Modelling Norms for Autonomous Agents

Fabiola López y López
Facultad de Ciencias de la Computación
Benemérita Universidad Autónoma de Puebla, México
fabiola@cs.buap.mx

Michael Luck
Department of Electronics and Computer Science
University of Southampton, UK
mml@ecs.soton.ac.uk

Abstract

Societies are regulated by norms and, consequently, autonomous agents that want to be part of them must be able to reason about norms. However, no reasoning can be done if agents lack a model of norms that allows them to know how by complying with norms some of their goals might be affected. In this paper, a general model of norms is proposed. By contrast with current models of norms, our model emphasises those aspects that autonomous agents might consider before taking decisions regarding norms. The model is applied to represent the most common kinds of norms that exist in a society, and its effectiveness to design systems of norms is shown.

1 Introduction

Nowadays, agent paradigm is one of the most successful areas of Computer Science [15]. Agents are autonomous problem-solving computational entities able to act in flexible and dynamic environments. Frequently, agents are required to work with other agents which do not necessarily share the same interests, and to avoid conflicts that appear among these autonomous and self-interested agents, *norms* are introduced in a system. Norms prescribe what is permitted and what is forbidden in a society. They specify the responsibilities and benefits for the members and, consequently, agents can make their plans for actions based on the expected behaviour of others. Norms also make formal the agreements between agents that promise to do something and agents that expect this to be done. In general, all kinds of activities that require the coordinated participation of more than one agent are possible thanks to the introduction of norms [10] and, therefore, their use is a necessity for multi-agent systems (MAS) to work effectively.

To incorporate norms in multi-agent systems, efforts have been done to describe and define the different types of norms that agents have to deal with [6, 19], and work that describes reasoning about obligations by using deontic logic [2, 11]. However, the work has not leaded towards a model that facilitates the computational representation of any kind of norm. Norms appear to be different each other, which also suggests that if we want to model agents able to deal with norms, different processes of reasoning must be proposed. This complicates rather than facilitates the modelling of the normative behaviour of agents.

There is also work that introduces norms in multi-agent systems to represent societies, institutions and organisations [9, 16]. There, norms represent the means to achieve coordination among agents which are assumed to be able to comply with norms, to adopt new norms, and to obey the authorities of the system as an end. Thus, although agents in such systems are said to be autonomous, their models of norms do not offer the means for them to understand why norms should be complied with. Consequently, if agents that autonomously decide whether to fulfill a norm are required, a model of norms that allows such decision is needed. That is, a model that allows the representation of different kinds of norms, and that includes elements that help agents in deciding what to do, is needed. In this paper, a general model of norms with these characteristics is proposed. The model is applied to represent the most common kinds of norms that exist in a society, and the way to design, by using the model, a complete system of norms is provided.

The organisation of this paper is as follows. In Section 2, a general model of norms is proposed. Section 3 presents the different categories of norms, whereas in Section 4, norms that make a system of norms are discussed. In Section 5 a multi-agent system regulated by norms is formally defined before providing our conclusions.

2 Norms

Autonomous Agents

In what follows, we use the Z specification language to construct a formal model of norms because it is a language that allows an easy transition from specification to implementation. 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. In addition, to avoid starting from scratch, we adopt the SMART agent framework [8] which provides the basis to understand agents and multi-agent systems. There, an *attribute* represents a perceivable feature of the agent's environment, *goals* are defined as a non-empty set of attributes that describe states of affairs in the world, *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, further details are not needed, so we simply consider them as given sets.

$$[Attribute, Goal, Motivation, Action, EnvState]$$

In SMART, an autonomous agent is described by a set of capabilities that it is able to perform, a set of beliefs that represent the state of its world, a non-empty set of goals that it wants to bring about, and a non-empty set of motivations representing its preferences. By omitting irrelevant details for this paper, we formalise an autonomous agent as follows.

AutonomousAgent

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

$$goals \neq \varnothing;\ motivations \neq \varnothing$$

Norm Model

Norms are mechanisms to drive the behaviour of agents especially in those cases when their behaviour might affect other agents. They can be characterised by their *prescriptiveness*, *sociality*, and *social pressure*. That is, a norm tells an agent how to behave (*prescriptiveness*); in situations where more than one agent is involved (*sociality*); and since it is always expected that norms conflict with the personal interests of agents, socially acceptable mechanisms to force agents to comply with norms are needed (*social pressure*). By analysing these properties, the essential components that enable agents to reason about why a norm should be complied with, can be identified.

Norms specify patterns of behaviour for a set of agents. These patterns are sometimes represented as actions to be performed [1, 21], or restrictions to be imposed over an agent's actions [16, 17]. At other times, patterns of behaviour are specified through goals that must be either satisfied or avoided by agents [5, 19]. Now, since actions are performed in order to change the state of an environment, goals are states that agents want to bring about, and restrictions can be seen as goals to be avoided, we argue that by considering goals the other two patterns of behaviour can be easily represented [13]. In brief, norms specify something that ought to be done and, consequently, a set of *normative goals* must be included in a norm. Sometimes, these normative goals must be directly intended, while at other times their role is to inhibit specific states (as in the case of prohibitions). Norms are always directed at a set of *addressee agents* which are directly responsible for the satisfaction of the normative goals. The set of addressee agents may contain all the agents in the system, as with a mutually understood social law, or it might just contain a single agent. Now, because sometimes, to take decisions regarding norms, agents not only consider what must be done but also for whom it must be done, agents that *benefit* from the satisfaction of normative goals may also be included in a norm.

In general, norms are not applied all the time, but only in particular circumstances or within a specific *context*. Thus, norms must always specify the situations in which addressee agents must fulfill them. For example, if an agent enters a library, the norm of being quiet must be triggered. *Exception* states may also be included, these exception states represent situations in which addressees cannot be punished when they *have not* complied with norms. Exceptions represent *immunity* states for all addressee agents in such a particular situation [18]. Moreover, to ensure that personal interests do not impede the fulfillment of norms, mechanisms either to promote compliance with norms, or to inhibit deviation from them, are needed. As a result, norms may include *rewards* to be given when normative goals become satisfied, or *punishments* to be applied when they are not. Both rewards and punishments are the means for addressee agents to know what might happen whatever the decision regarding norms they take. They are not the responsibility of addressees agents but of other agents already entitled to do it, and since they represent states to be achieved, it is natural to consider them as goals.

The formal specification of a norm is given in the schema *Norm*. All the components of norms described above are included, together with some constraints on them. First, it does not make any sense to have norms

specifying nothing, norms directed at nobody, or norms that either never or always become applied. Thus, the first three predicates in the schema state that all the set of normative goals, the set of addressee agents, and the context must never be empty. The fourth predicate states that the set of attributes describing both the context and exceptions must be disjoint to avoid inconsistencies in identifying whether a norm must be applied or not. The final constraint specify that punishments and rewards are also consistent and, therefore, they must be disjoint.

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

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, actions that are either *permitted* or *forbidden* by a norm are considered as follows. If there is a situation state in which a norm must be fulfilled, and the results of an action benefit the achievement of the associated normative goals, then such an action is *permitted* by the respective norm. For example, the action of leaving a building through an emergency exit is an action that is permitted by the norm of being outside every time a fire alarm becomes activated. Formally, we say that an action is *permitted* by a norm in a particular state of the environment, if and only if the context in which such a norm must be applied is a subset of this state, and the results of the action benefit one of the normative goals of the norm. There, *benefits* is a predicate that is true when an action leads to the satisfaction of a goal.

$$
\begin{array}{|l}
\textit{permitted\_} : \mathbb{P}(\textit{Action} \times \textit{Norm} \times \textit{EnvState}) \\
\hline
\forall\, a : \textit{Action};\ n : \textit{Norm};\ \textit{env} : \textit{EnvState} \bullet \\
\quad \textit{permitted}\ (a, n, \textit{env}) \Leftrightarrow n.\textit{context} \subseteq \textit{env} \wedge \\
\quad (\exists\, g : n.\textit{normativegoals} \bullet \textit{benefits}\ (a\ (\textit{env}), g)) \\
\end{array}
$$

By analogy, *forbidden* actions are defined as those actions leading to a situation which contradicts or hinders the normative goal. For example, the action *illegal parking* is an action forbidden by a norm whose normative goal is to avoid parking in front of a hospital entrance. This is formally expressed below, where *hinders* is a predicate that is true when the results of an action hinder a goal.

$$
\begin{array}{|l}
\textit{forbidden\_} : \mathbb{P}(\textit{Action} \times \textit{Norm} \times \textit{EnvState}) \\
\hline
\forall\, a : \textit{Action};\ n : \textit{Norm};\ \textit{env} : \textit{EnvState} \bullet \\
\quad \textit{forbidden}\ (a, n, \textit{env}) \Leftrightarrow n.\textit{context} \subseteq \textit{env} \wedge \\
\quad (\exists\, g : n.\textit{normativegoals} \bullet \textit{hinders}\ (a\ (\textit{env}), g)) \\
\end{array}
$$

In other words, if an action is applied in the context of a norm, and the results of this action benefit the normative goals, the action is permitted. However, when the action hinders the normative goals instead of providing benefits, that action is forbidden.

3 Categories of Norms

The term *norm* has been used as a synonym for obligations [7], prohibitions [6], social laws [16], social commitments [4, 10] and other kinds of rules imposed by societies (or by an authority). The position of our work is quite different. It considers that all these terms can be grouped in a general definition of a norm, because they have the same properties (i.e. prescriptiveness, sociality and social pressure), and they can be represented by using the same model. All of them represent responsibilities for addressee agents, and create expectations for other agents. They also are the means to support beneficiaries when they have to claim some compensation in the situations where norms are not fulfilled as expected. Moreover, whatever the kind of norm being considered, its fulfillment may be rewarded, and its unfulfillment may be penalised.

What makes one norm different from another is the way in which they are created, their persistence, and the elements that are obligatorily included in the norm. Thus, norms might be created by the agent designer as built-in norms, they can be the result of agreements between agents, or can be elaborated by a complex legal system. Regarding their persistence, norms might be taken into account during different periods of time, such as, until an agent dies, as long as an agent stays in a society, or just for a short period of time until its normative goals become satisfied. Finally, some components of a norm might not exist, there are norms that do not include either punishments or rewards, and even though they are complied with. Nevertheless these differences, all types of norms can be reasoned about in similar ways. Some of these characteristics can be used to provide a *classification* of norms into four main categories: *obligations*, *prohibitions*, *social commitments* and *social codes*.

Obligations and Prohibitions

Obligations and *prohibitions* are norms whose purpose is to ensure the coordination of individuals in a society, and which agents adopt once they become members of the society. Agents adopt these norms because they represent the means to satisfy other important goals. Generally, addressee agents do not participate in their creation, but there are agents entitled to do so. Obligations and prohibitions are considered by agents to be complied with, as long as they stay in a society. The main characteristic of these kinds of norms is that punishments are applied to those agents that offend them. Norms adopted by a secretary in an office, by workers in a factory, or by students in a university are some examples. Formally, an obligation is a norm which unfulfillment is always penalised. To represent it, the schema of a norm is used by imposing a constraint on punishments as follows.

$$Obligation \mathrel{\widehat=} [Norm \mid punishments \neq \varnothing]$$

Whereas obligations represent goals that addressees must bring about, prohibitions represent goals that should be avoided. Since goals are represented as desired states, and states are represented as predicates or their negation, normative goals of prohibitions can be easily represented as negated goals. Consequently, no further distinction between obligations and prohibitions is given. Formally, they have the same representation.

$$Prohibition == Obligation$$

Social Commitments

The second category of norms corresponds to *social commitments*. These are norms derived from agreements or negotiations between two or more agents [10]. They are part of a deal between two sets of agents and, consequently, addressees participate actively in their creation. Normative goals, rewards and punishments of this kind of norms are agreed rather than imposed. Once the normative goals of a social commitment are satisfied, a reward can be claimed. For this reason, social commitments sometimes come in pairs, one specifying what must be done in the first instance, and the other to specify what must be done when the first social commitment becomes fulfilled. Beneficiaries of a social commitment are, in general, responsible for monitoring its fulfillment. Contrary to obligations, social commitments are temporary, because they may disappear once the normative goals become satisfied. Social commitments are formally specified, in the schema below, as norms which fulfillment is always rewarded.

$$SocialCommitment \mathrel{\widehat=} [Norm \mid rewards \neq \varnothing]$$

Social Codes

Our third category of norms is *social codes*. These are norms which are accepted as general principles by the members of a society or a particular agent group. Rather than being forced through punishments or rewards, social codes are complied with as ends in themselves. They are motivated to be fulfilled because of the empathy or sympathy that addressee agents have towards other agents (specially towards agents that benefit from the norm), or because addressee agents want to express their social conformity. Examples of these kinds of norms can be, norms that prescribe that elderly people must have priority over the seats on buses, norms that state that garbage must not be threw on the street, or norms that state that any personal information provided to an institution is confidential. Formally, social codes are norms which have neither punishments nor rewards (at least explicitly). They can be represented as follows.

$$SocialCode \mathrel{\widehat=} [Norm \mid rewards = \varnothing \land punishments = \varnothing]$$

In the remainder of this paper, and in accordance with its definition, the term *norm* is used as an umbrella term to cover every type of norm, namely, obligations, prohibitions, social commitments and social codes. Now, a *normative agent* can be defined as an agent whose behaviour is shaped by obligations that it has to comply with, prohibitions that limit the kind of goals that it can pursue, social commitments that are created during its social interactions, and social codes whose fulfillment represents social satisfaction for the agent. Its formalisation is given as follows.

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

4 System of Norms

Norm Instances

To understand different events that occur in a system due to norms, it is necessary to talk about norms that are either fulfilled or unfulfilled. However, since the majority of the time norms are addressed at a set of agents rather than individuals, the meaning of fulfilling a norm might be subjective. The majority of the times it depends on the interpretation of either designers or analysers of a system. In small groups of

agents, it might be easy to say that a norm has been fulfilled when every addressee agent has fulfilled the norm; by contrast, in larger societies, a percentage of agents complying with a norm will be enough to declare it as fulfilled. As can be seen, instead of defining fulfilled norms in general, it is more practical to define norms being fulfilled by a particular addressee agent. To do so, the concept of norm instance is introduced.

Once a norm is adopted by an agent, a *norm instance* is created. It represents the internalisation of a norm that the agent makes. A norm instance is a copy of the original norm that now is used as a *mental attitude* from which new goals for the agent might be inferred. Norms and norm instances are the same concept used for different purposes. Norms exist in a society, and agents work with *instances* of these norms, consequently, there is an instance for each addressee of a norm. Formally, we do not make any distinction among a norm and its instances. An instance of a norm is formalised as follows

$$NormInstance == Norm$$

We say that a norm has been *fulfilled* by an addressee agent, if all the normative goals of the corresponding instance, have been already satisfied in a specific state. As can be observed, saying that an instance of a norm has been fulfilled is equivalent to say that its normative goals have been satisfied. Formally, we say that an instance of a norm is fulfilled when all its normative goals are satisfied. Its formal representation is given below. There, the *satisfied* predicate is true when the states represented by the goal are a logical consequence of the current environmental state.

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

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

Interlocking Norms

The norms of a system are not isolated each other; sometimes, the compliance with some of them is a condition to trigger (or activate) other norms. That is, there are norms that prescribe how some agents must behave in situations in which other agents either comply with a norm or do not comply with it [18]. For example, when employees comply with their obligations in an office, paying their salary becomes an obligation of the employer; or when a plane cannot take off, providing accommodation to passengers becomes a responsibility of the airline. Norms related in this way can make a complete chain of norms (a system of norms) because the new activated norms can, in turn, activate new ones. Now, since triggering a norm depends on the past compliance with another, we call these kinds of norms as *interlocking norms*. The norm that gives rise other norm is called the *primary* norm, whereas the norm activated as a result of either the fulfillment or unfulfillment of the first is called the *secondary* norm.

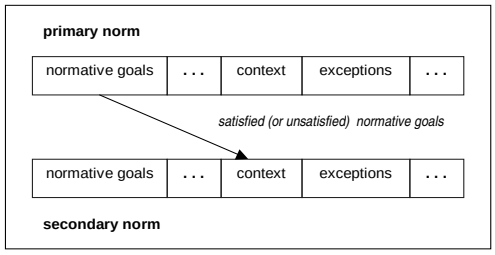

Figure 1. Interlocking norm structure

To describe the structure of these norms in terms of the norm model mentioned earlier, we remind that the *context* is a state that must hold for a norm to be complied with. Now, since the fulfillment of a norm is assessed through its normative goals, the context of the secondary norm must include the satisfaction (or non-satisfaction) of all the primary norm's normative goals. Figure 1 illustrates the structure of both the primary and the secondary norms and how they are interlocked through the primary norm's normative goals and the secondary norm's context.

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

$$\forall\, n_1, n_2 : Norm \bullet$$
$$lockedbynoncompliance\ (n_1, n_2) \Leftrightarrow$$
$$\exists\, ni : NormInstance \mid isnorminstance\ (ni, n_1) \bullet$$
$$\neg\ fulfilled\ (ni, n_2.context)$$

The above definition formally states that a norm is interlocked with another norm *by non-compliance* if in the context of the secondary norm, an instance of the primary norm can be considered as unfulfilled. This means that when any addressee of a norm does not fulfill the norm, the corresponding interlocking norm will be triggered. In the formal definition, n_1 represents the primary norm, whereas, n_2 is the secondary norm. The predicate *isnorminstance* is true when a norm instance is an instance of a norm.

Similarly, a norm is interlocked with another norm *by compliance*, if in the context of the secondary norm, an instance of the primary norm can be considered as fulfilled. Thus, any addressee of the norm that fulfills it will trigger the interlocking norm. The specification of this is given as follows.

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

$$\forall\, n_1, n_2 : Norm \bullet$$
$$lockedbycompliance\ (n_1, n_2) \Leftrightarrow$$
$$\exists\, ni : NormInstance \mid isnorminstance\ (ni, n_1) \bullet$$
$$fulfilled\ (ni, n_2.context)$$

Having the means to relate norms in this way allows us to model how the normative behaviour of agents that are addressees of a secondary norm is influenced by the normative behaviour of addressees of a primary norm.

Enforcement and Reward Norms

Particularly interesting are the norms triggered in order to punish offenders of other norms. We call them *enforcement norms* and their addressees are the *defenders* of a norm. These norms represent exerted social pressure because they specify not only who must apply the punishments, but also under which circumstances these punishments must be applied. That is, once the violation of a norm becomes identified by defenders, their duty is to start a process in which offender agents can be punished. 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 refuses to pay them.

As can be seen, norms that enforce other norms are an especial case of interlocking norms because besides being interlocked by non-compliance, the normative goals of the secondary norm must include every punishment of the primary norm (see Figure 2). By modelling enforcement norms in this way, we make an offender's punishments to be consistent with a defender's responsibilities. Addressees of an *enforced* norm (i.e. the primary norm) can know what might happen if the norm is not complied with, and addressees of an *enforcement* norm (i.e. the secondary norm) can know what must be done in order to punish the offenders of another norm. Enforcement norms allow the authority of defenders to be very well delimited.

Formally, the relationship between a norm directed to control the behaviour of some agents and the norm directed at punishing the offenders of such a norm can be defined as follows. A norm *enforces* another norm if the first norm is activated when the second becomes unfulfilled, and all punishments associated with the unfulfilled norm are part of the normative goals of the first. Every norm satisfying this property is known as an *enforcement* norm.

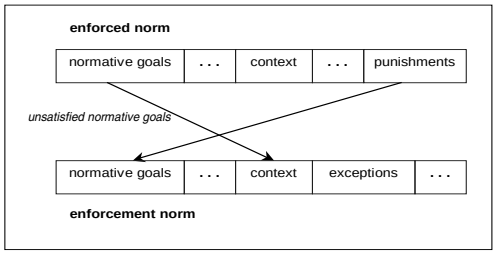

Figure 2. Enforcement norm structure

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

$$\forall\, n_1, n_2 : Norm \bullet enforces\ (n_1, n_2) \Leftrightarrow$$
$$lockedbynoncompliance\ (n_2, n_1) \wedge$$
$$n_2.punishments \subseteq n_1.normativegoals$$

So far we have described some interlocking 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 interlocking norms corresponding to the process of rewarding members doing their duties. These norms must be interlocked by compliance and all the rewards included in the primary norm (rewarded norm) must be included in the normative goals of the secondary norm (reward norm). Formally, we say that a norm *encourages* compliance with another norm 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 first norm. Every norm satisfying this property is known as a *reward* norm.

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

$$\forall\, n_1, n_2 : Norm \bullet rewardnorm\ (n_1, n_2) \Leftrightarrow$$
$$lockedbycompliance\ (n_2, n_1) \wedge$$
$$n_2.rewards \subseteq n_1.normativegoals$$

It is important to mention that this way of representing enforcement and reward norms can create an infinite chain of norms because we would also have to define norms to apply when authorities or defenders do not comply with their obligations either to punish those agents breaking rules or to reward those agents that fulfill their responsibilities [18]. The decision of when to stop this interlocking of norms is left to the creator of norms. If a system requires it, the 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.

Legislation Norms

In all societies, there must exist the possibility of creating new norms (to solve unexpected and recurrent conflicts among agents), modifying existing ones (to increase their effectiveness), or even abolishing those that become obsolete. Although it is possible that many of the members of a society have capabilities to do this, these capabilities must be restricted to be carried out by a particular set of agents in order to avoid everyone imposing norms, otherwise, conflicts of interest might emerge. That is, norms stating when actions to legislate are permitted must exist in a normative multi-agent system [12]. Formally, we say that a norm is a *legislation* norm if actions to issue and to abolish norms are permitted by this norm in the current environment. These constraints are specified as follows.

$$
\begin{array}{|l}
legislate\_ : \mathbb{P}(Norm \times EnvState) \\
\hline
\forall\, n : Norm;\ env : EnvState \bullet \\
\quad legislate\ (n, env) \Leftrightarrow \\
\quad (\exists\, issuingnorms, abolishnorms : Action \bullet \\
\quad\quad permitted\ (issuingnorms, n, env) \lor \\
\quad\quad permitted\ (abolishnorms, n, env))
\end{array}
$$

Enforcement, reward and legislation norms acquire particular relevance in systems regulated by norms because the abilities to punish, reward, and legislate must be restricted for use only by competent authorities (addressees of the respective norms). Otherwise, offenders might be punished twice or more times if many agents take this as their responsibility. It could also be the case that selfish agents demand unjust punishments or that selfish offenders reject being punished. That is, conflicts of interest might emerge in a society if such responsibilities are given either to no one or to anyone.

5 Normative Multi-Agent Systems

Norms cannot be studied independently of the systems for which they are created. Although social systems that are regulated by norms are different from one another, some general characteristics can be identified. They consist of a set of agents that are controlled by the same set of norms ranging from obligations and social commitments to social codes. However, whereas there are static systems in which all norms are defined in advance and agents in the system always comply with them [3, 16], a more realistic vision of these kinds of systems suggests that when *autonomous* agents are considered, neither can all norms be known in advance (since new recurrent conflicts among agents may emerge and, therefore, new norms may be needed), nor can compliance with norms be guaranteed (since agents can decide not to comply). We can say then, that systems regulated by norms must include mechanisms to deal with both the modification of norms and the unpredictable normative behaviour of autonomous agents. By using the concepts already defined, a general system regulated by norms can be formalised.

$$
\begin{array}{|l}
NormativeMAS \\
\hline
members : \mathbb{P}\, NormativeAgent \\
generalnorms, enforcenorms : \mathbb{P}\, Norm \\
rewardnorms, legislationnorms : \mathbb{P}\, Norm \\
environment : EnvState \\
\hline
\forall\, ag : members \bullet \\
\quad ag.norms \cap generalnorms \neq \varnothing \\
\forall\, sn : generalnorms \bullet sn.addressees \subseteq members \\
\forall\, en : enforcenorms \bullet \\
\quad (\exists\, n : generalnorms \bullet enforces\ (en, n)) \\
\forall\, rn : rewardnorms \bullet \\
\quad (\exists\, n : generalnorms \bullet rewardnorm\ (rn, n)) \\
\forall\, ln : legislationnorms \bullet \\
\quad legislate\ (ln, environment)
\end{array}
$$

The schema above represents a normative multi-agent system. It comprises a set of normative agent members (i.e. agents able to reason about norms) and a set of general norms that govern the behaviour of these agents (represented here by the variable *generalnorms*). There are also norms dedicated to enforce other norms (*enforcenorms*), norms directed to encourage compliance with norms through rewards (*rewardnorms*),and norms issued to allow the creation and abolition of norms (*legislationnorms*). The current state of the environment is represented by the variable *environment*. Constraints over these components are imposed as follows. Although it is possible that agents do not know all the norms in the system due to their own limitations, it is always expected that at least they adopt some of the norms, the first predicate in the schema represents this fact. The second predicate makes explicit that addressee agents of norms must be members of the system. Thus, 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. The last three predicates respectively describe the structure of enforcement and reward norms, 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 the society is legally entitled to punish an agent that does not fulfill such a norm.

6 Conclusions

There are many advantages of the model of norms presented in this paper over other models of norms. The model subsumes other norm models [4, 6, 10, 16, 22] where only some components of our model are considered. Rather than being directed at agents that always comply with norms, our model is directed at autonomous agents that decide by their own whether to do it because the model includes components that facilitate such a decision. It provides the means for agents to know how to behave in particular situations, the consequences of complying (or not) with the norm, and who benefits from this actions. Thus, the model sets up the basis to model the normative behaviour of autonomous agents [14]. The model also allows the representation of a complete system of norms where compliance with some norms activates other norms. By using interlocking norms, reward and enforcement norms are defined. A general model of normative multi-agent systems is proposed. In this kind of systems, norms to control the normative behaviour of agents are very well defined.

References

[1] R. Axelrod. An evolutionary approch to norms. *The American Political Science Review*, 80(4):1095–1111, 1986.

[2] M. Barbuceanu, T. Gray, and S. Mankovski. The role of obligations in multiagent coordination. *Applied Artificial Intelligence*, 13(1/2):11–38, 1999.

[3] M. Boman. Norms in artificial decision making. *Artificial Intelligence and Law*, 7(1):17–35, 1999.

[4] C. Castelfranchi. Commitments: From individual intentions to groups and organizations. In V. Lesser and L. Gasser, editors, *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95)*, pages 186–196. AAAI Press/MIT Press, 1995.

[5] 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.

[6] F. Dignum. Autonomous agents with norms. *Artificial Intelligence and Law*, 7(1):69–79, 1999.

[7] F. Dignum, D. Morley, E. Sonenberg, and L. Cavendon. Towards socially sophisticated BDI agents. In E. H. Durfee, editor, *Proceedings on the Fourth International Conference on Multi-Agent Systems (ICMAS-00)*, pages 111–118. IEEE Computer Society, 2000.

[8] M. d'Inverno and M. Luck. *Understanding Agent Systems*. Springer-Verlag, 2001.

[9] M. Esteva, J. Padget, and C. Sierra. Formalizing a language for institutions and norms. In J. Meyer and M. Tambe, editors, *Intelligent Agents VIII (ATAL'01)*, LNAI 2333, pages 348–366. Springer-Verlag, 2001.

[10] N. Jennings. Commitments and conventions: The foundation of coordination in multi-agent systems. *The Knowledge Engineering Review*, 8(3):223–250, 1993.

[11] A. Jones and M. Sergot. Deontic logic in the representation of law: Towards a methodology. *Artificial Intelligence and Law*, 1(1):45–64, 1992.

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

[13] F. López y López and M. Luck. Towards a model of the dynamics of normative multi-agent systems. In G. Lindemann, D. Moldt, M. Paolucci, and B. Yu, editors, *Proceedings of the International Workshop on Regulated Agent-Based Social Systems: Theories and Applications (RASTA'02) at AAMAS'02*, pages 175–193. University of Hamburg, 2002.

[14] F. López y López, M. Luck, and M. d'Inverno. Constraining autonomy through norms. In C. Castelfranchi and W. Johnson, editors, *Proceedings of The First International Joint Conference on Autonomous Agents and Multi Agent Systems AAMAS'02*, pages 674–681. ACM Press, 2002.

[15] M. Luck, P. McBurney, and C. Preist. Agent technology: Enabling next generation computing (a roadmap for agent based computing). Technical report, AgentLink, 2003.

[16] Y. Moses and M. Tennenholtz. Artificial social systems. Technical report CS90-12, Weizmann Institute, Israel, 1990.

[17] T. Norman, C. Sierra, and N. Jennings. Rights and commitments in multi-agent agreements. In Y. Demazeau, editor, *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98)*, pages 222–229. IEEE Computer Society Press, 1998.

[18] A. Ross. *Directives and Norms*. Routledge and Kegan Paul Ltd., 1968.

[19] M. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7(1):97–113, 1999.

[20] J. M. Spivey. *The Z Notation: A Reference Manual*. Prentice-Hall, 1992.

[21] R. Tuomela and M. Bonnevier-Toumela. Social norms, task, and roles. Technical report HL-97948, University of Helsinki, Helsinki, 1992.

[22] R. Tuomela and M. Bonnevier-Toumela. Norms and agreements. *European Journal of Law, Philosophy and Computer Sience*, 5:41–46, 1995.