MODELING NORMS IN ELECTRONIC INSTITUTIONS

Frank Dignum
dignum@cs.uu.nl
Utrecht University

Ismail Khalil Ibrahim
ismail@cs.uu.nl
University of Ghent
Netherlands

Edgar Weippl
edgar.weippl@scch.at

Wieland Schwinger
wieland.schwinger@scch.at

Software Competence Center Hagenberg
Hauptstr. 99, A-4232 Hagenberg
Austria

KEYWORDS

Agents, Multi agent systems, Norms, Cooperation

ABSTRACT

Two approaches have been advocated for the design and modeling of (social) norms in multi agent systems: in the coordination strategy, multi agent systems are defined as a set of entities regulated by mechanisms of social order and created by more or less autonomous actors to achieve common goals; in the cooperation strategy, agents model specific roles in the society and interact with each other as means to accomplish their goals. In this paper, we argue that there is a relative duality between the two approaches with respect to their use of norms as constraints on the social behavior of multi agents systems. We present and discuss a variant of an existing framework for modeling multi agent systems in an environment governed by norms for the cooperation strategy.

INTRODUCTION

Institutions, Norms and Multi-Agent Systems

The study and modeling of norms has attracted the interest of scientists from different disciplines such as sociology, economics, psychology, and computer science.

According to Encyclopedia Britannica, norms are rules or standards of behavior shared by members of a social group.

Several researchers have recognized that the design of agent societies can benefit from abstractions analogues to those employed by our robust and relatively successful societies and organizations [11]. There is a growing body of work that touches upon the concepts of norms and institutions in the context of multi agent systems [11, 12].

Human interactions very often follow conventions [12] that is, general agreements on language, meaning and behavior. By following conventions humans decrease uncertainties about the behavior of others, reduce conflicts of meaning, create expectations about the outcome of the interaction and simply the decision process by restricting to a limited set the potential actions that may be taken. These benefits explain why conventions have been so widely used in many aspects of human interaction: trade, law, games, etc.,

In most societies, norms are backed by a variety of social institutions [11] that enforce law and order (e.g., courts, police), monitor for and respond to emergencies (e.g., ambulance system), prevent and recover from unanticipated disasters (e.g., coast guard, fire-fighters), etc.,

The benefit of an institution resides in its potential to lend legitimacy and security to its members by establishing norms.

In this way civilized societies allow citizens to utilize relatively simple and efficient rules of behavior, offloading the prevention and recovery of many problem types to social institutions that can handle them efficiently and effectively by virtue of their economies of scale and widely accepted legitimacy. Successful civil societies have thus achieved a division of labor between individuals and institutions that decreases the barriers to survival for each citizen, while helping increase the welfare of the society as a whole.

The electronic counterpart of the physical institution does a similar task for software agents: it can engender trust through certification of an agent and by the guarantees that it provides to back collaboration [9, 10, 11, 12]. However, the electronic institution can also function as the independent place, in which all types of agent independent information about the interaction between the agents within the society is stored e.g., it defines the message types that can be used by the agents in their interactions, the rules of encounter, etc.,

In general, institutions enable to 1) specify the coordination structure that is used, 2) describe exchange mechanisms of the agent society, 3) determine interaction and communication forms within the agent society, 4) facilitate the perception of individual agents of the aims and norms of an agent society and 5) enforce the organizational aims of the agent society.

Modeling Norms in Multi-Agent Systems

Norms are expectations about what behavior, thoughts, or feelings are appropriate within a given group within a given context.

Two approaches have been advocated for the design and modeling of norms into multi agent systems: coordination [11] and cooperation [9].

Following the basic classification of coordination models from organizational theory, coordination in agent societies can be divided into markets, networks, and hierarchies. Different coordination models result in different frameworks for agent societies. The overall goals of a society are domain dependent but all societies depend on a facilitation layer that provides the social backbone of the organization. The objectives of the facilitation layer are the organization of the society itself and are dependent on the underlying coordination model and on the norms and conventions that hold in the domain. Social coordination describes the way interactions between roles are organized and the way the interface between the society and the outside world is defined. That is, the coordination model determines the institutional roles, social norms, and interactions forms in the society.

Cooperation models are based on the assumption that agents have some joint goal or intension. Such a joint goal enforces some type of cooperative behavior on all agents. The conventions according to which the agents coordinate their behavior is hard-wired into the protocols that the agents use to react to the behavior (cq. messages) of other agent.

Three levels of social behavior of an agent can be distinguished in the cooperation model: conventions, contract, and private levels. The level of conventions between agents can be compared with the prima facie obligations that arise from law. They provide a kind of moral background against which agents interact. The contracts level indicates how obligations and authorizations arise, how they are fulfilled (or expires) and what happens if they are violated. On the private level agents make private judgments between different obligations and/or goals and determine the actions they will take.

The coordination and cooperation strategies can be qualitatively and quantitatively compared in terms of their adequacy (1) to model a particular institution, (2) to cope with the autonomy of the agents, and the ability to (3) conform to norms.

Both coordination and cooperation models provide a setting for agent societies by setting out the goals of the society and the roles (what you can do) needed to achieve those roles. Institutions will enforce these models by setting out the scenes (where you can do it) and protocols (what you can say) for interaction in the society.

Example Scenario

The simplified example in this subsection attempts to illustrate the differences and commonalities between the coordination and cooperation strategies:

Example 1

Suppose we wish to model the following norms through an institution:

- norm 1 before participating in an auction, buyers and sellers are obliged to register.
- norm 2 if bid accepted in an auction, buyer is obliged to pay
- norm 3 if payment received in settlement, seller is obliged to deliver good
- norm 4 if has no obligation, buyer or seller may leave.

From this simplified scenario, we can abstract what we call scenes, one for the registration of buyers and sellers, another for the auction scene (where a standard downward bidding/Dutch auction format is employed), and the settlement scene where payments and delivery of goods can take place. Thus, for each activity that can

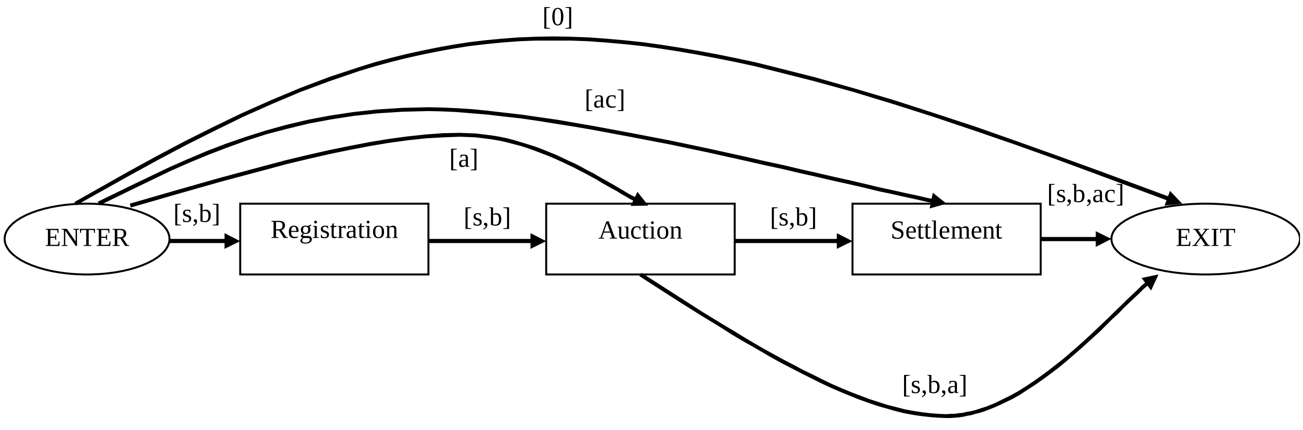

Figure 1: Institution Model of the Example Scenario.

take place in the institution, there is a corresponding scene, in which interactions between agents are articulated through agent group meetings that follow a well-defined protocol. This set of scenes and the connections between them – what roles agents may play in them, how many of each role, to which scenes they may move – constitute the performative structure for the electronic institution (Figure 1). The purpose of this diagram is to show the different scenes, which comprise the institution by means of a transition graph. Thus, the circle on the left hand denotes the start scene and that on the right hand side is the end scene. In between, there are scenes and arcs connecting them. The arcs are labeled by agent variable: role pairs , where a = auctioneer, b = buyer, s = seller, and ac = accountant. Agents enter the institution via the enter node and follow the paths assigned to the roles they have adopted.
Using a coordination approach, we model our solution in the following way:

$s(A, R) \leftarrow s_1(A_i,B_i,C_i), s_2(A_j,B_j,C_j), \ldots$

which states that for a specific role R, agent A is allowed to exist in the scenes $s_1, s_2, \ldots$
This definition is sufficient. However, in order to allow optimized modeling, we may want to state explicitly additional knowledge and assumptions about the scenes. For instance, we use a right arrow for norms in order to distinguish them from scene definitions if necessary.

- Registered agents are allowed to enter the auction
$reg(A_i,-,-) \rightarrow \exists B_j \exists C_j\ auc(A_j,B_j,C_j)$
- Agents in the auction scene are allowed to enter s:
$auc(A_i,-,-) \rightarrow \exists B_j \exists C_j\ sett(A_j,B_j,C_j)$

Using a cooperation approach, we model our solution in the following way: the two scenes are defined on the global model s(Agent, Role)

- $s_1(A_i,B_i,C_i) \leftarrow$ s(Agent, Role)
- $s_2(A_j,B_j,C_{j,}) \leftarrow$ s(Agent, Role)

To be consistent with the representation, which was captured in the first scenario, we can explicitly express additional knowledge stating that the decomposition is lossless. More specifically, in this example, the fact that A_i is the same agent as Aj.

- $s_1(A_i,B_i,C_i), s_2(A_j,B_j,C_{j,}) \rightarrow A_i = A_j$

The reader may have noticed the relative duality between the scene definitions and the norms in the two scenarios. In the first scenario, the scene definition constructs the global model s as a join: in a sense, by construction s1 and s2 are a lossless decomposition. In the second scenario, by construction of s1 and s2 the inclusion of s1 and s2 with respect to the set of norms must be verified by definition of the scenes under the Closed World Assumption.

Structure of the Paper

The reminder of the paper is organized as follow: in the next section we review the literature relevant to this line of work. In section 3, we present a transformation of the institution that allow us to formulate the cooperation strategy as an institution. Finally, we conclude with some remarks on our plans to design and implement the institution.

BACKGROUND

There is an extensive literature about agent theories concerning beliefs, goals and intension. However, there is not much theory available to incorporate norms into the behavior of agents [8]. On one hand, there is work on normative agents – that are agents, which have an explicit representation of norms and can reason about whether accepting and fulfilling them- but of an experimental nature and for the purpose of social simulation. In this type of work norms are built into the agent. The agent cannot change its behavior over time, based on experience. On the other hand, there are more complex normative agents for multi agents systems, mainly with the purpose of reducing or transaction costs but in these agents norms are simply built in constraints in the agent's architecture [18, 19] or rules and protocols the agent necessarily applies [14].
Boman [1] introduces norms in his agent architecture to overcome serious limitations of rational decision-making. However, in this architecture norms act only from outside the decision maker: they don't generate goals or meta-criteria to be taken into account during the decision. Either they simply modify the decision parameters, or they post hoc filter decisions and actions. Thus, we can neither say that norms are explicitly represented and reasoning about them takes place, nor that the agent describes deliberates to follow or violate a norm. The agent cannot really violate a norm, which is in fact just a complex constraint. As for [15] they take into account, for example, the collective interest in the agent's decision, but they do not account for the normative origin or the character of this goal: it is simply a pro-social attitude of the agent.
Conte and Castelfranchi proposed in [4, 6] a cognitive approach to norms in artificial agents, where norms are conceived as external (expectations, behaviors, and prescriptions) and internal (i.e., mental) entities. They show how norms are acknowledged and issued by the agents, and how they are translated into as normative beliefs and produce normative goals. They also characterize different kinds of norm adoption (parallel to goal adoption) based on different attitudes and motives about adopting the norm.

In [2], an explicit model is introduced for norm processing within an autonomous deliberate agent; their relations with beliefs, decisions, goals, plans, and actions. In other terms, a process model is presented formalizing how norms succeed in influencing the agent's behavior, although being possibly violated.
Of course, an important theory that could be used to incorporate norms into the agent theory is that of deontic logic [13, 16, 20]. A first attempt has been made in [9]. In this work several types of norms are distinguished and translated into obligations for the agent. All the obligations result into conditional goals for the agent. The decision whether to comply to a norm or not is made by ranking the goals. If the goal resulting from a norm is ranked on top the norm will be complied with, otherwise it might be violated. The theory does not provide explicit reasoning about complying with a norm or violating it, nor does it provide an operational architecture.
On the contrary, our approach takes a cooperation approach for modeling norms in electronic institutions which has the advantage that the designer of the modeling solution does not need to worry about how agents are implemented and which protocols are hard wired.

NORMS AS COOPERATION STRATEGY

Cooperation defined as an institution

For our purpose, an institution is solely defined by the name of the institution and its arity (number of scenes).
As we have seen cooperation strategy models the institution as a schema

$E = (A, I, S, N)$

where:

E is the name of the institution
A is the set of agent/role pairs
I is the union of local illocution schemas in the form of *<P, S, R, C>* where P is an illocutionary particle (request, accept, deny, inform, or pay), *S* is the sender agent identifier (agent/role pair), *R* is the receiver agent identifier (agent/role pair), and *C* is the content of the message.
S is the set of scene definitions for the component scenes in terms of the institution
and *N* is a set of norms on the component scenes.

Given an instance of the institution *A* i.e., a set of tuples for the Agent/Roles combinations, an instance of the institution is defined by the minimal model of $S \cup A$.
An instance of the institution is consistent if it is a model of *N*. In our case, we only need to verify that *A* is also a model of *N*. Since, the scenes are only materialized, say *J* is the actual instance of the *illocution schema*, an instance of the institution is an instance of $A \cup J$ such that it corresponds to the minimal model $S \cup A$, and *A* is consistent with *N*. Notice that *A* may not be unique. For the sake of simplicity, we will assume that the modeler of the solution has been careful and that a minimal instance is guaranteed to exist.

3.2 A dual view of the institution

In order to comply with the reality of the situation in the modeling, we would need to construct an institution $E'' = (A, I, S'', N'')$ with the same instances as *E*.
Using similar transformation as the one used in [12], we construct the institution as
$E' = (A \cup I, 0, 0, N')$ which has the same instances as *E*.
Given a scene definition for *s*:

- $s(\overline{X}) \leftarrow B(\overline{Y})$

we first make all the implicit quantifications explicit

- $\forall \overline{X}\ s(\overline{X}) \leftarrow \exists (\overline{Y} - \overline{X}), B(\overline{Y})$

the completed axiom is

- $s(\overline{X}) \Leftrightarrow \exists (\overline{Y} - \overline{X}), B(\overline{Y})$

which we can transform into two norms

- $s(\overline{X}) \rightarrow \exists (\overline{Y} - \overline{X}), B(\overline{Y})$
- (2) $B(\overline{Y}) \rightarrow s(\overline{X})$

if we call $\hat{N}$ the set of norms of type 2, the new institution
$E' = (A \cup I, 0, 0, S \cup N \cup \hat{N})$
We claim but do not prove that the original and transformed institutions have the same instances.
Let us look at the application of the transformation on an example.

Example 2

Let us consider the following institution
- $INST = (\{w/3\}, \{v/2\}, \{v(X,Y) \leftarrow p\{X,Y,Z)\}, 0\}$ where w is the name of an institution with three scenes, two illocution schemas have been uttered which lead that agents X, Y have moved from scene p to scene v and there are no norms specified for scene p. According to the transformation of the institution we have described, we generate the following two norms
- $n_1: p(X, Y, Z) \rightarrow v(X, Y)$

- n_2: $v(X, Y) \rightarrow \exists Z, w(X, Y, Z)$
-

We are now considering the institution

- $INST' = (\{w/3, v/2\}, 0, 0, \{n_{1,} n_2\})$

We also apply the transformation to our motivational example.

Example 3 (Example 1 cont.)

According to the transformation of the institution we have described, we generate the following four norms:

- $n_1 = reg(A_i,-,-) \rightarrow \exists B_j \exists C_j\ auc(A_j,B_j,C_j)$
- $n_2 = s_1(A_i,B_i,C_i) \leftarrow s(Agent, Role)$
- $n_3 = auc(A_i,-,-) \rightarrow \exists B_j \exists C_j\ sett(A_j,B_j,C_j)$
- $n_4 = s_2(A_j,B_j,C_{j,}) \leftarrow s(Agent, Role)$

we are now considering the institution schema as:

$INST' = (\{I/3, s/2\}, 0, 0, \{ n_1, n_2, n_3, n_4, n_{5\}}$

where n_5

- $s_1(A_i,B_i,C_i), s_2(A_j,B_j,C_{j,}) \rightarrow A_i = A_j$

CONCLUSIONS AND FUTURE WORK

Looking at the structure of the institution we can anticipate the types of interactions involved in interacting in a particular situation with respect to the defined norms. Thus, an institution defines a performative structure and a dialogic framework [12], by which we mean, it prescribes the actions agents can take and when and where they can perform those actions, and determines the form of conversations between agents. Therefore, the way norms and conventions are specified and enforced in a multi agent system depends on the coordination model. In hierarchies [11], norms and conventions can be embedded in the power relations. These relations determine which agent can demand an action from which other agent or which agent has priority over the resources. The controlling agent is supposed to uphold the norms of the society by managing the sub-ordinate agents according to them. In markets, norms and conventions are for a large part embedded in the market mechanism chosen e.g., the auction mechanisms try to ensure that all agents get an opportunity to require a resource relative to their private value of that resource. Cheating by over- or under-bidding does not lead to any benefits for the agent and thus is prevented by the mechanism itself. In network models explicit roles are defined to represent the institution does not lead to any trust and trace the fulfillment of contracts.

The purpose of the project from which this research stems is to build a complete monitoring system for electronic markets. Our objective is to integrate seamlessly the cooperation and coordination models as a basis for modeling electronic institutions.

REFERENCES

1. Boman, M. (1997). Norms as Constraints on Real_time Autonomous Agent Action. In: M. Boman, W. van de Velde (eds.), Multi-Agent Rationality, Proceedings of the 8th European Workshop on Modelling Autonomous Agents in a Multi- Agent World, MAAMAW'97, Lecture Notes in AI, vol. 1237, Springer Verlag, Berlin, 1997, pp. 36-44.
2. Castelfranchi, C., Dignum, F., Jonker, C., and Treur J., (1999). Deliberative Normative Agents: Principles and Architectures, Proc. of the Sixth International Workshop on Agent Theories, Architectures and Languages, ATAL'99. Orlando, pp. 206-220.
3. Cohen, P., and Levesque, H. (1991). Teamwork, Nous, vol.35, pp. 487-512.
4. Conte, R., and Castelfranchi, C., (1995a). Cognitive and Social Action. UCL Press, London.
5. Conte, R., and Castelfranchi, C. (1995b). Understanding the effects of norms in social groups through simulation. In Artificial societies: the computer simulation of social life. Eds. G.N. Gilbert and R. Conte, London, UCL Press.
6. Conte, R. and Castelfranchi, C. (1996). From Conventions to Prescriptions: Toward an integrated Theory of Norms, ModelAge'96 Workshop, Sesimbra, January, 1996 (AI&Law, forthcoming).
7. Conte, R., Castelfranchi, C., and Dignum, F. (1999). Autonomous Norm-Acceptance, In: J.P. Mueller, M.P. Singh, A.S. Rao (eds.), Intelligent Agents V, Proc. of the Fifth International Workshop on Agent Theories, Architectures and Languages, ATAL'98. Lecture Notes in AI, vol. 1555, Springer Verlag, Berlin, Springer Verlag.
8. Conte, R., Falcone, R., Sartor, G. (1999). Agents and norms: How to fill the gap? In: Conte, R., Falcone, R. and Sartor, G. (eds) Agents and Norms, special issue, AI & Law, Vol.7, No.1, pp. 1-15, 1999.
9. Dignum, F. (1996). Autonomous Agents and Social Norms. In: R. Falcone and R. Conte (eds.), ICMAS'96 Workshop on Norms, Obligations and Conventions, Kyoto, pp. 56-71. Revised version in In AI & Law, Vol.7, 1999, No.1, pp. 69-79.
10. Dignum, F. (2001). Agents, Markets, Institutions and Protocols. In: Dignum, F., and Sierra, C., (eds.), Workshop on Agent Mediated Electronic Commerce, LNAI 1991, Springer Verlag, Berlin, pp. 98-114.
11. Dignum, V., and Dignum F. (2001). Modeling Agents Societies: Coordination Frameworks and Institutions, Workshop on
12. Esteva, M., Padget, J., Sierra, C., (2001), Formalizing a language for Institutions and Norms, Proc of the Eighth International Workshop on Agent Theories, Architectures, and Languages, ATAL-2001, Seattle.
13. Hilpinen (ed.) (1971). Deontic Logic: Introductory and Systematic Readings. Reidel.

14. Jennings. N.R., (1993). Commitments and conventions: The foundation of coordination in multi-agent systems. The Knowledge Engineering Review, 3, 223-250.
15. Jennings. N.R. and Campos, J.R. (1997). Towards a Social Level Characterisation of Socially Responsible Agents. IEEE Proc. on Software Engineering, vol. 144, 1, pp.11-25
16. Meyer, J.-J., Ch., and Wieringa, R.J. (eds.) (1993). Deontic Logic in Computer Science. John Wiley and Sons Ltd., Chicester, UK.
17. Rao, A.S., and Georgeff, M.P. (1991). Modeling rational agents within a BDI architecture. In: R. Fikes and E. Sandewall (eds.), Proc. of the Second Conference on Knowledge Representation and Reasoning, Morgan Kaufman, pp. 473-484.
18. Shoham, Y and Tenneholtz, M. (1992a). On the synthesis of useful social laws for artificial agent societies (preliminary report). Proceedings of the AAAI Conference, 276-281.
19. Shoham, Y. and Tenneholtz, M. (1992b). Emergent conventions in multi-agent systems: Initial experimental results and observations. Proceedings of the 3rd International Conference on KR&R. Cambridge, MA, 225-232.
20. Wright, G. von (1951), Deontic Logic. Mind, vol. 60, pp. 58-74.

BIOGRAPHIES

FRANK DIGNUM got his Ph.D. in 1989 from the Free University of Amsterdam. After this he set up the Computer Science department of the University of Swaziland. After one year in the AI department of the Technical University of Lisbon he has worked at the Eindhoven University of Technology until September 2000. In Eindhoven he initiated an e-commerce research programme ("Enabling e-commerce") and also initiated an Esprit project (MeMo) that is geared to support SME's on electronic marketplaces. Since September 2000 he is an associate professor at the Utrecht University. His main research interests lay in the area of formal models of communication and social relations. The use of several modal logics is of prime importance in this research. The results are applied in the area of electronic commerce and multi-agent systems.

ISMAIL KHALIL IBRAHIM is a member of the Intelligent Systems Group at the Depatment of Information and Computing Sciences of Utrecht University. He holds a B.Sc. in Electrical Engineering, from the University of Technology, Iraq (1985), an M.Sc., and a Ph.D. (cum laude) in Computer Engineering and Information Systems from Gadjah Mada University, Indonesia.

EDGAR WEIPPL is project manager at the Software Competence Center Hagenberg. He received his PhD in Computer Science from the Vienna University of Technology. Besides teaching Database Systems and an introductory course on Security he focused his research on security in mobile applications.

WIELAND SCHWINGER recently graduated from the PhD program in Business Informatics writing a thesis on Modelling of Ubiquitous Web Applications. He is currently project manager at the SCCH.