Towards a Model of the Dynamics of Normative Multi-Agent Systems

Fabiola López y López Michael Luck

Department of Electronics and Computer Science
University of Southampton
Southampton, SO17 1BJ, UK
{flyl00r,mml}@ecs.soton.ac.uk

Abstract. For agents, one of the advantages of being in a society is the satisfaction of those goals whose success depends on the abilities of other agents. In turn, societies are controlled by norms and, consequently, agents must be able first to model the society in which they exist, and then to identify the different relationships, due to norms, in which they might be involved in order to act appropriately. Both of these could mean the difference between the success or failure of their goals. To this end, this paper focuses on the identification of the basic components of norm-based systems, and on representing and analysing the *dynamic* relationships between member agents which result from the processing of norms.

1 Introduction

Studying the effects of the incorporation of norms in agents and multi-agent systems is far from trivial. Research on norms and agents has ranged from fundamental work on the importance of norms in agent behaviour [7, 21], to proposing internal representations of norms [5, 6], analysing the different types of norms [9, 22, 23], considering their emergence in groups of agents [24], proposing logics for their formalisation [19, 25], and both analysing and representing institutions controlled by norms [1, 4, 17]. Norms can also be analysed from the internal point of view of agents and the role agents play in their processing. In this case, we can describe for example, how agents manage norm adoption and compliance [2, 10, 14], how agents responsible for enforcing norms must behave [3], and what the characteristics are of agents entitled to exert power in a society [12]. However, there are other aspects that have received little attention from the agent research community. In particular, little work has been done to explain, on the one hand, the relationships between agents that result from the issue of, compliance with, or violation of norms, and on the other, how an agent's decision-making process is modified as a result of these relationships.

For agents, one of the advantages of being in a society is the satisfaction of those goals whose success depends on the abilities of other agents. Societies in turn, are controlled by norms, and consequently, agents must be able first to model the society in which they exist, and then to identify the different relationships, due to norms, in which they might be involved in order to act appropriately. We argue that the correct identification of such relationships may be the difference between the success or failure of an agent's goals. For example, to select a plan, agents take into account not only their own obligations and prohibitions but also those of other agents. The identification of norm relationships also helps to avoid conflicts of authority among the members of a

system. For instance, norm defenders must be able to identify both when and to whom a punishment must be applied, and in turn, norm breakers need to know who is entitled to exert such a punishment, and under which circumstances it can be applied. So, addressing these aspects of norms is crucial for the modeling of different systems controlled by norms where norm compliance is not always expected, and also for the modeling of those agents able to exist in such a system.

Towards the main objective of having a framework that allows the representation of both norm-based systems and agents able to deal with norms, the aims of this paper are twofold. First, it aims to analyse the properties of norms, normative agents, and normative multi-agent systems in order to identify the basic components that should be considered in their representation, and second, it aims to describe the *dynamics* of norms and how, from the different stages in which norms are processed, different relations among agents can be identified. Besides the informal description, formal specifications of the main concepts and processes are given in order to avoid any ambiguity arising through the use of informal natural language. In particular, this avoids inconsistencies which might complicate the use and correct implementation of the theoretical framework provided. In this document, first a general structure of a norm, and the basic characteristics of normative agents are discussed in Section 2. Then, in Section 3 the main components of a multi-agent system controlled by norms are analysed, and their formalisation is provided. In the same section, some roles for agents, due to norms, are identified. After that, the changes that occur in a system when norms are issued, complied with, or violated are described and then formalised by following the different stages in the processing of norms (Section 4). Finally, in Section 5 a set of normative relationships between agents is provided, before to present our conclusions.

2 Norms and Normative Agents

In this section we describe the basic blocks from which to build up our theory of normative multi-agent systems. This conceptual infrastructure provides the basis for a broad theory, and underpins several aspects not included in this paper, but described elsewhere [14–16]. As a means to building up a formal model of a normative agent without being repetitive, we adopt the SMART *agent framework* described in [11]. In addition, in what follows, we also adopt the Z specification language to construct such a formal model, because Z schemas allow, among other facilities, an easy transition from specifications to programmes. A Z schema contains two parts: the declaration part which declares local variables, and the predicate part which expresses some properties of the values of these variables. Z is based on set-theory and first order logic, with details available in [20]. For brevity, however, we will not elaborate the use of Z further.

2.1 Agents

In the SMART *agent framework*, an *attribute* represents a perceivable feature of the agent's environment which, here, is represented as a predicate or its negation. Then, a particular *state* in the environment is described by a set of attributes, a *goal* represents situations that an agent wishes to bring about, *motivations* are desires or preferences that affect the outcome of the reasoning intended to satisfy an agent's goals, and *actions* are discrete events that change the state of the environment when performed. For the

purposes of this paper, we formally describe attributes, environmental states, goals and actions. Details of the remaining elements are not needed, so we simply consider them as given sets.

$$[Predicate, Motivation]$$

$$Attribute ::= pos\langle\langle Predicate\rangle\rangle \mid not\langle\langle Predicate\rangle\rangle$$

$$EnvState == \mathbb{P}_1\, Attribute \qquad Goal == \mathbb{P}_1\, Attribute$$
$$Action == EnvState \rightarrow EnvState$$

In addition, an entity is described by a non-empty set of attributes representing its permanent features, a set of goals that it wants to bring about, a set of capabilities that it is able to perform, and a set of motivations representing its preferences. Moreover, *agents* are entities whose set of goals is not empty, and *autonomous agents* are agents with non-empty sets of motivations. By omitting irrelevant details, we formalise them as follows.

Agent

$$capabilities : \mathbb{P}\, Action$$
$$goals : \mathbb{P}\, Goal$$
$$motivations : \mathbb{P}\, Motivation$$
$$beliefs : \mathbb{P}_1\, Attribute$$

$$goals \neq \varnothing$$

$$AutonomousAgent \mathrel{\widehat=} [Agent \mid motivations \neq \varnothing]$$

2.2 Norms

An agent may have access to certain norms which are represented as data structures relating to social rules. These may be common to all agents (such as with a mutually understood social law) or only available to some. *Norms* are mechanisms that a society has in order to influence the behaviour of agents within it. Norms can be created from different sources, varying from built-in norms to simple agreements between agents, or more complex legal systems. They may persist during different periods of time; for example until an agent dies, as long as an agent remains in the society for which the norms were issued, or just for a short period of time until normative goals become satisfied. There are different aspects that can be used for characterizing them. First, norms are always prescribed to be complied with by a set of *addressee* agents in order to *benefit* another set of agents (possibly empty). They specify something that ought to be done, and consequently they include *normative goals* that must be satisfied by addressees. Sometimes, these normative goals must be directly intended, whereas other times their role is to inhibit specific goals (as in the case of prohibitions). Second, norms are not always applicable, and their activation depends on the *context* in which agents are situated; there may be *exceptions* when agents are not obliged to comply with the norm. Finally, in some cases, norms suggest the existence of a set of *sanctions* or *punishments* to be imposed when agents do not satisfy the normative goals, and a set of *rewards* to be received when agents do. Both, punishment and rewards, are also

represented as goals that must be satisfied by someone else. Thus, the general structure of a norm can be formalised as follows. (Note that we specify normative goals as a set, to allow for the possibility of multiple goals in a norm, though we recognise that this will typically be a singleton set.)

$$
\begin{array}{|l}
\hline
Norm \\
addressees, beneficiaries : \mathbb{P}\, Agent \\
normativegoals, rewards, punishments : \mathbb{P}\, Goal \\
context, exceptions : EnvState \\
\hline
addressees \neq \varnothing \\
context \neq \varnothing \\
\hline
\end{array}
$$

Norms can be divided, without eliminating the possibility of having further categories, into four types: *obligations*, *prohibitions*, *social commitments* and *social codes*. Roughly, we can say that *obligations* and *prohibitions* are norms adopted once an agent becomes a member of a society, *social commitments* are norms derived from agreements or negotiations between two or more agents, and *social codes* are norms motivated by feelings such as love, pity, friendship, or social conformity. It is not the purpose of this paper to discuss the different categories of norms; consequently, in the remainder of this paper we will use the term *norm* as an umbrella term to cover every type of norm. However, we argue that all of them share the same structure. An important consideration at this point is that we understand *prohibitions* as norms whose normative goals must be avoided by addressee agents.

2.3 Normative Agents

Moreover, a normative agent is an autonomous agent whose behaviour is shaped by the obligations it must comply with, prohibitions that limit the kind of goals that it can pursue, social commitments that have been created during its social life and social codes which may not carry punishments, but whose fulfillment is a means of being identify as part of a community.

$$
\begin{array}{|l}
\hline
NormativeAgent \\
AutonomousAgent \\
norms : \mathbb{P}\, Norm \\
\hline
\end{array}
$$

2.4 Permitted and Forbidden Actions

Sometimes, it is useful to observe norms not through the normative goals that ought to be achieved, but through the actions that can lead to the satisfaction of such goals. Then, we can talk about actions that are either *permitted* or *forbidden* by a norm as follows. If there is a situation state which makes a norm become activated, and the results of an action benefit the achievement of the associated normative goal, then such an action is *permitted* by the respective norm. For example, if the normative goal of a norm is to have taxes paid then the action *paying taxes* is a permitted action if it changes an agent's situation of having taxes unpaid into a situation where taxes are paid. By analogy, we can define *forbidden* actions as those actions leading to a situation which contradicts or

hinders the normative goal. For example, the action *illegal parking* is a forbidden action by a norm whose normative goal is to avoid parking in front of a hospital entrance. In general, it is not trivial to observe how the results of an action might benefit or hinder the achievement of normative goals. For example, if we spend all our money and after that we try to pay our taxes, it might be not obvious that spending money may hinder our normative goal of paying taxes. To avoid drilling down into the intricate details of this important but somewhat secondary concern in relation to the focus of this paper, the associations between situation states that might either *benefit* or *hinder* goals are taken for granted and formalised as follows.

$$benefited\_ : \mathbb{P}(EnvState \times Goal)$$
$$hindered\_ : \mathbb{P}(EnvState \times Goal)$$

Now, we define two relations that hold among an action and a norm, which either permit or forbid the action, as follows.

$$permitted\_ : \mathbb{P}(Action \times Norm)$$
$$forbidden\_ : \mathbb{P}(Action \times Norm)$$

$$\forall\, a : Action;\ n : Norm;\ \bullet$$
$$permitted\ (a, n) \Leftrightarrow (\exists\, g : n.normativegoals \bullet$$
$$benefited\ (a\ n.context, g)) \wedge$$
$$forbidden\ (a, n) \Leftrightarrow (\exists\, g : n.normativegoals \bullet$$
$$hindered\ (a\ n.context, g))$$

In other words, if an action is applied in the context of a norm, and the results of this action benefit the normative goals, then the action is permitted, otherwise the action is forbidden.

3 Normative Multi-Agent Systems

Norms cannot be studied independently of the system for which they were created. Consequently, before describing how many processes due to norms are triggered, an analysis of the main components of a social system regulated by norms must be provided. A *normative multi-agent system* can be defined as a set of *normative* agents, which are controlled by a set of common *norms* ranging from obligations and social commitments, to social codes. This *control* can be observed in three different aspects.

- First, member agents must recognise themselves as part of the society.
- Second, complete control cannot be exerted if sanctions or incentives are not applied when norms are either violated or complied with.
- Third, changes in current normativity must be allowed as a way to solve unpredictable conflicts between agents and norms, or both.

Each one of these aspects is discussed in the subsections below.

3.1 Membership of Normative Societies

The performance of every structure of control relies on the capabilities of its members to recognise and follow its norms. However, given agents autonomy, fulfillment of norms can never be taken for granted. In fact, autonomous agents decide whether to comply

with norms based on their own current goals and motivations [14]. It is also possible that not all the norms that one agent has adopted belong to just one system because agents indeed can be part of more than one society at the same time. In addition, due to agent limitations, it is possible that not all the norms of the system can be known by any agent. These characteristics can be formally expressed by saying that the set of norms adopted by any member is not necessarily a subset of the norms of the system, and also that the intersection of both sets of norms is not empty (see Fig. 1). Now, part of being member of a society means that agents are subject to some of the norms in the system. In other words, the set of addressee agents of every norm must be included in the set of member agents, because it does not make any sense to have norms addressed to nonexistent agents (see formalisation in Subsection 3.4).

Fig. 1. An Agent's Norms

3.2 Enforcement and Encouragement of Norm Compliance

Complete control cannot be exerted if, for each norm in the system, there is no other norm that prescribes how some agents have to react when the original norm becomes unfulfilled [18]. For example, if there is an obligation to pay accommodation fees for all students in a university, there must also be a norm stating what hall managers must do when a student does not pay them. These kinds of norms are regularly called *secondary norms* because they are addressed to a specific group of agents responsible for punishing non-compliance with primary norms. It is only through these norms that some agents are entitled to punish other agents. Chaos might emerge in a society if such responsibility is given either to no one or to anyone. Addressee agents of this kind of norms are frequently called the *defenders* of a norm.

Describing these secondary norms in terms of the structure of a norm that has been proposed in Section 2, we observe that the violation of a norm can be detected by an agent when it realises that the associated normative goals were not satisfied. Once this event becomes identified by defenders, their duty is then to start a process in which rebellious agents can be punished. Consequently, the state representing the non-satisfaction of a normative goal must be included in the *context* of a secondary norm, because it is a condition to trigger that norm. Moreover, every punishment included in the unfulfilled norm must appear in the normative goals of the secondary norm, that is, defenders of norms must have the goal of punishing every offender of a norm. Fig. 2 shows how both, the structure of a norm and the norm which enforces it, are related. To formalise this kind of norms, some definitions are needed first. We say that a norm can be considered as *fulfilled* in a specific state of the system if its corresponding normative goals are a logical consequence of such a state.

$$logicalconsequence\_ : \mathbb{P}(EnvState \times EnvState)$$

$$fulfilled\_ : \mathbb{P}(Norm \times EnvState)$$

$$\forall\, n : Norm;\ st : EnvState \bullet$$
$$fulfilled\ (n, st) \Leftrightarrow (\forall\, g : n.normativegoals \bullet logicalconsequence\ (st, g))$$

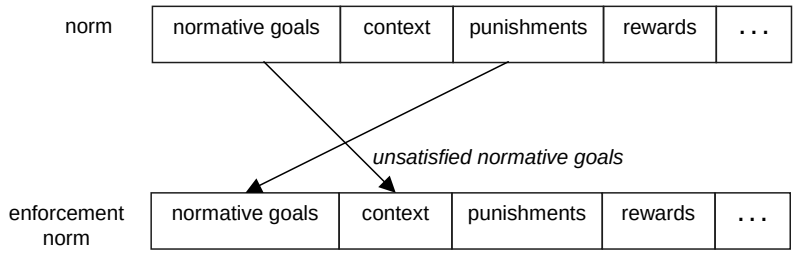

Fig. 2. Enforcement Norm Structure

Now, a relationship between a norm directed to control the behaviour of agents and a secondary norm can be defined as follows. A norm *enforces* another norm through punishments if the first norm is activated when the normative goal of the second becomes unfulfilled, and the punishments associated with the unfulfilled norm are part of the normative goals of the enforcement norm. We call these kinds of norms as *enforcement* norms.

$$enforce\_ : \mathbb{P}(Norm \times Norm)$$

$$\forall\, n_1, n_2 : Norm \bullet$$
$$enforce\ (n_1, n_2) \Leftrightarrow (\neg\, fulfilled(n_2, n_1.context) \wedge$$
$$n_2.punishments \subseteq n_1.normativegoals)$$

So far we have described secondary norms in term of punishments because punishments are one of the more commonly used mechanisms to enforce compliance with norms. However, a similar analysis can be done for secondary norms corresponding to the process of rewarding members doing their duties. The relations between norms and norms to reward their compliance are shown in Fig. 3. Formally we say that a norm *encourages* the compliance with another norm through rewards if the first norm is activated when the second norm becomes fulfilled, and the rewards associated with the fulfilled norm are part of the normative goals of the encourage norm. We call these kinds of norms as *reward* norms.

$$reward\_ : \mathbb{P}(Norm \times Norm)$$

$$\forall\, n_1, n_2 : Norm \bullet$$
$$reward\ (n_1, n_2) \Leftrightarrow (fulfilled(n_2, n_1.context) \wedge$$
$$n_2.rewards \subseteq n_1.normativegoals)$$

Now, an important point to mention here is that this way of representing enforcement norms can create an infinite chain of norms because we would also have to define norms to use when authorities or defenders do not comply with their obligations to either punish those agents breaking the rules or reward those agents who fulfill their

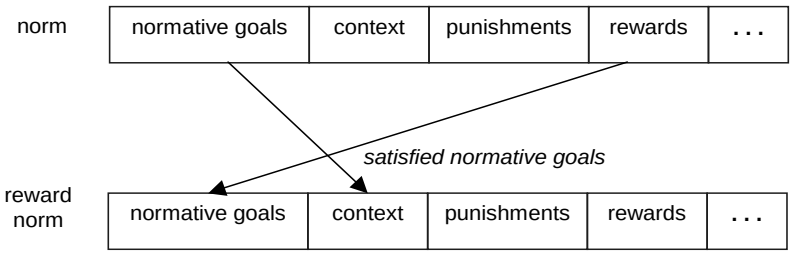

Fig. 3. Reward Norm Structure

responsibilities [18]. To avoid this chain of norms, and by taking the risk of being considered as absolutist, in what follows we consider that no punishments are applied when a enforcement norm is not fulfilled. That means, that neither authorities nor defenders can be judged (at least in this normative system) by dismissing their responsibilities. A similar reasoning for reward norms can be done. However, if a system requires it, our model and formalisation for enforcing and encouraging norms can be used recursively as necessary. There is nothing in the definition of the model itself to prevent this.

3.3 Dynamic Normativity and Legislation

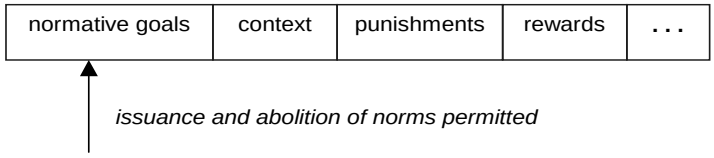

Fig. 4. Legislation Norm Structure

In general, norms are introduced into a society as a means to achieve social order. Some of them are intended to avoid conflicts between agents, others allow the establishment of commitments, and there also exist norms intended to unify the behaviour of agents as a way of social identification. However, neither all conflicts nor all commitments can be anticipated and controlled by norms. Consequently, in a *dynamic* multiagent system there must exist the possibility of creating new norms, modifying existing norms, or even abolishing those which become obsolete. Now, although it is possible that many of the members of a society have capabilities to do this, these actions must be restricted to be carry out by a particular set of agents in a particular situation in order to avoid that anyone can impose its norms, because some conflicts of interest might emerge. In other words, norms stating when actions to legislate are permitted must be also included [12]. These norms are called *legislation* norms, and they must specify that actions to issue and abolish norms are only permitted to a particular set of agents represented by its addressees (see Fig. 4). These constraints are specified in the following declaration.

$$legislate\_ : \mathbb{P}\, Norm$$

$$\forall\, n : Norm \bullet legislate\ (n) \Leftrightarrow (\exists\, issuingnorms, abolishnorms : Action \bullet permitted\ (issuingnorms, n) \vee permitted\ (abolishnorms, n))$$

3.4 Formal Model

All elements discussed above are now incorporated into the formal representation of a *normative multi-agent system* which schema is presented below.

$$
\begin{array}{|l}
\hline
NormativeMAS \\
members : \mathbb{P}\, NormativeAgent \\
normsNMAS : \mathbb{P}\, Norm \\
enforcenorms : \mathbb{P}\, Norm \\
rewardnorms : \mathbb{P}\, Norm \\
legislationnorms : \mathbb{P}\, Norm \\
\hline
\forall\, ag : members \bullet ag.norms \cap normsNMAS \neq \varnothing \\
\forall\, rg : normsNMAS \bullet rg.addressees \subseteq members \\
\forall\, en : enforcenorms \bullet (\exists\, n : normsNMAS \bullet enforce\ (en, n)) \\
\forall\, rn : rewardnorms \bullet (\exists\, n : normsNMAS \bullet reward\ (rn, n)) \\
\forall\, ln : legislationnorms \bullet legislate\ (ln) \\
\hline
\end{array}
$$

That is, a normative multi-agent system comprises the following elements: a set of members agents able to reason about norms, a set of norms directed to regulate the behaviour of these agents, and represented here by the variable $normsNMAS$, a set of norms directed to enforce and judge the latter set of norms ($enforcenorms$), the set of norms directed to encourage compliance with norms through rewards ($rewardnorms$), and the norms issued to allow the creation and abolition of norms ($legislationnorms$). In the schema, the first predicate states that all members must have adopted some of the norms of the normative multi-agent system, and the second makes explicit that addressee agents of this set of norms must be members of the system. The last three predicates describe respectively the structure of enforcement, reward and legislation norms. Notice that whereas every enforcement norm must have a norm to enforce, not every norm may have a corresponding enforcement norm, which means that no one in that society is legally entitled to punish an agent which does not fulfill such a norm.

3.5 Normative Roles

Defining a normative multi-agent system in this way allows the identification of different roles for agents. These roles depend on the kind of norms of which agents are responsible. Some of them are listed below:

- The set of agents who are entitled to create, modify, or abolish the set of norms of a society. No other members of the society are endowed with the power and authority to do so. This kind of agents can, in turn, be either elected or decreed, and we call them *legislators*.
- An agent is an *defender* if it is directly responsible for either applying punishments or giving rewards.
- The main responsibility of *police* agents is to monitor compliance with norms. They always watch the behaviour of other agents in order to detect transgressions and also to enforce the norms to be complied with.

– *Addressee* agents are directly responsible for the achievement of normative goals.
– Finally, *beneficiaries* are agents whose goals can be benefited when a normative goal becomes satisfied.

These *normative roles* for agents are not mutually exclusive. In fact, agents are able to have more than one normative role at the same time, depending on the kind of norm being considered. For example, in a social commitment, the beneficiary agent can be police and consequently encourage the fulfillment of a norm, or it can be a defender and either apply sanctions or give the rewards agreed in the past. In an office, the manager can be both a legislator, and then impose his own norms, and a defender entitled to punish his employees. The more complex a society is, the more elaborate these normative roles become, and in some cases all legislators, authorities (judges), and police make a complex structure of control generally named *government* with its own legal norms directed to control the rest of the society. For the purpose of this work, police agents are also grouped in what we call *defender* agents, with the authority to require compliance with norms, and either to give rewards or to apply punishments. In fact, being a defender is a relationship that holds between an agent and the enforcement norm that entitles it to defend the norm. Similarly, being a legislator means that there exists a norm that entitles an agent to modify the current legislation by creating new norms and abolishing some of the norms already created. Considering defenders and legislators in this way allows us to represent the fact that all these elements cannot be taken independently of each other, but are somehow complementary. All these characteristics are described in the following specification. First, a function to cast a normative agent as an agent is introduced due to type compatibility. After that, the set of relations for normative roles are given.

$$
\begin{array}{l}
theagent : NormativeAgent \rightarrow Agent \\
canpunish\_ : \mathbb{P}(NormativeAgent \times Norm \times NormativeMAS) \\
canreward\_ : \mathbb{P}(NormativeAgent \times Norm \times NormativeMAS) \\
isdefender\_ : \mathbb{P}(NormativeAgent \times Norm \times NormativeMAS) \\
islegislator\_ : \mathbb{P}(NormativeAgent \times NormativeMAS) \\
\hline
\forall\, ag : NormativeAgent;\ n : Norm;\ nmas : NormativeMAS \bullet \\
\quad canpunish\ (ag, n, nmas) \Leftrightarrow (n \in nmas.normsNMAS \wedge \\
\qquad (\exists\, en : Norm \bullet (en \in nmas.enforcenorms \wedge \\
\qquad\quad theagent\ ag \in en.addressees \wedge enforce\ (en, n)))) \wedge \\
\quad canreward\ (ag, n, nmas) \Leftrightarrow (n \in nmas.normsNMAS \wedge \\
\qquad (\exists\, en : Norm \bullet (en \in nmas.enforcenorms \wedge \\
\qquad\quad theagent\ ag \in en.addressees \wedge reward\ (en, n)))) \wedge \\
\quad isdefender\ (ag, n, nmas) \Leftrightarrow (canpunish\ (ag, n, nmas) \vee \\
\qquad canreward\ (ag, n, nmas)) \wedge \\
\quad islegislator\ (ag, nmas) \Leftrightarrow (\exists\, ln : Norm \bullet \\
\qquad ln \in nmas.legislationnorms \wedge theagent\ ag \in ln.addressees)
\end{array}
$$

That is, the first and second relations state who are the agents entitled to either punish or reward a norm in a specific normative multi-agent system. The third relation specifies which agents can be considered as defenders of a particular norm. Finally, the fourth relation states who a legislator is in the system being considered.

4 Dynamics of Norms

Norms are not a static concept. Once they are included in a system, they cause certain behaviour in each one of its agent members. In Figure 5, the different processes through which a norm passes since it is created until it becomes abolished can be observed. Arrows represent the transitions between one stage of norms to another. That is, first a legislator issues a norm. After that, the norm is spread by either indirect or direct communication. Then, adoption of norms takes place. Through this process an agent expresses its willingness to follow the norm as a way of being part of the society. Once a norm is adopted, it remains inactive, or in latency, until the applicability conditions are satisfied. Agents, in exception states, are not obligated to comply with norms, and consequently norms can be dismissed. However, in the majority of the cases, two different situations might occur after a norm becomes activated, a norm is either fulfilled or unfulfilled by addressee agents. After a norm is complied with, a reward could be offered. By contrast, if the norm is violated there are two possibilities: either a sanction is applied or it is not. Finally, as time progresses, some norms become either abolished or modified.

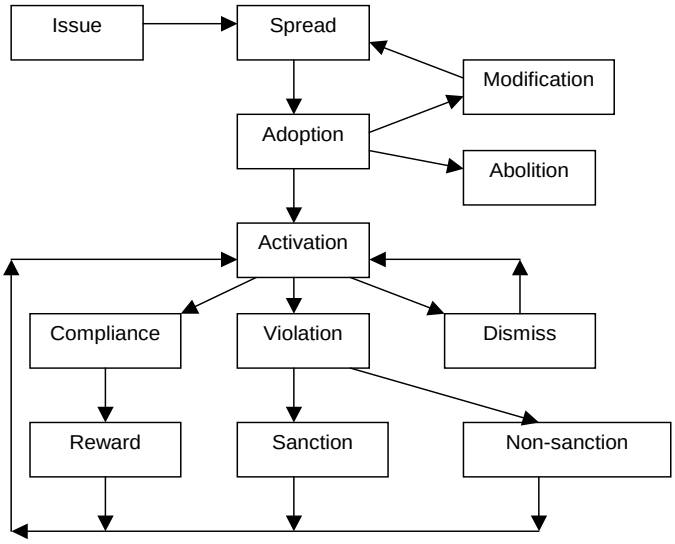

Fig. 5. Norm Dynamics

Considering the *dynamics* that result from norms is an important issue that deserve our attention, because interesting relations among agents can be identified in each one of them (see Section 5). In turn, according to these relationships different reactions of agents are expected. For instance, when a norm is activated, defenders are just entitled to require its fulfillment. However, if the norm becomes unfulfilled, defenders are now entitled to apply punishments. Consequently, we argue that the correct identification of the different stages of a norm is a key point to model the normative behaviour of

agents. In the following subsections, the transitions between these different stages are described and formalised from the point of view of an external observer.

4.1 Changing Legislation

Legislation of norms is a responsibility only attributed to legislator agents. Such a responsibility comprises at least three functions, namely: issuance, abolition, and modification of norms. Unfortunately, due to their complexity, details of how such functions are carried out cannot be given here. However, we can still introduce two functions to identify both all recently created norms ($newnorms$), and all norms that must be abolished ($obsoletenorms$). These functions might be equivalent to ask legislators about the result of their assigned tasks. In turn, modification of norms can be seen as the abolition of a subset of norms together with the issuance of another subset of norms with the same name, consequently a specific function to do that is not included here. Now, after new norms are created and others are abolished, spreading and updating of norms are needed. As a result of these changes at global level, the set of member agents must also change. That is, some of these norms become internally either adopted or abolished by addressee agents. This is represented by the functions $spreadnorms$ and $abolishnorms$ which can be seen as the processes through which agents are notified of the creation of new norms and the abolition of norms that become obsolete. The schema $NormLegislation$ formalises the functions associated with the legislation of norms. In it, the variable $nmas$ represents the normative multi-agent system in which changes in legislation occur.

$$
\begin{array}{|l}
\textit{NormLegislation} \\
\hline
nmas : NormativeMAS \\
legislators : \mathbb{P}\, NormativeAgent \\
newnorms : \mathbb{P}\, NormativeAgent \rightarrow \mathbb{P}\, Norm \\
obsoletenorms : \mathbb{P}\, NormativeAgent \rightarrow \mathbb{P}\, Norm \\
spreadnorms : (\mathbb{P}\, NormativeAgent \times \mathbb{P}\, Norm) \rightarrow \mathbb{P}\, NormativeAgent \\
abolishnorms : (\mathbb{P}\, NormativeAgent \times \mathbb{P}\, Norm) \rightarrow \mathbb{P}\, NormativeAgent \\
\hline
\forall\, ag : legislators \bullet islegislator\ (ag, nmas) \\
\mathrm{dom}\ newnorms = \mathbb{P}\, legislators \\
\mathrm{dom}\ obsoletenorms = \mathbb{P}\, legislators \\
\hline
\end{array}
$$

Now, the process that changes norms in both the system and all its members can be represented as follows.

$$
\begin{array}{|l}
\textit{ChangeLegislation} \\
\hline
\Delta NormLegislation \\
\hline
nmas'.normsNMAS = nmas.normsNMAS \setminus \\
\qquad obsoletenorms\ legislators\ \cup newnorms\ legislators \\
nmas'.members = spreadnorms\ (abolishnorms\ (nmas.members, \\
\qquad obsoletenorms\ legislators), newnorms\ legislators) \\
\hline
\end{array}
$$

The first predicate states that the set of norms, after a change in legislation, is composed of all the old norms except those recently abolished, joined with all norms recently created. The second predicate represents how all members are informed of legislation changes through a composition of functions. That is, first members are informed about norms that must be abolished because they are now considered as obsolete, and after that they receive information about all norms recently created.

4.2 Normative Multi-Agent System State

After norms are issued, spread, and then adopted, they enter into a circle in which different agents intervene. To capture the different stages in which a norm is processed, we specify the *state* of a normative multi-agent system as follows.

$$
\begin{array}{|l}
\hline
NMASState \\
\hline
NormativeMAS \\
currentsituation : EnvState \\
formeractivenorms : \mathbb{P}\, Norm \\
activenorms : \mathbb{P}\, Norm \\
fulfillednorms : \mathbb{P}\, Norm \\
unfulfillednorms : \mathbb{P}\, Norm \\
punishednorms : \mathbb{P}\, Norm \\
rewardednorms : \mathbb{P}\, Norm \\
\hline
activenorms \subseteq normsNMAS \\
\hline
\end{array}
$$

That is, in a particular instant of time some of the norms become $activated$, that means that the conditions under which a norm must be fulfilled are satisfied. Moreover, other previously activated norms become either $fulfilled$ or $unfulfilled$. Furthermore, some of the unfulfilled norms become $punished$, and some of the fulfilled ones become $rewarded$. Identifying these stages of norms is important because any change in them makes some other agents react. For example, addressee agents acquire new responsibilities because of active norms, and they deserve to be rewarded or punished due to fulfilled or unfulfilled norms respectively. In addition some agents might require compliance with active norms, or apply punishment to addressees of unfulfilled norms, etc. In the schema which represents the state of a normative multi-agent system the $formeractivenorms$ variable keeps the norms that were activated in a previous period of time, and the $currentsituation$ represents the state of the general environment.

4.3 Assessing Compliance with Norms

Although, not all norms change their stage at the same time, we take a particular point in the time to assess all of them. Now as mentioned before, the easy way to know if a norm has been fulfilled is by observing the current state of the system and then verifying if the associated normative goals are satisfied or not. This form of verifying compliance with norms can be used for any kind of norm, ranging from the norms of the normative

system to the norms to enforce compliance with. These changes are represented in the schema below.

$$
\begin{array}{|l}
\hline
\textit{AssessNorm} \\
\textit{NormativeMAS} \\
\Delta \textit{NMASState} \\
\textit{observedchanges} : \textit{EnvState} \rightarrow \textit{EnvState} \\
\textit{newactive} : \mathbb{P}\, \textit{Norm} \\
\textit{newfulfilled} : \mathbb{P}\, \textit{Norm} \\
\textit{newpunished} : \mathbb{P}\, \textit{Norm} \\
\textit{newrewarded} : \mathbb{P}\, \textit{Norm} \\
\hline
\textit{currentsituation}' = \textit{observedchanges}\ \ \textit{currentsituation} \\
\textbf{let}\ \textit{newactive} == \{n : \textit{normsNMAS} \mid \\
\quad \textit{logicalconsequence}\ (\textit{currentsituation}', n.\textit{context})\} \bullet \\
\textbf{let}\ \textit{newfulfilled} == \{n : \textit{activenorms} \mid \textit{fulfilled}\ (n, \textit{currentsituation}')\} \bullet \\
\textbf{let}\ \textit{newpunished} == \{n : \textit{unfulfillednorms} \mid (\exists\, \textit{en} : \textit{enforcenorms} \bullet \\
\quad (\textit{enforcepunish}\ (\textit{en}, n) \land \textit{fulfilled}\ (\textit{en}, \textit{currentsituation}')))\} \bullet \\
\textbf{let}\ \textit{newrewarded} == \{n : \textit{fulfillednorms} \mid (\exists\, \textit{en} : \textit{enforcenorms} \bullet \\
\quad (\textit{enforcereward}\ (\textit{en}, n) \land \textit{fulfilled}\ (\textit{en}, \textit{currentsituation}')))\} \bullet (\\
\qquad \textit{formeractivenorms}' = \textit{formeractivenorms} \cup \\
\qquad\quad \textit{activenorms} \setminus \textit{newactive} \land \\
\qquad \textit{activenorms}' = \textit{newactive} \land \\
\qquad \textit{fulfillednorms}' = \textit{fulfillednorms} \cup \textit{newfulfilled} \land \\
\qquad \textit{unfulfillednorms}' = \textit{unfulfillednorms} \cup \\
\qquad\quad (\textit{activenorms} \setminus \textit{newfulfilled}) \land \\
\qquad \textit{punishednorms}' = \textit{punishednorms} \cup \textit{newpunished} \land \\
\qquad \textit{rewardednorms}' = \textit{rewardednorms} \cup \textit{newrewarded}) \\
\hline
\end{array}
$$

In this schema, *observedchanges* is a function that reports the observed changes in the social environment, and by using such changes, the sets of norms are updated. First the set of new *active* norms is calculated by analysing if the context, to trigger a norm, is a *logical consequence* of the current situation of the system. After that, the set of active norms that were *fulfilled* by their corresponding addressee agents is calculated by verifying the satisfaction of the corresponding normative goals. Next, unfulfilled norms that were *punished* are found by verifying if the norm that enforces it, has already been satisfied. Something similar is done to verify if fulfilled norms were *rewarded*. After all these steps are done, the stages of norms are updated accordingly. Note that former norms are only added to by those norms that are no longer active.

5 Normative Relationships

As we said before, norms in their different stages create different kinds of relationships among agents. We identify four sets of them. The first is created due to the authority of certain agents in the system. The next is created once norms become activated. Norms that have been complied with also generate relations among agents through the offered

rewards. Finally, unfulfilled norms and their associated punishments make agents to be related in a different way. These relationships, in turn, are used by agents when reasoning about norms is needed, and a decision must be taken. Then, by using the proposed structure of the norm, the definition of a normative multi-agent system, and the different normative roles that agents might have in it, we describe the set of relationships between agents that we are interested in. These relations are illustrated in Fig. 6 in which circles and squares represent the type of norm and the state of a norm, respectively, and hexagons symbolize the relationships created by them.

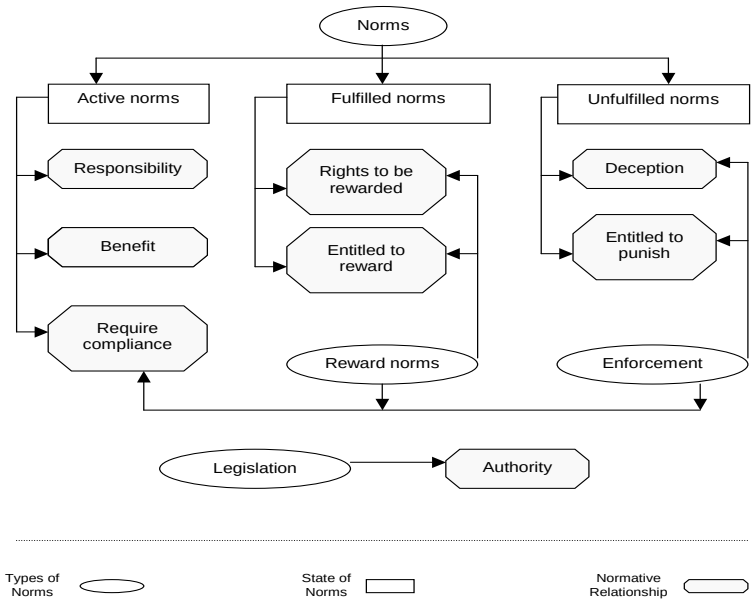

Fig. 6. Normative Relationships

5.1 Legislation Relations

As mentioned before, not all agents in a normative multi-agent system are entitled to legislate, and therefore, before a norm becomes adopted, agents must recognise the authority of the issuer, otherwise the validity of the norm could be questioned, and then rejected. Then formally we say that an agent is a *legal authority* for another agent, if it is a legislator in the normative multi-agent system which the second agent belongs to.

$$legalauthority\_ : \mathbb{P}(NormativeAgent \times NormativeAgent \times NormativeMAS)$$

$$\forall\, ag_1, ag_2 : NormativeAgent;\ nmas : NormativeMAS \bullet$$
$$legalauthority(ag_1, ag_2, nmas) \Leftrightarrow islegislator(ag_1, nmas) \wedge$$
$$ag_2 \in nmas.members$$

5.2 Active Norm Relations

Norms become *activated* when the current situations of an agent (or a group of agents) match the context in which a norm must be fulfilled. For example, if a driver wants to park its car in front of an entrance, the norm that forbids such an action is applied, otherwise agents do not need to be worried about them. From this situation four relations among agents can be inferred as follows.

It can be observed that some norms include exception states in which an agent is not obliged to respect them. An exception state could be treated as a state not included in the context of a norm, because in that case the norm would not be activated, and consequently, agents would not be obliged to comply with it. Although the results are similar, we prefer to make them explicit because it allows an agent to explain why it is not obliged to comply with that norm. This latter aspect can be useful if the norm is addressed to a set of agents, and only some of them are excepted from their responsibilities. Formally, we say that an agent *can dismiss* a norm in a particular state of the system, if that agent is an addressee of the norm, and the exception states of the norm are a logical consequence of the current state.

$$
\begin{array}{|l}
candismiss\_ : \mathbb{P}(NormativeAgent \times Norm \times EnvState) \\
\hline
\forall\, ag_1, ag_2 : NormativeAgent;\ n : Norm;\ nmas : NormativeMAS; \\
st : EnvState \bullet \\
\quad candismiss\ (ag_1, n, st) \Leftrightarrow (theagent\ ag_1 \in n.addressees \wedge \\
\qquad logicalconsequence\ (st, n.exceptions)) \\
\end{array}
$$

Another important relationship that can be observed here, is the relation between an addressee agent, a norm, and its defender. In this situation, it can be said that an agent is entitled to require compliance with norms either by threatening agents with future punishments, or by offering future rewards. Formally, it can be said that an agent *can require* another agent to fulfill a norm, if it is a designated defender in the system and the second agent is an addressee of the norm.

$$
\begin{array}{|l}
canrequire\_ : \mathbb{P}(NormativeAgent \times NormativeAgent \times Norm \\
\qquad\qquad \times NormativeMAS) \\
\hline
\forall\, ag_1, ag_2 : NormativeAgent;\ n : Norm;\ nmas : NormativeMAS; \\
st : EnvState \bullet \\
\quad canrequire\ (ag_1, ag_2, n, nmas) \Leftrightarrow (isdefender\ (ag_1, n, nmas) \wedge \\
\qquad ag_2 \in nmas.members \wedge theagent\ ag_2 \in n.addressees) \\
\end{array}
$$

Finally, there are two important relationships between agents that can be mentioned here. The first is the responsibility that an addressee agent has as soon as a norm becomes activated. Note, that although an agent has a responsibility to fulfill, it does not means that it is going to do so. That decision is only made by the agent itself. Formally, we say that an agent *has a responsibility* to another if there is a norm already addressed to the first agent, whose benefits may be enjoyed by the second.

$$
\begin{array}{|l}
hasresponsibility\_ : \mathbb{P}(NormativeAgent \times NormativeAgent \times Norm \\
\qquad \times NormativeMAS) \\
\hline
\forall\, ag_1, ag_2 : NormativeAgent;\ n : Norm;\ nmas : NormativeMAS; \\
st : EnvState \bullet \\
\quad hasresponsibility\ (ag_1, ag_2, n, nmas) \Leftrightarrow (n \in nmas.normsNMAS \wedge \\
\qquad ag_1 \in nmas.members \wedge theagent\ ag_2 \in nmas.members \wedge \\
\qquad theagent\ ag_1 \in n.addressees \wedge theagent\ ag_2 \in n.beneficiaries)
\end{array}
$$

The second relationship is its counterpart which relates to the expectations of a beneficiary agent to receive something from the responsibilities of others. Formally we say that an agent *expects benefits* from the responsibility of another agent if it is the beneficiary of a norm addressed to the second agent.

$$
\begin{array}{|l}
expectsbenefit\_ : \mathbb{P}(NormativeAgent \times NormativeAgent \times Norm \times \\
\qquad NormativeMAS) \\
\hline
\forall\, ag_1, ag_2 : NormativeAgent;\ n : Norm;\ nmas : NormativeMAS; \\
st : EnvState \bullet \\
\quad expectsbenefit\ (ag_1, ag_2, n, nmas) \Leftrightarrow (n \in nmas.normsNMAS \wedge \\
\qquad ag_1 \in nmas.members \wedge ag_2 \in nmas.members \wedge \\
\qquad theagent\ ag_2 \in n.addressees \wedge theagent\ ag_1 \in n.beneficiaries)
\end{array}
$$

5.3 Fulfilled Norm Relations

Once a norm is *fulfilled* no further action is necessary except maybe by addressee agents claiming rewards from a defender. Then two complementary relationships are identified as follows. First, we say that an agent has the responsibility of *rewarding* another agent, if the first agent is a defender of the norm and the second is an agent who has fulfilled it. In addition, an agent has the right to be *rewarded* by a defender of a norm, if the first agent has already complied with it.

$$
\begin{array}{|l}
rewards\_ : \mathbb{P}(NormativeAgent \times NormativeAgent \times Norm \times EnvState \\
\qquad \times NormativeMAS) \\
rewarded\_ : \mathbb{P}(NormativeAgent \times NormativeAgent \times Norm \times EnvState \\
\qquad \times NormativeMAS) \\
\hline
\forall\, ag_1, ag_2 : NormativeAgent;\ n : Norm;\ nmas : NormativeMAS; \\
st : EnvState \bullet \\
\quad rewards\ (ag_1, ag_2, n, st, nmas) \Leftrightarrow (theagent\ ag_2 \in n.addressees \wedge \\
\qquad ag_2 \in nmas.members \wedge canreward\ (ag_1, n, nmas) \wedge fulfilled\ (n, st)) \wedge \\
\quad rewarded\ (ag_1, ag_2, n, st, nmas) \Leftrightarrow (theagent\ ag_1 \in n.addressees \wedge \\
\qquad ag_1 \in nmas.members \wedge canreward\ (ag_2, n, nmas) \wedge fulfilled\ (n, st))
\end{array}
$$

5.4 Unfulfilled Norm Relations

By contrast, when a norm is *unfulfilled*, several events take place and therefore other kinds of relationships hold. Obviously, addressees of an unfulfilled norm will do nothing, and indeed they would prefer that their failure remains hidden, in order to avoid

facing the consequences of their actions. However, a *deception* situation emerges in which the interests of third agents, i.e. the beneficiaries, might be badly affected by the irresponsibility of offenders. Agents in this situation could claim compensation or something similar. Formally, it can be said that an agent is *deceived* by another agent if a norm was unfulfilled by the second agent, and the benefits could have been enjoyed by the first.

$$
\begin{array}{|l}
deceived\_ : \mathbb{P}(NormativeAgent \times NormativeAgent \times Norm \times EnvState) \\
\hline
\forall\, ag_1, ag_2 : NormativeAgent;\ n : Norm;\ nmas : NormativeMAS; \\
st : EnvState \bullet \\
\quad deceived\ (ag_1, ag_2, n, st) \Leftrightarrow (theagent\ ag_1 \in n.beneficiaries \land \\
\qquad theagent\ ag_2 \in n.addressees \land \neg\ fulfilled\ (n, st))
\end{array}
$$

In addition, defenders also have a different relation with addressees. When a norm becomes activated, defenders are entitled only to enforce a norm, but when the norm is broken they have the responsibility to start a sequence of events leading to punish rebellious agents. Nevertheless, it could be possible that none of the defenders realises the occurrence of these events, and consequently the rebellious agent never becomes punished. Then formally it can be said that an agent has to *punish* another agent if the first is a defender of the norm and the second one is an agent who has not fulfilled it.

$$
\begin{array}{|l}
punishes\_ : \mathbb{P}(NormativeAgent \times NormativeAgent \times Norm \times EnvState \\
\qquad \times NormativeMAS) \\
\hline
\forall\, ag_1, ag_2 : NormativeAgent;\ n : Norm;\ nmas : NormativeMAS; \\
st : EnvState \bullet \\
\quad punishes\ (ag_1, ag_2, n, st, nmas) \Leftrightarrow (isdefender\ (ag_1, n, nmas) \land \\
\qquad ag_2 \in nmas.members \land theagent\ ag_2 \in n.addressees \land \\
\qquad \neg\ fulfilled\ (n, st))
\end{array}
$$

As we can observe, all these relationships are relativised to a normative multi-agent system to which agents belong. That is, no relationships due to norms can be created when agents do not belong to the same system. To sum up, we say that in a normative multi-agent system where social control has been defined through norms, some relations can be identified. That is, in a particular point of time, there are responsibilities that agents acquire through norms, situations in which addressee agents can be excepted from such responsibilities, enforcement mechanisms that might be applied to agents with duties, rewards that must be given to respectful agents, punishments that must be applied to norm breakers, and deceived agents expecting compensations. All these relationships change as soon as new norms become activated, fulfilled or unfulfilled.

6 Conclusion

So far in our work, the basic components of a system controlled by norms have been identified. We call this kind of systems *normative multi-agent systems*, and we describe them as consisting of: a set of member agents whose compliance with norms is neither always enforced nor always expected, a set of norms directed at controlling the

behaviour of all members, a set of legal norms to enforce compliance with regulations through punishment, a set of legal norms directed to reward agents who fulfill norms, and a set of norms to entitle some agents to change the regulations. Moreover, the *dynamics* that occur in a system due to norms have been analysed, and according to the different stages in the processing of norms, some *normative relationships* have been identified. The key concept here is the normative behaviour of agents caused not only by the existence of norms, but also by their issuance, fulfillment or violation, which in turn must be the result of the decisions (autonomous or not) of each of the members.

By studying the characteristics of normative multi-agent systems, we have set up the basis for a framework to represent different kinds of social systems regulated by norms. In addition, the set of normative relationships identified in this paper might enable agents to take more effective decisions in situations where norms are involved. For example, agents who have benefited from a fulfilled norm might decide to reciprocate with the addressees of such a norm in their subsequent interactions. Normative relationships are also useful for identifying situations in which a subset of agents is legally empowered. These relations are used to inform the decision of when a new norm can be either adopted or complied with, and are the focus of the next stage in our work. We also aim to extend our work on norm compliance [14] to introduce strategies in which agents are externally influenced to comply with a norm. Additionally, we must also provide an analysis of those situations in which agents might adopt new norms. We believe that the normative roles that we have defined here can be used by agents to identify empowered agents, and therefore to identify from whom an order can be received.

Our analysis builds on much important work on norms. Ross, for example [18], describes some of the norms and relationships presented in this paper. In turn, Conte and Castelfranchi [5] have already mentioned some of the normative roles we present, and some of the processes involved in reasoning about norms. Jones and Sergot [12] also mention the characteristics of agents entitled to manage an institution. The closest work, however, is by Balzer and Toumela [1], who present the formalisation of an institution controlled by norms. However, neither of them consider the dynamics of norms nor the relationships that emerge from them, which in turn can be used by agents to reason about norms.

Acknowledgments The first author is supported by the Faculty Enhancement Program (PROMEP) of the Mexican Ministry of Public Education(SEP) and the Benemérita Universidad Autónoma de Puebla (México). Authors also express their gratitude to the valuable comments provided by the anonymous reviewers of this paper.

References

1. W. Balzer and R. Tuomela. Social institutions, norms and practices. In Dellarocas and Conte [8], pages 161–180.
2. G. Boella and L. Lesmo. Deliberative normative agents. In Dellarocas and Conte [8], pages 85–110.
3. C. Castelfranchi, R. Conte, and M. Paolucci. Normative reputation and the cost of compliance. *Journal of Artificial Societies and Social Simulation*, 1(3), 1998.

4. R. Conte. Emergent (info)institutions. *Journal of Cognitive Systems Research*, 2:97–110, 2001.
5. R. Conte and C. Castelfranchi. *Cognitive and Social Action*. UCL Press, London, 1995.
6. R. Conte and C. Castelfranchi. Norms as mental objects. From normative beliefs to normative goals. In C. Castelfranchi and J. P. Müller, editors, *From Reaction to Cognition (MAAMAW'93)*, LNAI 957, pages 186–196. Springer-Verlag, 1995.
7. R. Conte, R. Falcone, and G. Sartor. Agents and norms: How to fill the gap? *Artificial Intelligence and Law*, 7:1–15, 1999.
8. C. Dellarocas and R. Conte, editors. *Social Order in Multi-Agent Systems*. Kluwer Academic Publishers, 2001.
9. F. Dignum. Autonomous agents with norms. *Artificial Intelligence and Law*, 7:69–79, 1999.
10. F. Dignum, D. Morley, E. Sonenberg, and L. Cavendon. Towards socially sophisticated BDI agents. In *Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS'2000)*, pages 111–118. IEEE Computer Society, 2000.
11. M. d'Inverno and M. Luck. *Understanding Agent Systems*. Springer-Verlag, 2001.
12. A. Jones and M. Sergot. A formal characterisation of institutionalised power. *Logic Journal of the IGPL*, 4(3):429–445, 1996.
13. V. Lesser and L. Gasser, editors. *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95)*. AAAI Press/MIT Press, 1995.
14. F. López y López, M. Luck, and M. d'Inverno. Constraining autonomy through norms. In *Proceedings of The First International Joint Conference on Autonomous Agents and Multi Agent Systems AAMAS'02*, pages 674–681, 2002.
15. M. Luck and M. d'Inverno. A formal framework for agency and autonomy. In Lesser and Gasser [13], pages 254–260.
16. M. Luck and M. d'Inverno. A conceptual framework for agent definition and development. *The Computer Journal*, 44(1):1–20, 2001.
17. Y. Moses and M. Tennenholtz. Artificial social systems. Technical report cs90-12, Weizmann Institute, Israel, 1990.
18. A. Ross. *Directives and Norms*. Routledge and Kegan Paul Ltd., 1968.
19. M. Sergot. Normative positions. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems*, pages 289–308. IOS Press, 1999.
20. J. M. Spivey. *The Z Notation: A Reference Manual*. Prentice-Hall, 1992.
21. R. Tuomela. *The Importance of Us: A Philosophical Study of Basic Social Norms*. Stanford University Press, 1995.
22. R. Tuomela and M. Bonnevier-Toumela. Social norms, task, and roles. Technical report, University of Helsinki, Helsinki, 1992.
23. R. Tuomela and M. Bonnevier-Toumela. Norms and agreements. *European Journal of Law, Philosophy and Computer Sience*, 5:41–46, 1995.
24. A. Walker and M. Wooldridge. Understanding the emergence of conventions in multi-agent systems. In Lesser and Gasser [13], pages 384–389.
25. R. Wieringa, F. Dignum, J. Meyer, and R. Kuiper. A modal approach to intentions, commitments and obligations: Intention plus commitment yields obligation. In M. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems*, pages 80–97. Springer-Verlag, 1996.