Introduction to Normative Multiagent Systems

Guido Boella[⋆]
[⋆]Dipartimento di Informatica
Università di Torino
Italy
guido@di.unito.it

Leendert van der Torre[†]
[†]CWI Amsterdam and
Delft University of Technology
The Netherlands
torre@cwi.nl

Harko Verhagen[‡]
[‡]Dept of Computer and Systems Sciences
Stockholm University \ KTH,
Forum 100, SE-16440 Kista, Sweden
verhagen@dsv.su.se

Abstract

In this paper we give a short introduction to the emerging area of normative multiagent systems by presenting definitions and examples.

1 Introduction

Normative multiagent systems as a research area is best defined as the intersection of two established fields: normative systems and multiagent systems. This still leaves a lot of room for interpretation, as there are various definitions of both areas. However, it is clear that the intersection involves several issues which are highly relevant nowadays, such as coordination, multiagent organizations and agent security. This is witnessed both in the field of normative systems, where the last workshop on deontic logic in computer science (ΔEON04) had as topic "application to multiagent systems", and in the field of multiagent systems, where various subfields now are addressing normative systems, such as multiagent organization or regulated societies.

In this paper we make some observations with respect to the emerging research area of normative multiagent systems by addressing the following questions:

1. What is a normative multiagent system?
2. What are prototypical examples of normative multiagent systems?

To answer questions, we consider normative multiagent systems at the border of agent theory - both multiagent systems and autonomous agents - and the social sciences - sociology, philosophy, economics, et cetera. The forces at this border are considered in two directions: how do the social sciences influence agent theory, and how does agent theory influence the social sciences?

Social sciences $\Rightarrow$ Agent theory. The social sciences are regularly used in the development of theories and models of multiagent systems. It is used in two ways. The first and most obvious way is the use in agent theory of concepts developed in the social sciences, such as co-ordination, organization, convention, norm, trust, et cetera. A second and less popular way, but often at least as useful, is to contrast agent theory with social theory, based on the distinctions between artificial systems and humans. For example, humans cannot be programmed such that they never violate a norm or always co-operate, but artificial systems can.

Agent Theory $\Rightarrow$ Social sciences. According to Castelfranchi (1998), agent theory should also produce theories, models, and experimental, conceptual and theoretical new instruments, which can be used to revise and develop the social sciences. He summarises this point by stating that agent theory - and the related area of artificial intelligence - is not just an engineering discipline, but it is also a science.

The layout of this paper follows the questions. First we give some definitions, and then we discuss some examples.

2 Definitions

In this section we first introduce normative systems, then we consider norms in sociology, and finally we consider multiagent systems.

2.1 Normative systems

Normative systems have traditionally been studied by legal philosophers like Alchourròn and Bulygin (1971).

Meyer and Wieringa, who founded in 1991 the deontic logic in computer science workshops (known as the ΔEON workshops), define normative systems as

> "systems in the behavior of which norms play a role and which need normative concepts in order to be described or specified" (Meyer and Wieringa, 1993, preface).

They also explain why normative systems are intimately related with deontic logic:

> "Until recently in specifications of systems in computational environments the distinction between normative behavior (as it *should be*) and actual behavior (as it *is*) has been disregarded: mostly it is not possible to specify that some system behavior is non-normative (illegal) but nevertheless possible. Often illegal behavior is just ruled out by specification, although it is very important to be able to specify what should happen if such illegal but possible behaviors occurs! Deontic logic provides a means to do just this by using special modal operators that indicate the status of behavior: that is whether it is legal (normative) or not."

Deontic logic was founded by von Wright in 1951 as the formal study of ought. His main idea was that the deontic modalities of obligation and permission were related to each other in the same way as the alethic ones of necessity and possibility are related to each other. Thus, the logic of "It is obligatory to see to it that x" together with "it is permitted to see to it that x" is analogous to the logic of "it is necessary that x" together with "it is possible that x". Permissions are defined in terms of obligations just like possibility it defined in terms of necessity (it is not necessary that the absence of x is the case). However, whereas for necessity often a universal relation is taken (KD45), von Wright gave deontic logic a much weaker system (a weakened version of KD).

Another relation between deontic logic and alethic modal logic was pioneered by Anderson, who defined deontic logic in alethic modal logic with an additional constant (known as Anderson's reduction): "the intimate connection between obligations and sanctions in normative systems suggests that we might profitably begin by considering some penalty or sanction S, and define obligations as: p is obligatory if its falsity entails the sanction S". This was formalized by (a more complex variant of) $O(p) = \Box(\neg p \rightarrow S)$. When people objected that not all violations are sanctioned, Anderson replied that "S just means something bad has happened or a violation has occurred". Much later, Meyer (1988) used a similar reduction to dynamic logic.

Unfortunately, it soon became apparent that it was unclear how to formalize conditionals or rules in these modal systems, and many examples formalized in deontic logic had counterintuitive conclusions, known as the deontic paradoxes. The most notorious ones of them, the so-called contrary-to-duty paradoxes, are concerned with revision of obligations in case of violations.

Von Wright intended that the propositions of his deontic logic referred to actions, it was a logic that described what an actor has to do. Most people reinterpreted it as a system without any agency, but in seventies and eighties many temporal and action logics were introduced. Jones and Carmo (2001) recently define normative systems as, what we will call here, normative multiagent systems:

> "Sets of agents whose interactions are norm-governed; the norms prescribe how the agents ideally should and should not behave. [...] Importantly, the norms allow for the possibility that actual behavior may at times deviate from the ideal, i.e., that violations of obligations, or of agents' rights, may occur."

This does not mean, however, that this agrees with views on norms for multiagent systems. The most common view on norms in multiagent systems is that norms are constraints on behavior via social laws; an alternative view studies cognitively grounded norms. Conte and Castelfranchi (1995) mention three kinds of norms: norms as constraints on behavior, norms as ends (or goals) and norms as obligations. In the following section, we consider social systems.

2.2 Norms in sociology

In this section we analyze the use of norms within sociology. In sociology, the use of norms has been out

of fashion since the 1950's, apart from game theoretically inspired research. In a recent article, Therborn (2002) gives some reasons for this while at the same time presenting an overview of the use of and ideas about norms within sociology during the 1900's.
Within sociology, one can distinguish several action models. Here we present the overview developed by Habermas (1984) of the sociological action models.
The first action model is teleological action. Agents are goal directed and try to maximize their choice of means to obtain a goal. This is the rational choice model. The central issue in this action model is the choices the agent makes between different action alternatives, based on maximizing utility. Agents can thus - try to influence the world, and the rationality of the behavior of the agents can be evaluated with respect to the efficiency of their behavior. Adding other agents, with respect to whom the agent acts in a strategic manner (strategical action model), to the decision making model does not change the ontological principles. The agents may need to model the desires and actions of the other agents but these are still part of the objective world of existing states of affairs. Agents act with respect to this world according to their beliefs about the existing states of affairs and their intentions to bring about desired states of affairs in that world.

The second action model is the normatively regulated action model. Social agents are assumed to belong to a group and follow the norms that are obliged to be followed by members of that group. Following norms is taken as to behave according to expectations. The objective reality is extended by a social reality of obliging norms (acknowledged as such by the group). The rationality of the behavior of an agent is not only related to the objective reality (teleological and strategical action model), but also to the social reality. The conformity between the norms of the group and the behavior of the agents and the relation between the norms and the generalized interests of the agents (and thus if it is wise of the agents to confirm to those norms) are part of this social rationality. Agents act with respect to an objective world and a social world, namely the normative context that defines the possible interactions and legitimate inter-agent relationships between the agents.

The third action model is the dramaturgical action model. In this action model the inner world of the agents is considered. Based on the dramaturgical analysis of social life as developed by Goffman (1959), this action model has as a core the presentation of the self of an agent to an audience. This representation of the self may or may not be truthful. The agent makes use of the fact that its inner self is only admissible to itself. The inner self is defined as the constellation of beliefs, desires, intentions, feelings, and needs of an agent. Habermas views this inner self as a reality in its own right. When presented in a truthful and authentic way, and at the same time connected to the shared evaluation criteria and interpretations of needs, the subjective point of view of the agent can gain an intersubjective value. Truthful is not the same as true in objective sense, opening the door for lying and manipulation or *insincerity*. Agents act with respect to an objective world and a subjective world formed by the totality of subjective experience to which the agent has a privileged access. Examples of application of this action model in the field of MAS include lying agents and believable agents.

The fourth and final action model is the communicative action model. This action model unites the three functions of language specified by the three previous action models. In the strategical action model, language is used by an agent to reach its own goals possibly via influencing other agents by use of language, the normative action model uses language to actualize already existing normative agreements and the dramaturgical model uses language to allow for one to express oneself. In the communicative action model, language is used to bring about mutual understanding on all three previous levels. The agents use language to claim the truth of their utterances, the normative correctness of their speech acts in view of the context, and the sincerity of their intentions being formulated. Testing for rationality of actions is here no longer the privilege of the observer, but is done by the agents themselves to realize a common definition of the situation described in terms of relations between the speech act and the three worlds (i.e., the objective, social, and subjective world) this speech act has relations with. In the cooperative process of interpretation, all participating agents have to incorporate their own interpretation with that of the other agents so that the agents have a sufficiently shared view of the external (i.e., objective and social) world in order to coordinate their actions while pursuing their own goals.

For this article, we will focus upon the normative action model. Following Therborn Therborn (2002) we make some helpfull distinctions. For one, it cannot be stated that all actions that comply with norms can be called normative action in a more strict sense. Different reasons for complying with norms exist.

Normfollowing can be instrumental, where the reasons for compying are either for direct rewards or to avoid the costs of violation. Other reasons can be more socially oriented, such as the desire to belong to a group, not to loose face or esteem, avoid legal punishment, etc. i.e., socially instrumental reasons. Normative action is action where norms are followed for their own sake. This may be out of habit, in an unconscious way, or in a conscious or rational way, based upon an analysis of the consequences of actions within the social world. Other reasons for normfollowing include identification (e.g., with a group, institution or nation) and normfollowing out of self-respect. These reasons represent different levels of internatilization of norms. Norms correlated with self-respect are deeply rooted within the personality of the agent, whereas the identification norms are more shallowly rooted.
We may also look at norms from a functional point of view, what do norms result in? For one we have norms that are of a constitutive nature, they define the agent's membership in a system of action, ahd the system of action at large. Another funcion of norms is regulation, describing what members of a social system must and must not do. Thirdly, norms may have a distributive function, that is how rewards, costs and risks are to be divided among the social system's members.
Independent of the various types of norms, some main issues involved with discussions of norms are norm conformity and norm violoation and the dynamics of norms. If agents are to comply with norms, Norm conformity and violation issues

One characteristic that MAS research and social science, and sociology in particular, share is the interest in the relation between micro-level behaviour and macro-level effects. In MAS research, this boils down to the question "How to ensure efficiency at the level of the multiagent system whilst respecting individual autonomy?". Possible solutions to this problem comprise of:

- use of central control
- internalized control, e.g. the use of social laws Shoham and Tennenholtz (1992).
- structural coordination as proposed in Ossowski (1999)
- a set of norms and learning at all levels, including the level of norms, based on reflecting upon the results of actions.

2.3 Multiagent systems

In agent research or agent system development, omnipotent agents hardly exist, in fact if an agent can be omnipotent, we can do without the concept of agents. Agents are used in software development since knowledge is limited and local autonomy is needed. Also, the single agent paradigm is not really an implementation of agents. The basic agent definition as sketched out by ,e.g., Wooldridge (2002) states that an agent has the following characteristics:

> " it is a computer system that is situated in some environment and that is capable of autonomous action in this environment in order to meet its design objectives "

where autonomy means control over behaviour and internal state. This definition is further developed defining of weak versus strong agency:

> " Intelligent agent (Wooldridge 2002) - weak agency: an intelligent agent is capable of flexible autonomous action
>
> - flexibility meaning: reactivity: interact with environment pro-activeness: take initiative
> - social ability: interact with other agents/ co-operation o autonomy meaning: operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state "

Strong agency uses anthropomorphic notions based on mentalistic properties such as beliefs, desires, intentions, rationality and emotions.

2.4 Norms and multiagent systems

Norms are essential for artificial agents that are to display behavior comparable to human intelligent behavior or collaborate with humans, because the use of norms is the key of human social intelligence. Norms play a central role in many social phenomena such as coordination, cooperation, decision-making, etc. There is an increasing interest in the role of norms in societies, both inside as outside the agent community. The field of (multi)agent research is moving more and more from the individual, cognitive focussed agent models to models of socially situated agents.

Normative multiagent systems combine theories and frameworks for normative systems with multiagent systems. Thus, these systems provide a

promising model for human and artificial agent coordination, because they integrate norms and individual intelligence. They are a prime example of the use of sociological theories in multiagent systems, and therefore of the relation between agent theory and the social sciences, e.g., sociology, philosophy, economics, legal science, etc.

Below we describe our work on norm autonomous agents and our work on normative multiagent systems.

2.4.1 Norm autonomous agents

In the framework developed in Verhagen (2000), norm autonomous agents are described. In short, these agents are based upon cognitive (or goal autonomous agents as developed by Conte & Castelfranchi Conte and Castelfranchi (1995)) and are extended with norms. The agents are part of a normative framework, and at the same time reason about and are able to influence these norms. In this sense norm autonomous agents span the institutional (or even inter-institutional level) where norms get their meaning, the inter-individual level (groups of where norms are produced), and the individual level (where the individual decision making is taking place). These agents choose which goals are legitimate to pursue, based on a given system of norms. The agent has the autonomy of generating its own goals and to choose which it is going to pursue. Besides, the agent is equipped to judge the legitimacy of its own goals and other agents' goals. When a goal conflict arises (not to be confused with interest conflict), the agent may change its norm system thereby changing priorities of goals, abandoning a goal, changing a goal, generating another goal, etc. The reasoning capability of these agents at the level of norms is called normative reasoning. Norm autonomous agents generate norms they can use to evaluate states of the world in terms of whether or not they could be legitimate interests. Legitimacy is a social notion and is in the end determined by the norms of the agent with respect to the agent society it is part of.

2.4.2 Normative multiagent systems

Castelfranchi (1998, 2000) defines several social viewpoints on multiagent systems, in which he conceptualizes a multiagent system in terms of respectively the mind of the agent, the power of the agent, the dependencies between agents, and groups or coalitions. Moreover, he defines abstraction relations between them, related to emergence of social structures from individual agents.

In the Boella and van der Torre (to appear) model of normative multiagent systems, part of Castelfranchi's model is formalized in terms of rule based systems, based on deontic logic of van der Torre (2003) and Broersen et al. (2002, to appear)'s BOID architecture. These models are used to model a variety of phenomena, such as virtual communities, co-operation in groups, contracts, and constitutive norms.

The formal characteristic of the model is that it combines a logical framework with decision-theoretic and game-theoretic mechanisms, such that the behavior of the agents as well as the system can be formalized. For example, in the model of co-operation within groups the game-theoretic concepts are used to model the property of agents in a group are committed to mutual responsiveness, that is, they monitor the behavior of other agents and help them if possible to reach their goals.

The development of the model is driven by examples found in the social and legal literature. An example is Beccaria's argument that high penalties for minor offences increases the total set of norm violations, because once an agent has committed a violation he or she is no longer constrained to commit more violations. The model combines a logical framework to represent and reason about norms and norm violations, with a decision-theoretic mechanism to explain the behavior of the violator.

3 Examples

3.1 Coordination and cooperation

Shoham and Tennenholtz (1992) introduce artificial social systems to coordinate multiagent systems, using a kind of norms called *social laws*.

> In multiagent systems be they human societies or distributed computing systems different agents, people or processes, aim to achieve different goals and yet these agents must interact either directly by sharing information and services or indirectly by sharing system resources. In such distributed systems it is crucial that the agents agree on certain rules in order to decrease conflicts among them and promote cooperative behavior. Without such rules even the simplest goals might become unattainable by any of the agents or at least not efficiently attainable. Just imagine driving in the absence of traffic rules. These rules

> strike a balance between allowing agents sufficient freedom to achieve their goals and restricting them so that they do not interfere too much with one another.

They consider the possibility of limiting the agents to a subset of the original strategies of a given game thus inducing a subgame of the original one. They call such a restriction a social constraint if the restriction leaves only one strategy to each agent. Some social constraints are consistent with the principle of individual rationality in the sense that it is rational for agents to accept those assuming all others do as well.

A discussion in artificial social systems is whether social laws are hard or soft constraints. The distinction between hard and soft constraints corresponds to the distinction between preventative and detective control systems. In the former a system is built such that violations are impossible (you cannot enter metro station without a ticket) or that violations can be detected (you can enter train without a ticket but you may be checked and sanctioned).

3.2 Multiagent organizations

Organizations embody a powerful way to coordinate complex behavior in human society. Different models of organizations exist, from bureaucratic systems based on norms to competitive systems based on markets. Moreover, organizational concepts allow to structure the behavior of complex entities in a hierarchy of encapsulated entities: departments structured in roles, organizations structured in departments, and inter-organizational coordination structured in organizations. Organizations specify the interaction and communication possibilities of each of these entities, abstracting from the implementation of their behavior. Since these entities are autonomous, they can only be coordinated exogenously.

Organizational models have been popular in the last years in agent theory for modeling coordination in open systems, where departments and organizations are modeled as autonomous entities. This is also due to the need to ensure social order within MAS applications like Web Services, Grid Computing, and Ubiquitous Computing. In these settings, openness, heterogeneity, and scalability pose new challenges on traditional MAS organizational models. It becomes necessary to integrate organizational and individual perspectives and to promote the dynamic adaptation of models to organizational and environmental changes. Nowadays, practical applications of agents to organizational modeling are being widely developed.

Moreover, organizational concepts are used frequently for coordination purposes in different areas of Computer Science. For example, roles are used in access control, conceptual modeling, programming languages and patterns. Contracts are used in design by contract, and services are used in web services and service level agreements. Message based communication is used in networking. Finally, coordination techniques are used in formal models of organizations to analyze or simulate them. In contrast, most coordination languages refer mostly to different kinds of metaphors, like blackboards, shared data-spaces, component composition and channels.

Multiagent systems consist of a set of agents, which can be designed and implemented in a variety of ways. In particular, a system designer would like to control the emergent behavior of the system. In human societies, groups of humans are controlled by organizational structures, for example in business. However, it is more difficult to define the relations between agents, or properties of the whole system. Therefore multiagent organizations are defined, which describe the relations between agents.

References

C.E. Alchourròn and E. Bulygin. *Normative Systems*. Springer, 1971.

G. Boella and L. van der Torre. A game theoretic approach to contracts in multiagent systems. *IEEE Trans. SMC, Part C*, to appear.

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.

J. Broersen, M. Dastani, and L. van der Torre. Beliefs, obligations, intentison and desires as components in agent architectures. *International Journal of Intelligent Systems*, to appear.

C. Castelfranchi. Modelling Social Action for AI Agents. *Artificial Intelligence*, 103:157 – 182, 1998.

C. Castelfranchi. Founding an agent's autonomy on dependence theory. In *Proceedings of ECAI'00*, pages 353 – 357. Berlin, 2000.

R. Conte and C. Castelfranchi. *Cognitive and social action*. UCL Press London, 1995.

E. Goffman. *The Presentation of Self in Everyday Life*. Doubleday, 1959.

J. Habermas. *The Theory of Communicative Action, Volume One, Reason and the Rationalization of Society*. Beacon Press, Boston, 1984. transl McCarthy, orig publ as Theorie des Kommunikativen Handels, 1981.

A. Jones and J. Carmo. Deontic logic and contrary-to-duties. In D. Gabbay, editor, *Handbook of Philosophical Logic*, page 203279. Kluwer, 2001.

J-J. Meyer and R. Wieringa, editors. *Deontic logic in computer science: normative system specification*. Wiley, 1993.

J.-J.Ch. Meyer. A different approach to deontic logic: deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1988.

S. Ossowski. *Co-ordination in Artificial Agent Societies*. Springer-Verlag, Berlin, 1999.

Y. Shoham and M. Tennenholtz. On the Synthesis of Useful Social Laws for Artificial Agent Societies (Preliminary Report). In *Proceedings of the National Conference on Artificial Intelligence*, pages 276–281, San Jose, CA, July 1992.

Göran Therborn. Back to Norms! On the Scope and Dynamics of Norms and Normative Action. *Current Sociology*, 50(6):863 – 880, 2002.

L. van der Torre. Contextual deontic logic: Normative agents, violations and independence. *Annals of Mathematics and Artificial Intelligence*, 37 (1-2):33–63, 2003.

H. Verhagen. *Norm Autonomous Agents*. PhD thesis, Department of System and Computer Sciences, The Royal Institute of Technology and Stockholm University, Sweden, 2000.