Virtual Organizations as Normative Multiagent Systems

Guido Boella[1] Joris Hulstijn[2] Leendert van der Torre[3]

Abstract

In this paper we propose a conceptual model of virtual organizations as normative multiagent systems. The dynamic aspects of virtual organizations are modeled using aspects of speech act theory and Searle's theory of the construction of social reality. We illustrate the use of our model by discussing an example of distributed access control policies. We show how the model captures the distinction between local and global authorities, and between local and global norm enforcement policies.

1. Introduction

Recent developments on environments for computer supported collaborative work, distributed knowledge management and 'grid architectures' for sharing resources and computational services have lead to an increased interest in what has been termed *virtual organizations*: a set of individuals and institutions that need to coordinate resources and services across institutional boundaries [19]. Users of such a coordination infrastructure can form a community, with a shared set of objectives and norms about what constitutes accepted behavior. Nevertheless, control over services and resources largely remains local. The problem is to align the objectives and norms of the community of users, with the formal rules of such an infrastructure. Traditional client-server architectures can accommodate global objectives, but can not delegate basic local control. Fully distributed peer-to-peer architectures on the other hand, do not offer enough mechanisms to realize global objectives. Apparently we lack a conceptual model of virtual organizations with norms at different levels of control.

In this paper we present and discuss a conceptual model of virtual organizations based on what we call a *normative multiagent system* (NMAS) [3, 4, 5]. Unlike a central authority or a bureaucracy [28], our model allows a distributed control structure with normative agents operating at different levels of control. And unlike a peer-to-peer system, our model offers a more elaborate normative structure, which enables it to address global concerns. Given the proposal to model virtual organizations by normative multiagent systems, a number of questions remain.

1. How can the behavior of an individual agent in a virtual organization be described?
2. How can agents change a virtual organization?
3. How can agents in a virtual organization establish normative relations or contracts with each other?
4. How can we deal with norms that operate at different levels of control?

Our answer to these questions depends on the interaction among agents and normative systems. In general, the decisions of agents depend on recursive models of the expected behavior of other agents and normative systems. We consider interactions between an agent and a normative system, where an agent models the normative system as an autonomous agent, and the normative system defines roles that agents can play in the system. Moreover, we consider interaction among individual agents in the context of a normative system, like in contracting. Finally we consider interactions among normative systems, like the interaction between a global and a local authority with respect to global enforcement policies.

There are two ways to model virtual organizations. One may study existing virtual organizations, and try and find a descriptive model that best matches the observations. Alternatively, one may design and implement an architecture that enables a virtual organization. The architecture is based on a model, which thereby prescribes important aspects of future behavior. Our model is intended to be used in this prescriptive way, and has no empirical pretensions. More details of the formal logic that underlies the model can be found elsewhere [4, 5].

The layout of the paper is as follows. First we discuss the general structure of our model, and the methodology to model virtual organizations. We explain the basic idea behind the agent metaphor, and in what ways it can be applied. The questions 1 – 4 are addressed in the subsequent sections of the paper.

1 Dipartimento di Informatica, Università di Torino, guido@di.unito.it
2 Faculty of Economics and Business Administration, Vrije Universiteit, Amsterdam, jhulstijn@feweb.vu.nl
3 CWI, Amsterdam, torre@cwi.nl

2. Normative Multiagent Systems

The NMAS model introduced in this paper builds on a model developed for the design of multiagent systems (MAS). Multiagent systems are computer systems that are composed of a number of autonomous agents, i.e., pieces of software, that must interact in order to achieve the global system's objectives. Applications can for example be found in transport logistics [11], manufacturing scheduling [35], and social simulation [33]. Increasingly, developers of multiagent systems have turned to concepts from the social sciences, such as organization structures and norms as a guideline for design [15]. We believe that the principles of the NMAS model are general enough to be applied to human as well as artificial agents. In this paper we do not propose to deploy 'intelligent agents' to monitor or control aspects of virtual organizations. Whether such applications are successful is an empirical question [24]. In this paper the agent metaphor is only used as a conceptual tool.

We assume a partial unification of agents – human or artificial – and normative systems. A normative system is nothing but a set of explicit rules of behavior, norms, that apply to a set of agents. A normative system may provide additional rules for detecting violations and applying appropriate sanctions. The central idea is that both agents and normative systems can be viewed – and thus modeled – as autonomous decision makers [3]. Like an agent, a normative system serves some objective, for example to maintain property rights. And like agents, a normative system only has a limited capacity to pursue such objectives. A normative system can be considered as an autonomous agent from the viewpoint of an individual agent. For example, when an agent is deliberating whether or not to obey a norm, it will reason about the expected behavior of the normative system, given what it knows of the objectives and limitations of the normative system with regard to violation detection and sanctioning. A normative system can also be considered as an autonomous agent from the point of view of another normative system. However, normative systems are obviously different from ordinary agents in many other respects, and the two kinds of agents should be distinguished. In the end, the net behavior of a normative system is determined by individual agents playing pre-defined roles in the system, where the roles are governed by further norms. Only when the model abstracts over the individual properties of these agents, we may say that the normative system as such is making a decision. To stress the commonality, we sometimes refer to a normative system as a normative agent.

Starting point of the model is that there are different kinds of interactions in a normative multiagent system: interactions between an agent and a normative system, interactions among agents subject to a normative system, and interactions among normative systems. These interactions can take the form of communication between agents and normative systems, for example a normative system commanding an agent to do something .It can also mean an agent changing the normative system, or two agents establishing a contract in the context of the normative system. The interaction can be merely hypothetical, when an agent has a profile of another agent, and tries to predict its behavior.

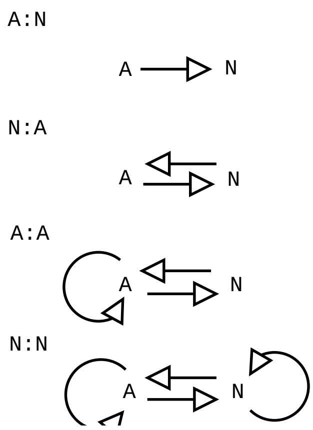

Figure 1: Types of Interaction between agents A and normative systems N

The core of the model is a relational view of two agents, either ordinary agents A or normative agents N. The model is structured according to the different kinds of interaction (Figure 1).

1. **A:N**: *How can the behavior of an individual agent in a virtual organization be described?*
   Interaction. The normative system can command an individual agent using prescriptive norms and imperative speech acts. For general norms, the normative system has to decide whether the agent is subjected to the norm. In a social system, the agent has a choice to accept or reject the norm. The normative system monitors the agent and observes whether behavior would count as a violation. If a violation is detected, the normative system can sanction the agent.
   Recursive modeling. The view on normative systems from the perspective of individual agents depends on the agent model used. We discuss the *cognitive foundations* of our model, and how norms can influence individual decisions [13] (Section 3), based on an existing cognitive agent architecture called BOID [13]. In addition to modeling the customary beliefs, desires and intentions of an agent (abbreviated BDI), also the effect on decision making of normative concepts like obligations (referred to as O) is accounted for.

2. **N:A**: *How can agents change a virtual organization?* **Interaction.** First it must be decided which agents are allowed to do what in the normative system. Some agents have the power to introduce new norms, to decide to count behavior as a violation, or to sanction other agents. These powers are related to roles like legislator, policemen, defender, etc. Due to the separation of duties, some combinations of roles are excluded. Second, agents in certain roles and under certain circumstances can change the rules of a normative system, by uttering a performative speech act that constitutes a change (Section 4).

   Recursive modeling. A normative system not only views agents as subjects which it can oblige, prohibit, permit, sanction and control, but it also views the agents as subjects that can play other roles in the system, for example to modify the normative system. We also look at constitutive norms, that can be contrasted with regulative norms, and at norms regulating interaction itself, such as dialog games or interaction protocols.

3. **A:A**: *How can agents in a virtual organization establish normative relations or contracts with each other?*

   Interaction. When two agents in a virtual organization together would like to cooperate, but individual rationality dictates them to defect (as in the Prisoner's Dilemma), they can agree to draw up a contract. The normative system may contain a contractual framework. The agents can propose a contract, and negotiate its conditions. Typically when two agents are in a contracting relationship there may be disagreement about the conditions of the contract, whether behavior counts as a violation, and how such behavior should be sanctioned. Contracting involves social phenomena in normative multiagent systems like directed obligations and the formation of groups (Section 5).

   Recursive modeling. We consider agents from the viewpoint of other agents, given a normative system in which the interaction is embedded. This involves social phenomena like trust, as well as control mechanisms to enforce contracts.

4. **N:N**: *How can we deal with norms that operate at different levels of control?*

   Interaction. In Section 6 we study the example of distributed access control policies [36]. We propose a distinction between local and global authorities, and between local and global norm enforcement policies. We distinguish three kinds of agents: subjects, whose behavior is governed by norms; defenders, who should monitor violations of norms and apply sanctions; and autonomous normative systems, who issue norms and supervise defender agents.

   Recursive modeling. The example of distributed access control policies [36] crucially involves normative systems that contain multiple other normative systems, which can interact at various levels.

3. Cognitive Foundations

In this section we consider the viewpoint of an arbitrary agent. We assume that the core of the agent's deliberation is based on mental attitudes like beliefs, desires, goals and plans, following the tradition in cognitive science, e.g., [10] and on obligations, studied in deontic logic [32] – the logic of normative systems. The mental attitudes are interpreted as follows [13]. Beliefs are informational attitudes; they represent what the world is expected to be like. Desires represent internal motivations, which may lead to possible goals for the agent to pursue. When a goal is adopted by the agent, a plan is formed to achieve the goal. Finally, obligations model external motivations for an agent, such as norms.

3.1. Formal Models

In our model, these mental attitudes are not represented as sets of sentences as is customary, e.g. [37], but as sets of conditionals or production rules. This expresses the fact that mental attitudes are context dependent by nature [22], and that their application is conditional on certain constraints. So each attitude Bel, $Goal$, Obl, etc., is represented by a set of rules of the form $A \longrightarrow B$, where both A and B are sequences of facts. Moreover, B may contain special decision variables, or actions. The values of decision variables are under the control of the agent. For simplicity, both facts and decision variables are represented by boolean variables, being either true or false. The decision making of an agent is represented by a forward reasoning loop, which runs roughly as follows.[1] The agent receives input from observation, represented as a set of facts S. Alternatively, the agent may start with a set of initial goals, also represented by a set of boolean variables S. Now the agent tries to match each rule $A \longrightarrow B$ against S. If A is contained in S, and the facts of B do not contradict a fact in S, the rule is applicable. However, there may be several rules applicable to S, from the same and from different mental attitudes, each with a different possible outcome. Using a priority ordering, the agent selects one rule – this is called conflict resolution – and applies it: the result B is added to S. This process continues, until a stable set of facts is reached, to which no further additions can be made. Such a stable set, an extension, represents the result of the decision making. For

1 Technical details of the reasoning is expressed using input/output logics [29, 30]. Their application to Normative Multiagent Systems is explained in [4, 5].

agent a deliberates about optimal decision
– considers optimal decision of agent b
agent b deliberates about optimal decision
– considers optimal decision of agent a
agent a deliberates about optimal decision
– considers optimal decision of agent b
...

Figure 2: Recursive modeling using agent profiles

example, a selfish agent will prefer desires to social obligations; a social agent will let obligations take priority. The decision making behavior of an agent therefore crucially depends on the way the conflicts among the mental attitudes are resolved. Different priority orders lead to different extensions, which represent different kinds of behavior.

3.2. Recursive Modeling

Interaction between agents is based on the idea that agents also have profiles of other agents, which specify the mental attitudes of others. Technically, profiles also contain sets of production rules. Individual agents can therefore play hypothetical games with other agents to predict their behavior, in response to their own behavior (Figure 2). This is called recursive modeling. Such recursive modeling can become rather complex. Therefore we assume that our agents are resource bounded, such that for example the number of levels at which an agent can reason about other agents is limited. Typically, an agent can reason about another agent's expectations about a third agent; more complicated nestings are not allowed.

Recursive modeling is especially useful to model trust, deception and threats. Consider an agent a that receives information from agent b. Now should a trust b regarding the truth of that information? Suppose a already has a profile of b, to the effect that b has no desire to be truthful to a. In that case, one likely outcome of the recursive simulation will be that b will not speak the truth. So a should not trust b.

Interaction between an agent and a normative system is modeled just like the interaction between two agents. The agent thus assumes that the normative system has motivational attitudes, and that these desires or goals of the normative system are like commands for the agent. One example can be found in a traditional client-server architecture, where a 'wish' for the server, is in fact a 'command' for the client. Moreover, the agent assumes that the normative system has informational attitudes, which partly determine the state of the world. Following Searle [39], instead of saying that the normative system *believes* that A implies B in context S, we would say that A *counts as* B in context S. The normative system creates and records the facts that make up social reality. For example, within the conventional context of going to the theater, a ticket to a performance counts as evidence that you are entitled to occupy a seat during the performance. More about such constitutive norms in section 4.3 below.

Unlike desires or goals, norms are external to an agent. So how can norms and obligations affect decision making? The idea is that behavior or a situation which does not satisfy the obligation, is seen as a social fact that counts as a violation in the normative system. The idea to use violation conditions to model obligations, is known as Anderson's reduction. So instead of specifying that access to some resource is forbidden, we specify that accessing the resource counts as a violation. Thus a rule of the form $\texttt{access}(a, r) \longrightarrow \texttt{viol}(a)$ will be part of the 'beliefs' of the normative system, where a can be any agent and r any resource. Moreover, with each obligation we can associate a sanction. A sanction is an action or fact that is undesirable for each agent that is subject to the norm. It must also be undesirable for the normative agent to impose a sanction, and sanctions must only be triggered by detected violations. These restrictions prevent arbitrary sanctioning [3]. Note that there is a distinction between behavior that counts as a violation and behavior that is actually being sanctioned, as these are distinct decisions of the normative system.

4. Dynamics

How can we account for the dynamics of a normative system? Again, we apply ideas from Searle [39], including older work on speech acts [1, 38]. Roughly, there are speech acts, so called performatives, that in a given community 'count as' a change to the rules and social facts governing the normative system. For example, by explicitly authorizing an agent to access a resource, that agent acquires the right to access the resource. This means that the access policies of the community have in fact been altered. So just like certain facts may count as evidence of other social facts, as we have seen in the example of a theater ticket, certain actions may count as changes to the social facts. Typical examples are making a promise, or accepting a request. However, performatives only have their intended effect, when so called felicity conditions or applicability conditions are satisfied. Consider for example a command, which only has effect in case the commanding agent has authority over the agent commanded. In the distributed access policy example, the right to entitle others to access a resource, is granted to the owner of a resource. In other words, only if the speaker and the addressees of a speech act enact appropriate roles in the normative system that underlies the interaction, the speech act will be successful.

4.1. Roles

Agents can play different roles in a normative system. A role specifies the rights and duties of each agent that enacts that role (e.g., access rights, safety maintenance duties), as well as social relationships with other roles (e.g., authority, dependence). Thus, normative systems are modeled as a structured entity, where the structure is provided by the roles and the role relations. This is useful to model relations in institutions as studied in organizational theory, as well as roles in legal theory. For example, in the *trias politica* there are different roles associated with the executive, legislative and judicial powers. In a similar way, one can conceive of a normative system in which the right and duty to see to it that norms are obeyed and that violations are detected and sanctioned, have been delegated to so called *defender* agents. These defender agents control *subjects*, but are themselves controlled by the autonomous *normative system* as a whole.

The possibility to delegate powers to agents in different roles, is a distinctive property of our model. Normative systems have thus far been modeled as abstract entities. If normative systems or institutions are represented at all, then often they are just referred to as indices S, for example in the phrase "A counts as B in institution S" [27]. However, it is crucial to find out what aspects of the structure of an institution make sure that such a constitutive norm is in place.

4.2. Institutional Change

Virtual organizations are subject to change. Changes take place at different places in the model, and can have a different scope and impact. Speech acts may be exchanged between participants in an interaction. And in case a speech act counts as a particular move in a dialog game (see below), changes to the institutional facts are established. For example, when accepted, a promise creates a commitment for the promiser to keep the promise. Similarly, when new agents enter a community, a registration action creates their new status as a member. The way such a registration action is carried out differs between applications. In computational settings, often a login is created. A login can later be used as evidence of the identity and role of the agent.

Technically, updates to a normative multiagent system can be accounted for as follows: instead of one model, we use an ordered series of models representing the state of affairs for each possible moment in time. In case it is important to compare different outcomes, a so called branching-time logic like CTL or CTL* is used: each possible future course of events is represented by a 'branch' of the tree. See [14] for an application of this idea to normative models. Alternatively, we can use linear time models, using the different extensions generated by the recursive game simulations as representations of alternative outcomes.

The examples above concern a change that takes place within a normative system. The system itself remains relatively stable. But norms and social structures can evolve too. However, such changes are much more elaborate, and are subject to restrictions. For instance, a norm must be known to all agents that are subject to it. Therefore, before a new norm is established, all agents must be informed. Such restrictions on the formation of norms can be laid down in a second normative system. Thus, law scholars like Hart [23] distinguish between primary laws, whose purpose is to direct the behavior of citizens, and secondary laws, which, among other functions, serve the maintenance of the normative system. These rules are instrumental to the primary system, in that they regulate the regulation. For example, art. 2 of the Italian Civil Code states: "the creation of laws [...] is regulated by constitutional laws" (Cost. 70). This subsystem, according to Hart, does not include only the rules of change which specify how new laws are introduced or old ones removed, but it also includes rules about powers for the private citizen. These rules are at the basis of civil code and allow testaments and contracts; for Hart they allow the exercise of limited legislative powers by the citizens. These rules do not create or remove general laws, but they introduce and remove individual obligations and permissions: e.g., the Italian Civil Code art. 1173 (sources of obligations) specifies that obligations are created by contracts, where a contract is meant as an agreement among two or more parties to regulate a juridical relationship about valuables by art. 1321.

4.3. Constitutive Norms

Regarding institutional change, we must first explain the distinction between *regulative norms* and *constitutive norms*, that was introduced by Searle in his work on speech acts [38], and further developed in his later work on the construction of social reality [39]. Some rules regulate antecedently existing forms of behavior. For example, the rules of polite table behavior regulate eating, but eating exists independently of these rules. Some rules, on the other hand, do not merely regulate an antecedently existing activity; they, as it were, create the possibility of or define that activity. For example, the activity of playing chess is constituted by the rules of the game. The institutions of marriage, money, and promising are like the institutions of baseball and chess in that they are systems of such constitutive rules or conventions [38, p. 131]. Within normative multiagent systems we distinguish between regulative norms that describe obligations, prohibitions and permissions, and constitutive norms that regulate the creation of institutional facts like property, marriage and money, as well as the modification of normative system itself.

Constitutive norms are introduced in our normative multiagent systems for the following three reasons.

First of all, regulative norms are not categorical, but conditional: they specify all their applicability conditions. In case of complex and rapidly evolving systems new situations arise which should be considered in the conditions of the norms. Thus, new regulative norms must be introduced each time the applicability conditions must be extended to include new cases. In order to avoid changing existing norms or adding new ones, it would be more economic that regulative norms could factor out particular cases and refer, instead, to more abstract concepts only. Hence, the normative system should include some mechanism to introduce new institutional categories of abstract entities for classifying possible states of affairs. Norms could refer to this institutional classification of reality rather than to the commonsense classification [12]: changes to the conditions of the norms would be reduced to changes to the institutional classification of reality.

Second, the dynamics of the social order which the normative system aims to achieve is due to the evolution of the normative system over time, which introduces new norms, abrogates outdated ones, and, changes its institutional classification of reality. So the normative system must specify how the normative system itself can be changed by introducing new regulative norms and new institutional categories, and specify by whom the changes can be done.

Third, the dynamics of a normative system also includes the possibility that ordinary agents create new obligations, prohibitions and permissions concerning specific agents. This activity is particularly important in applications for e-commerce where it is necessary to model contracts which introduce new normative relations among agents, like the duty to pay a fee for a service [17, 34].

4.4. Interaction Norms

We are working on extensions of the model to incorporate agent communication languages and dialog games [2]. Interaction itself can be seen as governed by a set of normative rules too. Such rules may be formulated in the shape of a dialog game, e.g., [41, 26, 31]. A dialog game defines the roles of the participants, the repertoire of speech acts, the sequences of speech acts that are considered well-formed, the meaning that is given to each speech act in terms of the resulting effect on the underlying normative system, and the entry and end conditions. Given a message containing a particular speech act, a particular response may be obligatory. For example, a request must be either granted or denied. A command must be carried out, at least, when there is an authority relationship between the commander and person being commanded. However, to determine what counts as an appropriate response, there is room for interpretation by the receiver. For example, a remark like “Open the window!” can be taken as a command, or as a suggestion. What kind of interpretation is selected by the receiver, is indicated by the response. So a response like “good idea!” would signal that the receiver takes it as a suggestion, and not as a command. Such a response also denies the existence of an authority relation. This effect of the receiver influencing the dialog, is called uptake [16]. An preliminary account of uptake was given using the BOID model [25].

5. Contracts and Social Relations

In our approach, the behavior of agents is modeled through a number of recursive games. An agent has a profile of other agents and of the normative system, and will base its decisions on the reactions that come out of a simulation of the game using the profiles. Agents in a normative system are subject to obligations. When an agent considers the normative system as part of its recursive modeling, this may influence its behavior. When agents consider each other in the context of a normative system, then obligations can also be directed towards other agents, and not just towards the normative system [14]. For example, a promise creates an obligation, not towards society as a whole, which provides the normative system against the background of which promises conventionally hold, but towards the specific agent the promise was made to. Such directed obligations almost always coincide with some social relationship between agents, that helps to define a community or an organization. In the case of a promise, that relationship might be one of friendship. In the case of a contract, the relationship might be a long term trading relation, etc. A contract is typically drawn up by agents that are of equal authority, but that depend on each other. Although there is an incentive to honor the contract, there must be a guarantee against opportunistic behavior. Agents gather power over each other due to the directed obligations they can create and delegate. These powers are part of a social reality based on institutional facts. The distinction is between agents that are individually rational, and social agents that are aware of other agents, set against a background in which the institutional relations between agents have been fixed.

A particularly interesting social relationship, is that of being member of the same group. Social agents can operate as a group, team or coalition and have joint goals. In case there is a joint goal, this generates further obligations. In particular, in addition to the fact that members must have the goal of completing their part of the task, they are obliged to assist each other and to communicate failure or completion of their part [40]. If a group has collective obligations, this generates the problem of how to distribute these obligations over the members of the group [20]. The main issue for the agents is how to deal with social laws. A related issue is due to social norms and commitment policies, see e.g. [21]. Moreover, when social norms are institutionalized, then agents can create obligations directed towards each other by creating contracts. In human society, the creation and execution of contracts is regulated by contract law. In virtual organizations, rules with a similar function should exist, the define the meaning of newly created or abolished obligations, about violations and possible sanctions and about exceptions.

6. Local and Global Policies

In this section we consider normative systems from the viewpoint of other normative systems. Though normative reasoning often assumes a single normative system, as in moral theory, in reality there are typically many interacting normative systems and their relation has to be explained by any useful model of normative systems. The best example is legal theory, which is typically based on hierarchies of norms. There are for example the laws of the European Union and of the individual member states, with carefully delimited areas of jurisdiction. The interaction between normative systems is also important in virtual communities, where we can distinguish between local and global authorities. This idea is elaborated in the example below.

6.1. Distributed Access Control

Developments in network technology have raised new problems concerning the control of peer-to-peer systems and grid architectures [19]. In these computing paradigms there is no central administrator; access to resources is controlled at local levels. Participants can take the role of a consumer, or of a provider of resources. As a provider, participants have the power to set their own access policy. If a set of participants wants to form a virtual community to share resources, local access policies should be organized according to some global policy which specifies how resources can be shared. In the traditional client-server approach access policies are controlled by a central authority. According to Pearlman et al. “a key problem associated with the formation and operation of distributed virtual communities is that of how to specify and enforce community policies” [36]. Since there is no plausible way to enforce a global policy by technically constraining a distributed architecture, a normative system is needed to specify global policies about local policies. Moreover, a mere specification of what is permitted and what is obligatory is not sufficient: “exercise of rights is effective only if the resource provider has granted those rights to the community” [36]. The normative system must therefore enforce the norms at each level. Not only users must be given an incentive to respect the norms, but also the local providers must be motivated to issue local policies that respect the global ones.

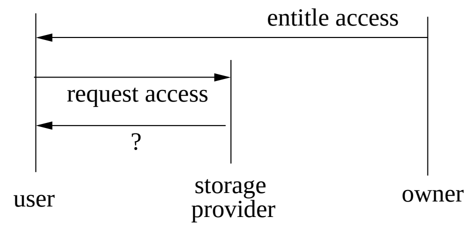

Figure 3: User a_1 requests access to resource d from storage provider a_2, having been entitled access to d by owner a_3

The following example is taken from Firozabadi and Sergot [18]. The example hinges on the difference between the permission and the right to access a resource. Suppose there are agents a_1, a_2 and a_3, and there is no central access control. In order achieve its purpose p, agent a_1 wants to access a resource d which is stored remotely, in control of agent a_2. Now suppose a_2 does not at the moment permit a_1 to access the resource. When a_1 requests permission to access the resource, a_2 will have to make a decision, based on its local access policies combined with the global access policies.

There are several possible scenarios:

1. Suppose a_1 is (globally) permitted to access resource d, but there is no obligation on a_2 to (locally) grant that access. That means a_2 will not violate any policy regardless of whether it grants or denies the access.

   In our model, the decision depends entirely on the individual desires and goals of the local authority. If the local authority has a desire to limit the data traffic, for example, it might choose to deny access.

2. Suppose a_1 is not only permitted to access d, but is also entitled to it. This means that a_2 has an obligation to grant access whenever a_1 requests it. This may happen when a_1 is the owner of d, or when d is owned by another agent a_3 and a_3 has entitled a_1 to access d on a_2 (Figure 3).

   Consider the case in which a_1 claims to be the owner. First, a_2 must establish whether a_1 really does count as the owner of d. It will need some evidence that constitutes the identity of a_1. For example, the global authority may issue ownership certificates. Second, a_2 must establish whether or not denying access will count as a violation. If so, it must establish if the chances of getting detected and the severity of the corresponding sanction, outweigh the benefit of not granting access. To this end, it will employ a version of recursive modeling. If a_2 decides to grant access, it must perform a social act that counts as the permission for a_1 to access d, and that also enables a_1 to physically access the data.

Consider the case in which a_1 claims to have been entitled by owner a_3 to access d. Amongst other things, it must be specified which speech acts constitute an entitlement. In one possible set-up, the owner may order the local authority to grant an agent access for a period of time. In that case the local authority must verify the authenticity of the order, and store the details. The power for owners to entitle agents' access is guaranteed by the global normative system. Everything else is handled locally. In another set-up, the owner may create an institutional fact that counts as evidence of the agent being entitled access, similar to the identity certificates discussed above. In this case, both the evidence and the obligation for local authorities to grant access, are guaranteed by the global normative system.

3. Suppose a_1 has no permission to access d, so a_2 is globally forbidden to grant a_1 access. Such a prohibition is represented by rules specifying what counts a violation, and by a sanctioning goal for the global normative system. Since a_2 physically controls access, it may still grant a_1 access. This may happen for example, when a_2 has desires or goals, that would make it beneficial to grant access, and the chances of being detected and sanctioned are small. Bribes fall into this category, as well as kind librarians that want to help out a student with no library card.

This brief exposition indicates, that in principle the conceptual model of normative multiagent systems is rich enough to capture the crucial aspects of the scenarios. Moreover, all of the concepts (agents, roles, regulative norms, constitutive norms, speech acts) are needed to cover some aspect of the scenario. So on first sight, the model is not too rich either. Having said that, we realize that a proper empirical validation of the use of the model for designing and evaluating distributed access control systems, would require a detailed field study.

Note furthermore, that for many applications, the kind of rational deliberation shown by the agents in assessing the normative system's ability to detect and sanction a violation, is not necessary. In that case, norm abiding behavior is 'hard-wired' into the application [14].

7. Conclusions

We conclude by summarizing the advantages of our conceptual model. As discussed above, our model is based on a partial unification of computer systems and normative systems, both called agents. It is this partial unification that clarifies their relation, and the distinctions. The second, third and fourth advantage have first been observed in [6]. The advantages are as follows.

1. The complex structural relation between agents and normative systems has been decomposed into four separate relations, which each can be studied on their own: normative systems viewed from the perspective of an individual agent, individual agents viewed from the perspective of normative systems, the view from agents towards other agents in the context of a normative system, and the view of normative systems from the perspective of other normative systems. This point has not been discussed in our framework before.

2. The behavior of agents and normative systems can be modeled as recursive games between agents in a standard game theory. It has been shown in [7] that many subtle forms of fraud and deception can be modeled in this way.

3. Obligations can be modeled in the standard BDI setting using the metaphor "Your wish is my command". Thus, the wishes of a normative agent become commands for the subjects of the normative system. This can be explained using the social delegation metaphor.

4. The model allows a distinction to be made between violability and sanctions. Behavior which counts as a violation is distinguished from behavior that is sanctioned. The normative system may autonomously decide which behavior counts as a violation, and whether violations are sanctioned.

5. The model captures several kinds of dynamics of a normative system. Our model combines the dynamics of normative systems with the dynamics of multiagent systems, and explains what roles an agent can play in the dynamics of normative systems.

6. The model can capture the most important aspects of dialogue games, i.e., roles of participants, dialogue game rules (conditional obligations) and the status of the dialogue context (institutional facts). However, more research is needed to fully incorporate dialogue games into normative multiagent systems.

We believe that this conceptual model of virtual organizations as normative multiagent systems may have important consequences for the design and evaluation of virtual organizations. Crucial is that norms are dealt with explicitly, and that system designs do not assume that norms – protocols, standards, contracts – will automatically be obeyed. Norms will almost certainly be violated if there is an incentive to do so. And since virtual organizations are open systems, in the sense that they are inhabited by both human and artificial agents of whom nothing may be known in advance, it is better to be prepared. Modeling violations explicitly makes it possible to design countermeasures.

References

[1] J.L. Austin. *How to do things with words*. Harvard University Press, Cambridge Mass., 1962.

[2] G. Boella, J. Hulstijn, and L. van der Torre. Persuasion strategies in dialogue. In F. Grasso and C. Reed, editors, *Proceedings of the ECAI workshop on Computational Models of Natural Argument (CMNA'04), Valencia.*, 2004.

[3] G. Boella and L. Lesmo. A game theoretic approach to norms. *Cognitive Science Quarterly*, 2(3-4):492–512, 2002.

[4] G. Boella and L. van der Torre. Contracts as legal institutions in organizations of autonomous agents. In *Proceedings of the Third International Joint Conference on Autonomous, Agents and Multi Agent Systems (AAMAS'04)*, 2004.

[5] G. Boella and L. van der Torre. Regulative and constitutive norms in normative multiagent systems. In *Procs. of 9th International Conference on the Principles of Knowledge Representation and Reasoning*, Whistler (CA), 2004.

[6] G. Boella and Leendert W. N. van der Torre. Attributing mental attitudes to normative systems. In *Proceedings of the Third International Joint Conference on Autonomous, Agents and Multi Agent Systems (AAMAS'04)*, pages 942–943, 2003.

[7] G. Boella and Leendert W. N. van der Torre. Normative multiagent systems. In *Proceedings of the AAMAS-workshop on Trust in Agent Societies (TRUST'04)*, 2004.

[8] Guido Boella and Leendert W. N. van der Torre. Decentralized control: Obligations and permissions in virtual communities of agents. In *Proceedings of the Fourteenth International Symposium on Methodologies for Intelligent Systems (ISMIS 2003)*, pages 618–622, 2003.

[9] Guido Boella and Leendert W. N. van der Torre. Local policies for the control of virtual communities. In *Proceedings of the IEEE/WIC Web Intelligence Conference*, pages 161–167, 2003.

[10] Bratman. *Intentions, plans, and practical reason*. Harvard University Press, Harvard (Massachusetts), 1987.

[11] H.-J. Brckert, K. Fischer, and G. Vierke. Holonic transport scheduling with TeleTruck. *Applied Artificial Intelligence*, 14(7):697725, 2000.

[12] J. Breuker, A. Valente, and R. Winkels. Legal ontologes: A functional view. In *Proceedings of the 1st Workshop on Legal Ontologies (LegOnt)*, page 2336, 1997.

[13] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly*, 2(3-4):428–447, 2002.

[14] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Specifying multiagent organizations. In *Proceedings of DEON'04*, 2004.

[15] Cristiano. Castelfranchi. Modelling social actions for ai agents. *Artificial Intelligence*, 103:157–182, 1998.

[16] Herbert H. Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.

[17] Chrysanthos Dellarocas. Analyzing the economic efficiency of ebay-like online reputation reporting mechanisms. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 171 – 179. ACM, 2001.

[18] B. S. Firozabadi and M. Sergot. Contractual access control. In *Procseedings of the 10th International Workshop on Security Protocols*, 2002.

[19] I. Foster, C. Kesselman, and S. Tuecke. The anatomy of the grid: Enabling scalable virtual organizations. *International Journal of Supercomputer Applications*, 15(3):?–?, 2001.

[20] D. Grossi, F. Dignum, L. Royakkers, and J-J. Ch. Meyer. Collective obligations and agents: Who gets the blame? In *Proceedings of DEON'04*, 2004.

[21] B. J. Grosz, S. Kraus., D. G. Sullivan, and S. Das. The influence of social norms and social consciousness on intention reconsiliation. *Artificial Intelligence*, 142(2):147–177, 2002.

[22] B. Hansson. An analysis of some deontic logics. *Nôus*, 3:373–398, 1969.

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

[24] Will Hill, Larry Stead, Mark Rosenstein, and George Furnas. Recommending and evaluating choices in a virtual community of use. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI-95)*, pages 194 – 201. ACM, 1995.

[25] J. Hulstijn and N. Maudet. Uptake by conditional obligations. In P. Dekker and R. van Rooy, editors, *Proceedings of the 14th Amsterdam Colloquium*, pages 141 – 146. ILLC, University of Amsterdam, 2003.

[26] Joris Hulstijn. *Dialogue Models for Inquiry and Transaction*. PhD thesis, University of Twente, Enschede, 2000.

[27] A.J.I. Jones and M. Sergot. A formal characterisation of institutionalised power. *Journal of the Interest Group in Pure and Applied Logic*, 3:427–443, 1996.

[28] Ronald M. Lee. Bureaucracies as deontic systems. *ACM Transactions on Information Systems*, 6(2):87 – 108, 1988.

[29] D. Makinson and L. van der Torre. Input/output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.

[30] D. Makinson and L. van der Torre. Constraints for input/output logics. *Journal of Philosophical Logic*, 30:155–185, 2001.

[31] Peter McBurney and Simon Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11(3):315–334, 2002.

[32] J.-J. Meyer and R. Wieringa. *Deontic Logic in Computer Science: Normative System Specification*. John Wiley & Sons, 1993.

[33] Scott Moss, Helen Gaylard, Steve Wallis, and Bruce Edmonds. SDML: A multi-agent language for organizational modelling. *Computational and Mathematical Organization Theory*, 4(1):43–70, 1998.

[34] S. Neal, J. Cole, P. F. Linington, Z. Milosevic, S. Gibson, and S. Kulkarni. Identifying requirements for business contract language: a monitoring perspective. In *Proceedings of the Seventh International Enterprise Distributed Object Computing Conference (EDOC03)*, page 5062. IEEE press, 2003.

[35] H. Van Dyke Parunak. A practitioners' review of industrial agent applications. *Autonomous Agents and Multi-Agent Systems*, 3(4):389–407, 2000.

[36] L. Pearlman, V. Welch, I. Foster, C. Kesselman, and S. Tuecke. A community authorization service for group collaboration. In *Proceedings of the IEEE 3rd International Workshop on Policies for Distributed Systems and Networks (POLICY'02)*, pages 50–59. IEEE Press, 2002.

[37] Anand S. Rao and Michael P. Georgeff. Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the International Workshop on Knowledge Representation (KR91)*, pages 473–484. Morgan Kaufmann, San Mateo CA, 1991.

[38] John R. Searle. *Speech acts: an Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, 1969.

[39] J.R. Searle. *The Construction of Social Reality*. The Free Press, New York, 1995.

[40] R. Tuomela. *Cooperation: A Philosophical Study*. Kluwer Academic Pulishers, Dordrecht, 2000.

[41] D. N. Walton and E. C. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, 1995.