*. PhD thesis, Faculty of Communication Sciences, University of Lugano, Switzerland, 2003. http://doc.rero.ch.

[7] N. Fornara and M. Colombetti. A commitment-based approach to agent communication. *Applied Artificial Intelligence an International Journal*, 18(9-10):853–866, 2004.

[8] N. Fornara, F. Viganò, and M. Colombetti. Agent communication and institutional reality. In 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 LNCS 3396, pages 1–17. Springer, 2005.

[9] N. Fornara, F. Viganò, and M. Colombetti. Agent communication and artificial institutions. *Autonomous Agents and Multi-Agent Systems*, preprint:http://dx.doi.org/10.1007/s10458–006–0017–8, 2006.

[10] N. Fornara, F. Viganò, M. Verdicchio, and M. Colombetti. Artificial institutions: A model of institutional reality for open multiagent systems. Technical Report 4, Institute for Communication Technologies, Università della Svizzera Italiana, 2006.

[11] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns*. Addison Wesley, 1995.

[12] A. Garcia-Camino, P. Noriega, and J. A. Rodriguez-Aguilar. Implementing norms in electronic institutions. In *Proceedings of te 4th International Joint Conference on Autonomous agents and Multi-Agent Systems (AAMAS 2005)*, pages 667–673, New York, NY, USA, 2005. ACM Press.

[13] D. Grossi, H. Aldewereld, and F. Dignum. Ubi lex, ibi poena: Designing norm enforcement in e-institutions. In V. Dignum, N. Fornara, and P. Noriega, editors, *Proceedings of the AAMAS06 Workshop Coordination, Organization, Institutions and Norms in Agent Systems (COIN)*, pages 107–120, 2006.

[14] H. L. A. Hart. *The Concept of Law*. Clarendon Press, Oxford, 1961.

[15] A. Jones and M. J. Sergot. A formal characterisation of institutionalised power. *Journal of the IGPL*, 4(3):429–445, 1996.

[16] L. Kagal and T. Finin. Modeling Conversation Policies using Permissions and Obligations. In R. van Eijk, M. Huget, and F. Dignum, editors, *Developments in Agent Communication*, volume 3396 of *LNCA*, pages 123–133. Springer, 2005.

[17] A. Lomuscio and M. Sergot. A formulation of violation, error recovery, and enforcement in the bit transmission problem. *Journal of Applied Logic (Selected articles from DEON02 - London)*, 1(2):93–116, 2002.

[18] 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.

[19] F. López y López, M. Luck, and M. d'Inverno. A Normative Framework for Agent-Based Systems. In *Proceedings of the First International Symposium on Normative Multi-Agent Systems, Hatfield*, 2005.

[20] Y. Moses and M. Tennenholtz. Artificial social systems. *Computers and AI*, 14(6):533–562, 1995.

[21] O. Pacheco and J. Carmo. 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] J. R. Searle. *The construction of social reality*. Free Press, New York, 1995.

[23] J. Vázquez-Salceda, H. Aldewereld, and F. Dignum. Implementing Norms in Multiagent Systems. In I. G. Lindemann, J. Denzinger, I. J. Timm, and R. Unland, editors, *Multiagent System Technologies: Second German Conference (MATES 2004)*, volume 3187 of *LNAI*, pages 313–327, Berlin, Germany, 2004. Springer Verlag.

[24] J. Vázquez-Salceda, V. Dignum, and F. Dignum. Organizing multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 11(3):307–360, Nov 2005.

[25] F. Viganò. A Framework for Model Checking Institutions. In *Proceedings of the ECAI Workshop on Model checking and Artificial Intelligence (MOCHART IV)*, pages 31–46, 2006. To appear. Available from: www.istituti.usilu.net/viganof.

[26] F. Viganò, N. Fornara, and M. Colombetti. An Event Driven Approach to Norms in Artificial Institutions. In O. Boissier, J. Padget, V. Dignum, G. Lindemann, E. Matson, S. Ossowski, J. Simao Sichman, and J. Vázquez-Salceda, editors, *Coordination, Organization, Institutions and Norms in Multi-Agent Systems I, Proc. ANIREM'05/OOOP'05, Utrecht, The Netherlands, July 2005*, volume LNAI 3913, pages 142–154. Springer Berlin, 2006.

[27] F. Zambonelli, N. R. Jennings, and M. Wooldridge. Developing multiagent systems: The Gaia methodology. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 12(3):317–370, 2003.