ICMAS '96: Norms, Obligations, and Conventions

Rosaria Conte and Rino Falcone

The use of a normative vocabulary and concepts (obligations, deontic operators, norms, and so on), as well as the insertion of norms in knowledge bases for expert systems, has a well-established tradition in many AI subfields (legal expert systems, artificial normative reasoning, and so on). In adjacent domains (logical philosophy, social philosophy, decision theory), both legal and social norms have received considerable, if not satisfactory, attention.

The advent of large communication networks, civic networks, and so on, has contributed dramatically to the attention given by the scientific community to several normative issues: authorization, access regulation, privacy maintenance, respect of decency, and so on, not to mention the more obvious problems associated with the regulation of the use and purposes of networks.

In the multiagent system (MAS) discipline, social norms and laws are perceived to help improve coordination and cooperation (Conte and Castelfranchi 1995; Jennings 1995; Walker and Wooldridge 1995; Jennings and Mandami 1992; Shoham and Tennenholtz 1992). Indeed, the efforts done by MAS researchers and designers to construct autonomous agents (Wooldridge and Jennings 1995) carry a number of interesting but difficult tasks, including how to avoid interferences and collisions among agents moving around and separately acting in a common space and how to ensure that negotiations and transactions fulfill the norm of reciprocity. Imagine a software assistant delegated to conduct transactions on behalf of its user. Because of its loyalty (benevolence), the assistant will behave as a shark with regard to potential partners, always looking for the most convenient transaction, thereby infringing on existing commitments. Other difficult tasks, more generally, are how to obtain a robust performance in teamworks (Cohen and Levesque 1990); how to prevent agents from dropping their commitments; or better, how to regulate agents dropping their commitments to a joint action to not disrupt the common activity and preclude the common goal being achieved (Jennings 1995; Singh 1995; Kinny and Georgeff 1991).

These tasks have now entered the MAS field's common knowledge. Other problems are perhaps less obvious. For example, the existence of so-called

---

The Second International Conference on Multiagent Systems (ICMAS '96) Workshop on Norms, Obligations, and Conventions was held in Kyoto, Japan, from 10 to 13 December 1996. Participants included scientists from deontic logic, database framework, decision theory, agent architecture, cognitive modeling, and legal expert systems. This article summarizes the contributions chosen for presentation and their links to these areas.

---

virtual representatives brings about the question of delegation. Software assistants and mobile agents are intended to act as virtual representatives of network clients, but the role of representative implies that some normative mechanism is at work, such as responsibility (Jennings 1992) and delegation (Santos and Carmo 1996). Analogously, the concepts of role (Werner 1990) and role tasks, which are so crucial for the implementation of organizational work, require a model of authorization and (institutional) empowerment (Jones and Sergot 1995).

That social norms are of vital importance for AI is witnessed by the variety of subcommunities dealing with normative concepts and questions: legal expert systems, artificial normative reasoning, AI and law, deontic logics for computer science, MAS, and distributed artificial intelligence (DAI) are some examples. However, these communities suffer from a poor level of communication and confrontation among one another.

Bringing together people from these communities was one major reason of interest behind the Second International Conference on Multiagent Systems (ICMAS'96) Workshop on Norms, Obligations, and Conventions (held in Kyoto, Japan, on 10–13 December 1996). Both the program committee and the contributors included scientists from different backgrounds (deontic logic, database framework, decision theory, agent architecture, cognitive modeling, legal expert systems, and so on). The discussion addressed several issues: (1) What is the relationship between action and decision? Often action is reduced to decision (that is, a choice among one's preferences), but how are preferences formed when they are not built into the system? (2) Should norms be represented as a specific mental object, and if so, how? What is the role and nature of such an object, and how do we describe it? (3) How and why can a self-interested agent decide to comply with norms? (4) Commitment is crucial for teamwork robustness. However, many authors (Jennings 1995; Kinny and Georgeff 1991) have shown that we need a flexible notion of commitment. How do we reconcile flexibility and commitment, autonomy and obedience? (5) Present theories (Rao and Georgeff 1991; Cohen and Levesque 1990) seem to share a notion of a commitment to communicate one's possible decision to abandon a joint action rather than a commitment to the action as such. Why? Don't we need a more general definition of commitment that applies to both notions? Let us summarize the contributions chosen for presentation and their links to

Only if agents acknowledge that some actions or world states are enforced by (social or legal) norms will they be able to apply norms consistently and influence other agents in their neighborhood to do the same.

some of the issues mentioned previously.

As for the question of whether norms should be treated as specific mental objects, Frank Dignum (Eindhoven University of Technology, The Netherlands) addressed it with a deontic-based approach. He suggested some good ideas about an agent architecture incorporating deontic operators in a beliefs-desires-intentions framework. He used the speech-act theory to model the generation of norms and deontic logic to model the concepts that are necessary for autonomous agents in an environment that is governed by norms. Using deontic logic, the author not only explicitly described those norms that can be used to implement the interactions among agents but also both norm violations and possible reactions to such violations.

Cristiano Castelfranchi (IP-CNR, Rome, Italy) showed how the representation of the hearer's mind in the speaker's mind is, in fact, much richer than is usually supposed. While applying this point of view to normative prescriptions, the author argued that what is required by a norm is not only a given behavior but also a mental attitude; therefore, the real task that we are facing is how to model normative minds rather than mere behaviors. The concluding remark of the author was that under any circumstance, a norm is aimed at influencing the agent, that is, at changing his/her goals: Norms should lead not only to factual conformity but also to cognitive obedience.

Commitment and flexibility in commitment were addressed by both Milind Tambe (University of Southern California) and Munindar Singh (North Carolina State University). In the paper presented by Tambe, flexible teamwork is considered more than a union of agents' simultaneous execution of individual plans. The

central hypothesis is that for an effective teamwork, agents should be provided explicit team plans and an underlying model of teamwork that explicitly outlines their commitments and responsibilities as participants in team activities.

In his paper, Singh proposed a notion of commitment that satisfies both principles from DAI and those from spheres of control, a conceptual approach (introduced in databases) used for structuring activities. Commitments are an important abstraction for characterizing, understanding, analyzing, and designing MASs. They also arise in distributed databases. However, traditional distributed databases implement a highly restrictive form of commitment, but their modern application requires greater organizational flexibility reflected on more flexible forms of commitment.

Omar Belakhdar (Swiss Federal Institute of Technology, Lausanne, Switzerland) and Jacqueline Ayel (Université de Savoie, France) described how the notion of ontology sharing can be used in MASs to ensure that agents have the same interpretation of the concepts used during a cooperation process. Because the ontologies are specified inside the cooperation protocols shared by the agents, the ontologies are also shared. Sharing these protocols means that the agents are committed to respect the norms and rules implied by them.

The issue of how a set of autonomous agents in an MAS can be forced to act in accordance with norms was addressed by Magnus Boman (Stockholm University and Royal Institute of Technology, Sweden). He proposed a solution in terms of decision theory, implementing norms as input to the evaluations performed by a decision module. Then, no action that violates a norm will be suggested to any agent.

A computer-simulation methodology was utilized by Mario Paolucci (ISTAT, Rome, Italy) to explore the functions of norms in reducing aggression among agents living in a common world scattered with scarce but self-replenishing resources. Three strategies were compared: (1) *utilitarian* (never attack stronger agents), (2) *norm based* (never attack agents eating their own food), and (3) *ultimatum based* (attack only agents refusing to share their food with you). The work aimed at exploring the role of an antisocial strategy based on an ultimatum and comparing it with both a prosocial strategy based on norms and a merely utilitarian strategy.

Although the issues raised during the discussion could hardly have received conclusive treatment, the participants seem to have reached a sort of agreement on a number of issues. At the metatheoretical level, the following were acknowledged: First, because of the emphasis on autonomy, MASs provide a fundamental test bed for the study of norms. Second, the logic-based approach (for example, deontic logic) is a useful instrument for speaking about social norms, provided that it is a language for describing not only the effects of norms (compliance versus violation) but also the process leading to them and that operators for obligations be interfaced with operators for mental states and possibly integrated into an agent architecture. Third, decision theory has much to say about social norms and especially self-interested (in the sense of autonomous) normative decision. Actually, it is a necessary, but insufficient, approach.

The reasons for these statements, especially the last evaluation, become clear when looking at the theoretical conclusions reached by the discussion: First, the efficacy of norms depends on their being acknowledged as such by the agents. Only if agents acknowledge that some actions or world states are enforced by (social or legal) norms will they be able to apply norms consistently and influence other agents in their neighborhood to do the same. Second, it is insufficient to implement social norms as action constraints (Shoham and Tennenholtz 1992) be-

cause this solution does not allow for both the violation and the acquisition of the norms. Third, a suggestion coming from decision theory is to define norms as input to the agents' subjective assessments of their utilities. This solution deals nicely with norm violation. However, two questions are still open: (1) How is utility assessment modified by means of norms? (2) How are alternatives for decision acquired?

These questions, as well as others, seem to show that a promising research direction is opening up in the field: the autonomous normative agent model and architecture. Confirming the interest that this research direction has received, a special issue of *AI & Law* entitled "Agents and Norms" is forthcoming.

References

Cohen, P., and Levesque, H. 1990. Intention Is Choice with Commitment. *Artificial Intelligence* 42(3): 213–261.

Conte, R., and Castelfranchi, C. 1995. *Cognitive and Social Action.* London: UCL.

Jennings N. 1995. Commitment and Conventions: The Foundation of Coordination in Multiagent Systems. *Knowledge Engineering Review* 8:135–148.

Jennings, N. 1992. On Being Responsible. In *Decentralized Artificial Intelligence 3*, 93-102. Amsterdam: Elsevier Science.

Jennings, N. R., and Mandami, E. H. 1992. Using Joint Responsibility to Coordinate Collaborative Problem Solving in Dynamic Environments. In Proceedings of the Tenth National Conference on Artificial Intelligence, 269–275. Menlo Park, Calif.: American Association for Artificial Intelligence.

Jones, A. J. I., and Sergot, M. 1995. Norm-Governed and Institutionalized Agent Interaction. Paper presented at ModelAge'95: General Meeting of ESPRIT wg 8319, 22–24 January, Sophia Antipolis, France.

Kinny, D., and Georgeff, M. 1991. Commitment and Effectiveness of Situated Agents. In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, 82–88. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.

Rao, A. S., and Georgeff, M. P. 1991. Modeling Rational Agents within a BDI Architecture. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning,* eds. J. Allen, R. Fikes, and E. Sandewall, 473–485. San Francisco, Calif.: Morgan Kaufmann.

Santos, F., and Carmo, J. 1996. *Indirect Action, Influence, and Responsibility.* In *Deontic Logic, Agency, and Normative Systems,* eds. M. Brown and J. Carmo, 194–215. New York: Springer Verlag.

Shoham, Y., and Tennenholtz M. 1992. On the Synthesis of Useful Social Laws in Artificial Societies. In Proceedings of the Tenth National Conference on Artificial Intelligence, 276–282. Menlo Park, Calif.: American Association for Artificial Intelligence.

Singh, M. P. 1995. Multi-Agent Systems: A Theoretical Framework for Intentions, Know-How, and Communications. Lecture Notes in Computer Science, Volume 799. New York: Springer Verlag.

Walker, A., and Wooldridge, M. 1995. Understanding the Emergence of Conventions in Multiagent Systems. In *Proceedings of the First International Conference on Multi-Agent Systems,* 384–389. Cambridge, Mass.: MIT Press.

Werner, E. 1990. Cooperating Agents: A Unified Theory of Communication and Social Structure. In *Distributed Artificial Intelligence, Volume 2,* eds. L. Gasser and M. N. Huhns, 84–101. San Francisco, Calif.: Morgan Kaufmann.

Wooldridge, M., and Jennings, N., eds. 1995. *Intelligent Agents.* Lecture Notes in Artificial Intelligence, Volume 890. New York: Springer Verlag.

Rosaria Conte is a cognitive scientist. She is responsible for the Division of AI, Cognitive, and Interaction Modeling at the Institute of Psychology of the Italian National Research Council. Her main research interests include multiagent systems, computer simulation, and formal models of mental attitudes. Her e-mail address is rosaria@pscs2.irmkant.rm.cnr.it.

Rino Falcone is a physicist. He has worked in a research department of Olivetti and is now a researcher for the Division of AI, Cognitive, and Interaction Modeling at the Institute of Psychology of the Italian Research Council. His research interests range from natural language processing to multiagent systems and agent theory. His e-mail address is falcone@pscs2.irmkant.rm.cnr.it.