A Game-Theoretic Approach to Normative Multi-Agent Systems

Guido Boella[1] and Leendert van der Torre[2]

[1] Università di Torino, Dipartimento di Informatica
10149, Torino, Cso Svizzera 185, Italia
guido@di.unito.it
[2] University of Luxembourg, Computer Science and Communications (CSC)
1359, Luxembourg, 6 rue Richard Coudenhove Kalergi, Luxembourg
leendert@vandertorre.com

Abstract. We explain the *raison d'être* and basic ideas of our game-theoretic approach to normative multiagent systems, sketching the central elements with pointers to other publications for detailed developments.

Keywords. Normative multiagent systems, deontic logic, input/output logic

Introduction

We explain the *raison d'être* and basic ideas of our game-theoretic approach to normative multi-agent systems, sketching the central elements with pointers to other publications for detailed developments. In particular, we address the following questions:

Motivation. Why do we need a game-theoretic approach to normative multi-agent systems?
Objectives. What do we want to achieve with the theory of normative multi-agent systems?
Methodology. How do we achieve the objectives?
Results. Which results have been obtained thus far?
Interdisciplinarity. How are various disciplines used in the theory?

We aim to explain our own approach, and we are therefore very brief with respect to recent related approaches in the area of normative multiagent systems. For these other approaches, see the special issue on normative multiagent systems in *Computational and Mathematical Organization Theory* [68], these DROPS proceedings, the proceedings of the biannual workshops on deontic logic in computer science (ΔEON) and of the COIN workshop series.[1]

The layout of this paper follows the five questions above, addressing each of them in a new section.

[1] http://www.ia.urjc.es/COIN2007/

Dagstuhl Seminar Proceedings 07122
Normative Multi-agent Systems
http://drops.dagstuhl.de/opus/volltexte/2007/937

1 Motivation for a new approach to normative systems

In Section 1.1 we explain why we need a theory of norms by arguing that norms are a special class of constraints deserving special analysis. In Section 1.2 we define what we mean by a norm, distinguishing among regulative, constitutive and procedural ones, and in Section 1.3 we explain why a normative system in multi-agent systems is seen as a mechanism, in particular to obtain desirable agent behavior or to structure organizations. Finally we explain in Section 1.4 what we mean by game-theoretic scenarios in normative multi-agent systems, and in Section 1.5 we discuss an important advantage of our game-theoretic approach, which we call the game-theoretic analysis of normative multi-agent systems.

1.1 Norms are a class of constraints deserving special analysis

Meyer and Wieringa 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" [100, preface]. Alchourròn and Bulygin [2] define a normative system inspired by Tarskian deductive systems:

> "When a deductive correlation is such that the first sentence of the ordered pair is a case and the second is a solution, it will be called normative. If among the deductive correlations of the set α there is at least one normative correlation, we shall say that the set α has normative consequences. A system of sentences which has some normative consequences will be called a normative system." [2, p.55].

Jones and Carmo [89] introduce agents in the definition of a normative system by defining it as "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." Since the agents' control over the norms is not explicit here, we use the following definition.

> A normative multi-agent system is a multi-agent system together with normative systems in which agents can decide whether to follow the explicitly represented norms or not, and the normative systems specify how and in which extent the agents can modify the norms. [68]

Note that this definition makes no presumptions about the internal architecture of an agent or of the way norms find their expression in agent's behavior.

Representation of norms Since norms are explicitly represented, according to our definition of a normative multi-agent system, the question should be raised how norms are represented. Norms can be interpreted as a special kind of constraint, and represented depending on the domain in which they occur.

However, the representation of norms by domain dependent constraints runs into the question what happens when norms are violated. Not all agents behave according to the norm, and the system has to deal with it. In other words, norms are not hard constraints, but soft constraints. For example, the system may sanction violations or reward good behavior. Thus, the normative system has to monitor the behavior of agents and enforce the sanctions. Also, when norms are represented as domain dependent constraints, the question will be raised how to represent permissive norms, and how they relate to obligations. Whereas obligations and prohibitions can be represented as constraints, this does not seem to hold for permissions. For example, how to represent the permission to access a resource under an access control system? Finally, when norms are represented as domain dependent constraints, the question can be raised how norms evolve.

We therefore believe that norms should be represented as a domain independent theory. For example, deontic logic [94,95,96,109,110,117] studies logical relations among obligations and permissions, and more in particular violations and contrary-to-duty obligations, permissions and their relation to obligations, and the dynamics of obligations over time. Therefore, insights from deontic logic can be used to represent and reason with norms in multi-agent systems. Deontic logic also offers representations of norms as rules or conditionals. However, there are several aspects of norms which are not covered by constraints nor by deontic logic, such as the relation among the cognitive abilities of agents and the global properties of norms. Meyer and Wieringa 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" [100, preface].

Norms and agents Conte *et al.* [76] distinguish two distinct sets of problems in normative multi-agent systems research. On the one hand, they claim that legal theory and deontic logic supply a theory of norm-governed interaction of autonomous agents while at the same time lacking a model that integrates the different social and normative concepts of this theory. On the other hand, they claim that three other problems are of interest in multi-agent systems research on norms: how agents can acquire norms, how agents can violate norms, and how an agent can be autonomous. Agent decision making in normative systems and the relation between desires and obligations has been studied in agent architectures [72], which thus explain how norms and obligations influence agent behavior.

An important question in normative multi-agent systems is where norms come from. Norms are not necessarily created by legislators, but they can also be negotiated among agents, or they can emerge spontaneously, making the agents norm autonomous [112]. In electronic commerce research, for example, cognitive foundations of social norms and contracts are studied [58]. Protocols and social mechanisms are now being developed to support such creations of norms in multi-agent systems. Moreover, agents like legislators playing a role in the normative system have to be regulated themselves by procedural norms [67], raising the question how these new kind of norms are related to the other kinds of norms.

When norms are created, the question can be raised how they are enforced. For example, when a contract is violated, the violator may have to pay a penalty. But then there has to be a monitoring and sanctioning system, for example police agents in an electronic institution. Such protocols or roles in a multi-agent system are part of the construction of social reality, and Searle [105] has argued that such social realities are constructed by constitutive norms. This raises the question how to represent such constitutive or counts-as norms, and how they are related to regulative norms like obligations and permissions [62].

Norms and other concepts Not only the relation between norms and agents must be studied, but also the relation between norms and other social and legal concepts. How do norms structure organizations? How do norms coordinate groups and societies? How about the contract frames in which contracts live? How about the relation between legal courts? Though in some normative multi-agent systems there is only a single normative system, there can also be several of them, raising the question how normative systems interact. For example, in a virtual community of resource providers each provider may have its own normative system, which raises the question how one system can authorize access in another system, or how global policies can be defined to regulate these local policies [62].

1.2 Kinds of norms

Normative multiagent systems as a research area can be defined as the intersection of normative systems and multi-agent systems [68]. With 'normative' we mean 'conforming to or based on norms', as in *normative behavior* or *normative judgments*. According to the Merriam-Webster Online Dictionary [99], other meanings of normative not considered here are 'of, relating to, or determining norms or standards', as in *normative tests*, or 'prescribing norms', as in *normative rules of ethics* or *normative grammar*. With 'norm' we mean 'a principle of right action binding upon the members of a group and serving to guide, control, or regulate proper and acceptable behavior'. Other meanings of 'norm' given by the Merriam-Webster Online Dictionary but not considered here are 'an authoritative standard or model', 'an average like a standard, typical pattern, widespread practice or rule in a group', and various definitions used in mathematics. Kinds of norms which are usually distinguished are regulative norms

like obligations, permissions and prohibitions, constitutive norms like counts-as conditionals, and more, as discussed below.

Regulative norms: obligations, permissions, prohibitions Regulative norms specify the ideal and varying degrees of sub-ideal behavior of a system by means of obligations, prohibitions and permissions. Deontic logic [6,118] considers logical relations among obligations and permissions and focuses on the description of the ideal or optimal situation to achieve, driven by representation problems expressed by the so-called deontic paradoxes, most notoriously the contrary-to-duty paradoxes, see, for example, [89,110].

Constitutive norms: counts-as conditionals Constitutive norms are based on the notion that "X counts-as Y in context C" and are used to support regulative norms by introducing institutional facts in the representation of legal reality.

The notion of counts-as introduced by Searle [105] has been interpreted in deontic logic in different ways and it seems to refer to different albeit related phenomena [85]. For example, Jones and Sergot [90] consider counts-as from the constitutive point of view. According to Jones and Sergot, the fact that A counts-as B in context C is read as a statement to the effect that A represents conditions for guaranteeing the applicability of particular classificatory categories. The counts-as guarantees the soundness of that inference, and enables "new" classifications which would otherwise not hold.

An alternative view of the counts-as relation is proposed by Grossi *et al.* [84]: according to the classificatory perspective A counts-as B in context C is interpreted as: A is classified as B in context C. In other words, the occurrence of A is a sufficient condition, in context C, for the occurrence of B. Via counts-as statements, normative systems can establish the ontology they use in order to distribute obligations, rights, prohibitions, permissions, *etc.* See [54] for a discussion on the relation between count-as conditionals, classification and context.

In [42,52,58] we propose a different view of counts-as which focuses on the fact that counts-as often provides an abstraction mechanism in terms of institutional facts, allowing the regulative rules to refer to legal notions which abstract from details. Counts-as conditionals can be used to define other concepts, such as role-based Rights in Artificial Social Systems [50]. In [51,60] we study the relation between obligations, permissions and constitutive norms using a logical architecture.

Procedural norms The distinction between substantive and procedural norms is well known in legal theory [98]. Substantive norms define the legal relationships of people with other people and the state in terms of regulative and constitutive norms, where regulative norms are obligations, prohibitions and permissions, and constitutive norms state what counts as institutional facts in a normative system. Procedural norms are instrumental norms, addressed to agents playing

roles in the normative system, which aim at achieving the social order specified in terms of substantive norms. Procedural law encompasses legal rules governing the process for settlement of disputes (criminal and civil). Procedural and substantive law are complementary. Procedural law brings substantive law to life and enables rights and duties to be enforced and defended. For example, procedural norms explain how a trial should be carried out and which are the duties, rights and powers of judges, lawyers and defendants.

The role that agents have in enforcing the social order the normative system aims to by creating norms has been recognized in normative multi-agent systems [58,62], and agents are considered which are in charge of sanctioning violations on behalf of the normative system [33,39]. Moreover, obligations are associated with procedural norms which are instrumental - to use Hart [86]'s terminology - to distribute the tasks to agents like judges and policemen, who have to decide whether and how to fulfill them.

In [55,67] we introduce a logical framework for substantive and procedural norms, and we use it to study the relation between these two kinds of norms and to answer the following three questions. First, how are regulative and constitutive norms related in a normative system with substantive and procedural norms? Second, by which mechanism are procedural norms created to motivate agents to recognize violations, apply sanctions, or to recognize institutional facts? Third, how can the formal framework be used to model various applications of normative multi-agent systems, where only some of them may need procedural norms?

1.3 Normative system as a mechanism

In this section we discuss why there are norms in social systems like multi-agent systems. We have distinguished various kinds of norms, such as obligations or counts-as conditionals, but this does not explain their existence. We assume that a norm is a mechanism to obtain desired multi-agent system behavior. In other words, it is an incentive, which brings us directly into the study of incentives, called economics.

Norms as a mechanism to obtain desirable agent behavior Norms have for long been considered as one of the possible incentives to motivate agents. Consider the economist Levitt [92, p.18-20], discussing an example of Gneezy and Rustichini [81].

> Imagine for a moment that you are the manager of a day-care center. You have a clearly stated policy that children are supposed to be picked up by 4 p.m. But very often parents are late. The result: at day's end, you have some anxious children and at least a teacher must wait around for the parents to arrive. What to do?
> A pair of economists who heard of this dilemma – it turned out to be a rather common one – offered a solution: fine the tardy parents. Why, after all, should the day-care center take care of these kids for free?

> The economists decided to test their solution by conducting a study of ten day-care centers in Haifa, Israel. The study lasted twenty weeks, but the fine was not introduced immediately. For the first four weeks, the economists simply kept track of the number of participants who came late; there were, on average, eight pickups per week per day-center. In the fifth week, the fine was enacted. It was announced that any parent arriving more than ten minutes late would pay $3 per child for each incident. The fee would be added to the parents' monthly bill, which was roughly $380.
> After the fine was enacted, the number of late pickups promptly went ... up. Before long there were twenty late pickups per week, more than double the original average. The incentive had plainly backfired.
> Economics is, at root, the study of incentives: how people get what they want, or need, especially when other people want or need the same thing. Economists love incentives. They love to dream them up and enact them, study them and tinker with them. The typical economist believes the world has not yet invented a problem that he cannot fix if given a free hand to design the proper incentive scheme. His solution may not always be pretty–but the original problem, rest assured, will be fixed. An incentive is a bullet, a lever, a key: an often tiny object with astonishing power to change a situation.
> ...
> There are three basic flavors of incentive: economic, social, and moral. Very often a single incentive scheme will include all three varieties. Think about the anti-smoking campaign of recent years. The addition of $3-per-pack "sin tax" is a strong economic incentive against buying cigarettes. The banning of cigarettes in restaurants and bars is a powerful social incentive. And when the U.S. government asserts that terrorists raise money by selling black-market cigarettes, that acts as a rather jarring moral incentive.

The daycare example illustrates that norms can be used as a mechanism to obtain desirable behavior of a multiagent system, because it is used as one of the incentives. It suggests also that the main tools to study incentives in economics, classical decision and game theory, may be useful tools to study the role of normative incentives too. Note that the daycare example illustrates also that economic theory is concerned with normative reasoning too, and that an analysis of incentives should not naively restrict itself to economic incentives, because it should also take norms into account.

The fact that norms can be used as an mechanism to obtain desirable system behavior, i.e. that norms can be used as incentives for agents, implies that in some circumstances economic incentives are not sufficient to obtain such behavior. For example, in a widely discussed example of the so-called centipede game, there is a pile of thousand pennies, and two agents can in turn either take one or two pennies. If an agent takes one then the other agent takes turn, if it takes two then the game ends. A backward induction argument implies that it is rational only

to take two at the first turn. Norms and trust have been discussed to analyze this behavior, see [87] for a discussion.

Norms as a mechanism to organize systems To manage properly complex systems like multiagent systems, it is necessary that they have a modular design. While in traditional software systems, modularity is addressed via the notions of class and object, in multiagent systems the notion of organization is borrowed from the ontology of social systems. Organizing a multiagent system allows to decompose it and defining different levels of abstraction when designing it.

According to Zambonelli *et al.* [119] "a multiagent system can be conceived in terms of an organized society of individuals in which each agent plays specific roles and interacts with other agents". At the same time, they claim that "an organization is more than simply a collection of roles (as most methodologies assume) [...] further organization-oriented abstractions need to be devised and placed in the context of a methodology [...] As soon as the complexity increases, modularity and encapsulation principles suggest dividing the system into different suborganizations". According to Jennings [88], however, most current approaches "possess insufficient mechanisms for dealing with organisational structure". Moreover, what is the semantic principle which allows decomposing organizations into suborganizations must be still made precise. Organizations are modelled as collections of agents, gathered in groups [78], playing roles [88,97] or regulated by organizational rules [119].

Norms are another answer to the question of how to model organizations as first class citizens in multiagent systems. Norms are not usually addressed to individual agents, but rather they are addressed to roles played by agents [65]. In this way, norms from a mechanism to obtain the behavior of agents, also become a mechanism to create the organizational structure of multiagent systems. The aim of an organizational structure is to coordinate the behavior of agents so to perform complex tasks which cannot be done by individual agents. In organizing a system all types of norms are necessary, in particular, constitutive norms, which are used to assign powers to agents playing roles inside the organization. Such powers allow to give commands to other agents, make formal communications and to restructure the organization itself, for example, by managing the assignment of agents to roles.

Moreover, normative systems allow to model also the structure of an organization and not only the interdependences among the agents of an organization. Consider a simple example from organizational theory in Economics: an enterprise which is composed by a direction area and a production area. The direction area is composed by the CEO and the board. The board is composed by a set of administrators. The production area is composed by two production units; each production unit by a set of workers. The direction area, the board, the production area and the production units are *functional areas*. In particular, the direction area and the production areas belong to the organization, the board to the direction area, *etc.* The CEO, the administrators and the members of

the production units are *roles*, each one belonging to a functional area, e.g., the CEO is part of the direction area.

This recursive decomposition terminates with roles: roles, unlike organizations and functional areas, are not composed by further social entities. Rather, roles are played by other agents, real agents (human or software) who have to act as expected by their role.

Each of these elements can be seen as an institution in a normative system, where legal institutions are defined by Ruiter [104] as "systems of [regulative and constitutive] rules that provide frameworks for social action within larger rule-governed settings". They are "relatively independent *institutional legal orders* within the comprehensive legal orders".

1.4 Game-theoretic scenarios of normative multiagent systems

In this section we explain our game-theoretic foundations for norms [59].

Interactions among agents in a normative multiagent system Many examples are given in the literature on the interaction among agents using regulative norms. For example, consider the following simple scenario due to Ron Lee of the access to a photo copier [91]:

1. A tells B to permit C to do use photocopier
2. B permits C to do use photocopier
3. C cannot do use copier, since door is closed
4. A complains to B about C not able to use it
5. B tells A that he permitted C, as requested

The standard analysis of this example includes the idea that C has the right to use the photocopies in the sense that he is entitled to use it, and B is therefore obliged to open the door. Moreover, as the photocopier has an access code, then B has to tell the code to C (even though C would be able to use the copier if he where to guess the code, which is the point where knowledge gets into this scenario). There are many variants of this example due to Sergot and colleagues, such as borrowing books in the library regulations formalized, parking cars in the parking lot in the parking regulations, and so on. Typically, many more agents are involved in these real world examples than just A, B and C. In multiagent systems, similar scenarios can be found in access control systems, for example to access a web service.

A similar example is often discussed where permission is replaced by obligation (where A like to transmit its will to influence the behavior of C via B):

1. A tells B to oblige C to do copies of a paper.
2. B obliges C to do copies of a paper
3. C does not do copies of a paper
4. A complains to B about C not doing copies of a paper
5. B tells A that he obliged C, as requested

Contracts are based on norms and occur in strategic interaction scenarios found in e-commerce, as studied by Tan and colleagues [83]. In an escrow service or a bill of lading typically several buyers, sellers, transporters, financial institutions and other agents are involved, which regulate their interactions via complex contracts. Here norms are used to give agents the power to achieve things.

When more agents are involved, their social interactions may give rise to the emergence of norms. For example, in case there is no trust there is no deal due to lack of equilibrium. If there is a joint goal based on agent desires then the agents can propose or negotiate norms leading to a new equilibrium, in which they accept the norms.

In most human social systems it takes a long time to emerge, but in computer systems they can be created much more quickly. For example, consider a peer to peer ad hoc network used for incident management. In case of an incident such as fire in a tunnel, cars, police, firemen, hospitals and so on have to be coordinated.

Another source for social interaction scenarios can be found in the popular reality games such as second life, where we expect many applications of normative multiagent systems.

Complexity and abstraction All these examples of interaction scenarios show highly complex and dynamic systems, and the question is how we can model the examples - whether it is to develop multi-agent systems in agent based software engineering, or to analyze the multi-agent system in agent theory. There are two main approaches to reduce the complexity.

First, the usual approach to reduce the complexity is to describe agents using a simple and uniform formalization. For example, classical game theory describes all agents by utility function and probability distribution, together with the decision rule to maximize expected utility, and alternative agent models developed in artificial intelligence and cognitive science are based on models such as belief-intention-desire (BDI) model.

Second, an alternative approach to reduce the complexity is to restrict the number of agents which are considered in the interaction. Here the multi-agent structure of normative systems can be used. For example, legal systems are based on the Trias Politica [39]. From the perspective of a normative system, there are two kinds of agents. First, the agents who are subject to the norms, and the agents who play a role in the system to make it function. This generates to the distinction between substantive norms used to regulate agents subject to the system, and procedural norms used to regulate agents playing a role in the system.

Normative system as a level of abstraction One way we can simplify interaction in a normative multiagent system is to abstract away the agents playing a role in the normative system, and keeping the normative system as an entity

interacting with the agents subject to it. The agents playing a role in the system empower the normative system, and the normative system delegates again some of its powers to these agents; this is known as *mutual empowerment*. This level of abstraction is most clear in organizational theory, where an organization can be seen as a normative multiagent system as well as a legal entity. One can use the agent metaphor for such abstracted normative systems, for example by attributing mental attitudes to normative systems [32].

The sociologist Goffman sees norms as producing a form of strategic interaction between the agent and the normative system. In a normative system, the "enforcement power is taken from mother nature and invested in a social office specialized for this purpose, namely a body of officials empowered to make final judgements and to institute payments" [82, p.115]. Such a game is unusual since "the judges and their actions will not be fully fixed in the environment, many unnatural things are possible. [...] the payment for a player's move ceases to be automatic but is decided on and made by the judges where everything is over" [82, p.115]. "Strategic interaction" here means the, according to Goffman unavoidable, taking into consideration of the other agents' actions.

> "When an agent considers which course of action to follow, before he takes a decision, he depicts in his mind the consequences of his action for the other involved agents, their likely reaction, and the influence of this reaction on his own welfare" [82, p. 12].

At this level of abstraction, the simplest game which can be played is an agent deliberating about an action, and the normative system reacting to it. Since the goal of the agent is typically to violate the norms without being sanctioned, we call this kind of interaction a *violation game*, and we represent it by A:N. Other kinds of interactions at this abstraction level are extensions of violation games. consider for example a (legislator in a) normative system deliberating which norm to create. He can introduce a norm, then an agent will play a violation game. The goal of the normative system is that the agent is motivated such that the norm is not violated. We call it a norm creation game, and we abstractly represent it by N:A:N. Also more complex interactions among agents can be modelled in this way, for example involving control hierarchies such as defender agents, various kinds of authorities like norm source hierarchies, and so on.

1.5 The game-theoretic analysis of norms

Norms should satisfy various properties to be effective as a mechanism to obtain desirable behavior. For example, the system should not sanction without reason, as for example Caligula or Nero did in the ancient Roman times, otherwise the norms would loose their force to motivate agents. Moreover, sanctions should not be too low, as in the daycare example, but they also should not be too high, as shown by the argument of Beccaria [14] on death penalty. Otherwise, once a norm is violated, there is no way to prevent further norm violations. In [59] we list the following requirements for such an analysis.

The first requirement is that norms influence the behavior of agents. However, they only have to do so under normal or typical circumstances. For example, if other agents are not obeying the norm, then we cannot expect an agent to do so. This norm acceptance has been studied by [76], and in a game-theoretic setting for social laws by [108].

The second requirement is that even if a norm is accepted in the sense that the other agents obey the norm, an agent should be able to violate the norms. A normative multi-agent system is a "set of agents [...] whose interactions can be regarded as 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" [89]. In other words, the norms of global policies must be represented as soft constraints, which are used in detective control systems where violations can be detected, instead of hard constraints restricted to preventative control systems in which violations are impossible. The typical example of the former is that you can enter a train without a ticket, but you may be checked and sanctioned, and an example of the latter is that you cannot enter a metro station without a ticket. Moreover, detective control is the result of actions of agents and therefore subject to errors and influenceable by actions of other agents. Therefore, it may be the case that violations are not often enough detected, that law enforcement is lazy or can be bribed, there are conflicting obligations in the normative system, that agents are able to block the sanction, block the prosecution, update the normative system, etc. A game-theoretic analysis can be used to study these issues of fraud and deception.

The third requirement is that norms should apply to a variety of agent types, since agents can be motivated in various ways, as the daycare example illustrates. We assume that a norm is a mechanism to obtain desired multi-agent system behavior, and must therefore under normal or typical circumstances be fulfilled for a range of agent types. Castelfranchi argues that sanctions are only one of the means which motivate agents to respect obligations, besides "pro-active actions, prevention from deviation and reinforcement of correct behavior, and then also 'positive sanctions', social approval" [74]. Castelfranchi [74] argues that an agent should fulfill an obligation because it is an obligation, not because there is a sanction associated with it.

> "True norms are aimed in fact at the internal control by the addressee itself as a cognitive deliberative agent, able to understand a norm as such and adopt it. [...] The use of external control and sanction is only a sub-ideal situation and obligation." [74]

We therefore use the distinction between violations and sanctions to distinguish between the agent's interpretation of the obligation, and its personal characteristics or agent type. The agent types are inspired by the use of agent types in the goal generation components of Broersen et al.'s BOID architecture [73]. *Roughly*, we distinguish among *norm internalizing agents*, *respectful* agents that attempt to evade norm violations and that are motivated by what counts and

does not count as a violation, and *selfish* agents that obey norms only due to the associated sanctions, i.e. that are motivated by sanctions only. An obligation without a sanction *should* be fulfilled, as Castelfranchi argues. But if fulfilling the obligations has a cost then it *is* only fulfilled by respectful agents, not by selfish agents, unless some incentives are provided or the agents dislike some social consequences of the violations. A respectful agent fulfills its obligations due to the existence of the obligation, whereas a selfish agent fulfills its obligations due to fear of consequences.

Respectful agents: agents that base their decisions solely on whether their behavior respects the goals of the normative agents. They put their duties before their own goals and desires: they maximize the fulfilment of obligations regardless to what happens to its own goals; even if the agent **n** did not sanction them, the agent **a** would prefer to respect the obligation. We say that respectful agents *adopt* the goal of the normative agent as their preference.

Selfish agents: agents that base their decisions solely on the consequences of their actions. If the obligation is respected, it is because agent **a** predicts that the situation resulting from the fulfillment is preferred according to its own goals and desires only: e.g., if it does not share its files, it knows that it can be sanctioned, a situation it does not desire or want. But it is possible also that there are not only material reasons, that is, not only for the damage caused by the sanction. Nothing prevents that the content of the norm is already a goal of the agent; moreover, agent **a** could have the desire not to be considered a violator, or it knows that being considered a violator gives it a bad reputation, so that it would not be trusted by other agent. However, to stick to the obligation, the goal of not being a violator or of not being sanctioned must be preferred by the agent to the desire or goal not to respect the obligation (obligations usually have a cost): a weak sanction, as it often happens, does not enforce the respect of a norm (e.g., the sanction is that the access to a website is forbidden, but the agent has already downloaded what it wanted).

To distinguish these cases, we distinguish between the decision to count behavior as a violation, and to sanction it.

Of course, most agents are mixed types of agents between these two extremes. Sometimes an agent is respectful and in other cases it is selfish. Balancing these two extremes is an important part of the agent's deliberation. The adoption of the obligation as a goal can be considered as an additional factor when the different alternatives are weighed according to its own goals and desires: so that the newly added motivations can affect the decision of agent **a** and move it towards an obligation-abiding behavior besides its own attitude towards that goal and the possible consequences of its alternative decisions. However, if a norm is effective in each case of the agent types, it is also effective for mixed agents. therefore we can restrict ourselves to the extreme agent types in a game-theoretical analysis.

Given possible conditions for a norm, the fourth requirement is that norms are as weak as possible, in the sense that the norms should not apply in cases where this is undesired, and that sanctions should not be too severe. The latter is motivated by a classical economic argument due to Beccaria, which says that if sanctions are too high, they can no longer be used in cases where agents already have violated a norm. Sanctions should be high enough to motivate selfish agents, but they should not be too high.

Designing norms satisfying these requirements is an area of game theory called mechanism design. In, amongst others, [59] we provide the following informal definition of obligation, extending Boella and Lesmo [26]'s proposal. According to legal studies what distinguishes norms from mere power to damage an agent is that sanctions are possible only in case of violations and which situations can be considered as violations is defined by the law: "*nullum crimen, nulla poena sine lege*". In our definition a norm specifies what will be considered as a violation by the normative agent (item 2) and that the normative agent will sanction only in case of violations (item 3). In this paper, we consider the set of norms as given. Given a set of norms N, agent **a** is obliged by the normative agent **n** to do x with sanction s, iff there is a norm n of N such that:

- The content x of the obligation is a desire and goal of **n** and agent **n** wants that agent **a** adopts this as its decision since it considers agent **a** as responsible for x.
- agent **n** has the desire and the goal that, if the obligation is not respected by agent **a**, a prosecution process is started to determine if the situation "counts as" a violation of the obligation and that, if a violation is recognized, agent **a** is sanctioned.
- Both agent **a** and agent **n** do not desire the sanction: for agent **a** the sanction is an incentive to respect the obligation, while agent **n** has no immediate advantage from sanctioning.

This definition is extended in various papers in a number of ways. For example, goals and desires are formalized as conditional rules, because norms and obligations are typically represented by conditional rules.

2 Objectives

In this section we discuss the four objectives of our work on game-theoretical approach to normative multiagent systems.

2.1 A representation of a normative multiagent systems that combines the three existing representations of normative multiagent systems

There are many approaches to conceptualizing and developing normative multiagent systems. The most popular class of approaches starts from logical relations

among obligations (and sometimes permissions) in a deontic logic, and then extends the formalism with agent concepts like actions and time. Since the norms in normative multiagent systems are typically represented explicitly, we may say that these approaches start from the representation of norms. We call it the deontic logic approach. Drawback of this approach is that there is no guideline to tell how norms affect behavior.

The second class of approaches in normative multiagent systems starts from the use of norms in agent decision making and interaction. In other words, given a set of norms, how do the agents behave? We call it the normative agent architecture approach. This more dynamic approach, focussing on behavior rather than normative system structure, has the drawback that it does not give a guideline how the normative system changes over time due to behavior of agents.

The third class of approaches in normative multiagent systems takes the strategic interaction among agents and (representatives of) the normative system as a starting point. We call it the game-theoretic approach. An example is the distinction between controllable and uncontrollable agents in the Tennenholtz and Brafman's model. A drawback of their quantitative model is that it is not easily combined with the other two approaches, such that explicit representation of norms and normative decision making remains a problem.

Our first objective is to build a game theoretic model of normative multiagent systems that extends both the deontic logic and the normative decision making approach. We therefore refer to Goffman's notion of strategic interaction rather than the dominant economic equilibrium analysis. Particular challenges are:

KR-ASS Combining the qualitative formalisms used in knowledge representation and the quantitative ones used in artificial social systems,

AA-ASS Combining the micro representation of agent architectures with the macro representations in social theories (the micro-macro dichotomy)

KR-AA The use of logic in knowledge representation and the use of architecture in agent theory.

Typically there are several ways to align two theories in a common framework. For example, for the latter problem, we may develop logical representations of agents [73], or we may develop an architecture for a normative system [60].

2.2 A logical framework for qualitative risk analysis, building on ideas in security and risk management

Risk analysis goes beyond traditional security by not only stating what is forbidden, but also accepting that things go wrong sometimes, and how to deal with them. Therefore, risk analysis not only needs constraints like security, but also contrary-to-duty or more generally normative reasoning.

Traditional risk analysis is quantitative and based on statistics. However, it does not take the organizational structure, legal consequences and so on into account. If we model a system, we may also take a more qualitative approach. What we have to do to analyze risk is to build normative systems and agent

models, and combine them. This is precisely what we do in our normative multiagent systems. Thus, we replace classical statistical risk analysis by our game theoretic analysis.

This thus explains why contrary-to-duty reasoning is essential for risk management, but there is much more to risk management than contrary-to-duty reasoning. In particular, given that agents can violate norms and can be sanctioned, will agents violate the norms? Moreover, in such cases, will they be sanctioned?

2.3 A classification of situations in which one needs which elements of normative systems as a social mechanism

In Section 1.2 we discussed various kinds of norms such as obligations, permissions, counts-as conditionals, and procedural norms. Moreover, there are social laws, conventions, rights, entitlements, legal institutions, and many more related concepts. Each of these concepts may be considered as another kind of mechanism. At this moment there is no consensus when we need these mechanisms, and when we can use a simpler normative multiagent systems. Thus far only two kinds of arguments: obligations are needed when there is contrary-to-duty reasoning [89], and permissions are needed for multiple authorities [2].

2.4 Examples and classification which kind of normative multiagent systems are to be used for which kind of applications in computer science

Many new technologies use essentially the same kind of normative multiagent system as older ones, but each application domain tends to reinvent its own normative system, typically in a very naive way. For example, consider the use of rollback and compensation in web technology, which may be seen as preventative and detective control systems. We therefore aim to illustrate each kind of normative multiagent system by a typical example, such as fraud and deception, electronic commerce, secure knowledge management, and so on.

2.5 Scope of the objectives

We consider the organizational structures with its explicit roles as an orthogonal issue to the game-theoretic approach to normative multi-agent systems. In other words, we can study normative multi-agent systems without making these aspects explicit. We discuss some of our results in this area in Section 4.6.

We do not discuss the design or implementation of normative multiagent systems, though we believe that our programming language powerJava may be used to implement normative multiagent systems.

Though norms are used to control the emergent behavior in MAS, and evolutionary game theory is a useful tool to study this emergence, we do not address this issue.

3 Methodology

In our approach we use input/output logic for deontic logic, BOID architecture for the agent architecture, and both game theory and recursive modeling for artificial social systems.

3.1 Input/output logic

Input/output logic takes its origin in the study of conditional norms. These may express desired features of a situation, obligations under some legal, moral or practical code, goals, contingency plans, advice, etc. Typically they may be expressed in terms like: *In such-and-such a situation, so-and-so should be the case*, or *... should be brought about*, or *... should be worked towards*, or *... should be followed* – these locutions corresponding roughly to the kinds of norm mentioned.

3.2 BOID Architecture

The BOID architecture [73] is an extension of the BDI architecture with obligations (O). Each mental attitude is represented by a component in the architecture, whose behavior is described by input/output logic. Moreover, qualitative decision theories have been developed which extend this rule based formalism with the decision rule to achieve goals and to evade goal violations.

In the BOID architecture there is no set of norms and norm descriptions, but instead the agent description is adapted such that obligations (O) are added to the mental states of agents. This can be interpreted as a kind of internalization of the normative system by the agents, or as an abstraction which abstracts away the normative system. This is the dominant approach in deontic logic [100], in which typically one abstracts away from the norms to study logical relations between obligations (though for criticism and alternative approaches see [1,116]). Alternatively, in approaches based on the so-called Anderson reduction [4,5], which defines obligation of p as the necessity that the absence of p leads to a violation, $O(p) = \Box(\neg p \rightarrow V)$, obligations are defined in terms of violability and the state of the world. In the variant proposed by Meyer [101], who defines obligation of action α as 'the absence of α leads to a violation state, or $O(\alpha) = [\overline{\alpha}]V$, obligation is defined in terms of the agent's behavior.

3.3 Tennenholtz' classical game-theoretic approach to artificial social systems

We first consider the so-called partially controlled multi-agent system (PCMAS) approach of Brafman and Tennenholtz [70], one of the classical game-theoretical studies of social laws in so-called artificial social systems developed by Tennenholtz and colleagues, because incentives like sanctions and rewards play a central role in this theory. So-called controllable agents – agents controlled by the system programmer – enforce social behavior by punishing and rewarding agents,

and thus can be seen as representatives of the normative system. For example, consider an iterative prisoner dilemma. A controlled agent can be programmed such that it defects when it happens to encounter an agent which has defected in a previous round.

The PCMAS model thus distinguishes between two kinds of agent interaction in the game theory, namely between two normal (so-called uncontrollable) agents, and between a normal and a controllable agent. We show in this paper that this makes it a very useful model to give game-theoretic foundations to norms. Whereas classical game theory is only concerned with interaction among normal agents, it is the interaction among normal and controllable agents which we use in our game theoretic foundations.

The PCMAS approach not only clarifies the design of punishments, but it also illustrates the iterative and multi-agent character of social laws. However, there are also drawbacks of the model, such that it cannot be used to give a completely satisfactory game-theoretic foundation for norms. We would like to express that a norm can be used for various kinds of agents, such as norm internalizing agents, respectful agents that attempt to evade norm violations, and selfish agents that obey norms only due to the associated sanctions. Therefore, as classical game theory is too abstract to satisfactorily distinguish among agent types, we consider also cognitive agents and qualitative game theory.

Several game-theoretic studies on social laws have been made by Tennenholtz and colleagues, for example based on off-line design of social laws [106], the emergence of conventions [107], and the stability of social laws [108]. The approach of Brafman and Tennenholtz [70] distinguishes between controllable and uncontrollable agents, analogous to the distinction between controllable and uncontrollable events in discrete event systems.

Controllable agents are agents controlled by the system programmer to enforce social behavior by punishing and rewarding agents. The game-theoretic model is the most common model for representing emergent behavior in a population. A single game consists of the usual payoff matrix. For example, the prisoner's dilemma is a two person game where each agent can either cooperate or defect.

Definition 1. *A k-person game g is defined by a k-dimensional matrix M of size $n_1 \times \ldots \times n_k$, where n_m is the number of possible actions (or strategies) of the m'th agent. The entries of M are vectors of length k of real numbers, called pay-off vectors. A joint strategy in M is a tuple $(i_1, i_2, \ldots, i_k)$, where for each $i \leq j \leq k$, it is the case that $1 \leq i_j \leq n_j$.*

An iterative game consists of a sequence of single games.

Definition 2. *A n-k-g iterative game consists of a set of n agents and a given k person game g. The game is played repetitively an unbounded number of times. At each iteration, a random k-tuple of agents play an instance of the game, where the members of this k-tuple are selected with uniform distribution from the set of agents.*

Efficiency is a global criterion for judging the "goodness" of outcomes from the system's perspective, unlike single payoffs which describe a single agent's perspective.

Definition 3. *A joint strategy of a game g is called* efficient *if the sum of the players pay-offs is maximal.*

New in the Brafman-Tennenholtz model are the notions of punishment and reward w.r.t. some joint strategy s, measuring the gain (benefit) or loss (punishment) of an agent if we can somehow change the joint behavior of the agents from a chosen efficient solution s to s'.

Definition 4. *Let s be a fixed joint strategy for a given game g, with pay-off $p_i(s)$ for player i; in an instance of g in which a joint strategy s' was played, if $p_i(s) \geq p_i(s')$ we say that i's* punishment *w.r.t. s is $p_i(s) - p_i(s')$, and otherwise we say that its* benefit *w.r.t. s is $p_i(s') - p_i(s)$.*

Agents may need to be constrained to behave in a way that is locally suboptimal such that the multi-agent system is as efficient as possible. Brafman and Tennenholtz call such a constraint a social law. Then they informally define controlled agents:

> "Agents not conforming to the social law are referred to as *malicious agents*. In order to prevent the temptation to exploit the social law, we introduce a number of *punishing agents*, designed by the initial designer, that will play 'irrationally' if they detect behavior not conforming to the social law, attempting to minimize the payoff of malicious agents. The knowledge that future participants have of the punishment policy would deter deviations and eliminate the need for carrying it out. Hence, the punishing behavior is used as a threat aimed at deterring other agents from violating the social law. This threat is (part of) the control strategy adopted by the controllable agents in order to influence the behavior of the uncontrollable agents. Notice that this control strategy relies on the structural assumption that uncontrollable agents are expected utility maximizers."

They consider the design of punishments, and show, for example, necessary and sufficient conditions for the existence of a punishing strategy.

We believe that PCMAS can be used to give game-theoretic foundations to norms, though Brafman and Tennenholtz do not use or consider the terminology of normative systems or deontic logic. The model fulfills our two requirements by explaining several aspects of norms, such as the fact that they can be used iteratively, that sanctions are associated to it, and that they can be applied to various kinds of agents.

In particular, a useful property of the PCMAS model is that it uses the game-theoretic machinery to study not only interaction among normal agents, but also interaction among the controlled agents and the normal agents. Since

the controlled agents are representatives of the normative system, this means that the game-theoretic machinery is used to study the interaction among the normative system and the agents.

However, the emphasis on modeling uncontrollable agents as utility maximizers implies that they only obey the norm because they are afraid of the sanction. Thus the model does not fulfill the third requirement because it seems to exclude the possibility that an agent obeys the norm simply due to its existence. In social theory, for example, agents have been studied which internalize norms in the sense that they incorporate norms as their own goal, or respectful agents trying to obey the norms without internalizing them.

Maybe the game-theoretic machinery can be extended to take such social agents into account. For example, a norm internalizing agent may be defined as an uncontrollable agent which simply copies the utility function of a punishing agent, and a respectful agent which avoids sanctions even when the number of punishing agents is too low, for example by assuming the number of punishing agents is much higher than it is in reality. They may for example be ashamed to be caught while driving without a train ticket.

However, such a solution does not seem very satisfactory. For the norm internalizing agents, they not only obey the norm but they also start to act as policemen, which seems to go to far. Moreover, even when punishment is low or absent, a respectful agent may obey the norm (as in the daycare example). There seem to be several alternative ways to define respectful agents, but they seem to have their own drawbacks.

Moreover, there are also some more technical problems. For PCMAS to give game-theoretic foundations to norms, we first have to define the syntax of a norm. Typically norms are expressed as modal sentences expressing that p is obliged, $O(p)$ in a deontic logic, or p is permitted, $P(p)$. Since in the PCMAS setting we have actions or strategies only, we define $O_i(\alpha, p)$ for agent i is obliged to do action α, otherwise he is sanctioned with punishment p. Since a punishment p is defined as $p_i(s) - p_i(s')$, the first problem is how to define the chosen efficient solution s'. It is implicit in the condition that in the situation in which no norm is violated, no agent is punished (the Nero/Caligula example of the introduction).

Finally, whether an obligation $O_i(\alpha, p)$ holds in PCMAS or not cannot be seen from the game's definition, but only from the behavior of the controlled agents. In other words, it can only be derived from the design of punishments not explicit in the game theory.

3.4 Recursive modeling

Classical decision and game theory have been criticized for their assumptions of ideality. Several alternatives have been proposed that take the limited or bounded rationality of decision makers into account. For example, Newell [102] and others develop theories in artificial intelligence and agent theory replace probabilities and utilities by informational (knowledge, belief) and motivational attitudes (goal, desire), and the decision rule by a process of deliberation. Brat-

man [71] further extends such theories with intentions for sequential decisions and norms for multiagent decision making.

Alternatively, Gmytrasiewitcz and Durfee [80] replace the equilibria analysis in game theory by recursive modelling, which considers the practical limitations of agents in realistic settings such as acquiring knowledge and reasoning so that an agent can build only a finite nesting of models about other agents' decisions.

> "Recursive modelling method views a multi agent situation from the perspective of an agent that is individually trying to decide what physical and/or communicative actions it should take right now." [80]

Boella and Lesmo [26] therefore propose the following definition of a sanction-based obligation in terms of beliefs, goals and desires, inspired by Goffman's game-theoretic interpretation of obligations and by recursive modelling. Boella and Lesmo formalize this definition in a game-theoretic framework in which they recursively model the normative agent's behavior. The formalization is based on multi-attribute utility theory for taking into account the different aspects of the world for which agents have preferences. An important advantage of Boella and Lesmo's definition is that it does not introduce additional mental attitudes, but it defines obligations in terms of beliefs, desires and goals. Moreover, they distinguish various reasons why agents fulfil or violate obligations.

> "An obligation holds when there is an agent A, the *normative* agent, who has a goal that another (or more than one) agent B, the *bearer* agent, satisfies a goal G and who, in case he knows that the agent B has not adopted the goal G, can decide to perform an action Act which (negatively) affects some aspect of the world which (presumably) interests B. Both agents know these facts." [26, p.496]

The approach overcomes some of the limitations discussed in the previous section. First, obligations of the agents can be formalized as desires or goals of the normative agent. This representation may be paraphrased as "Your wish is my command", the title of this paper, because the desires or wishes of the normative agent are the obligations or commands of the other agents. The goals of the normative system describe the ideal behavior of the system.

Second, structural relations between agents playing roles in a normative system like legislators, who create norms, judges, who count behavior as violations and associate sanctions, and policemen who enforce sanctions, can be represented by the standard hierarchical structure of agents. For example, the normative agent a_{n+1} may contain the role of a legislator a_1, a judge a_2 and a policeman a_3. Such a relation between the normative agent and other agents is probably not reductive, as the normative system also contains properties which cannot be reduced to roles of normal agents.

Third, as illustrated by Boella and Lesmo by several examples, agents take the normative system into account by playing games with it. For example, an agent considers whether its actions will lead to a reaction of the normative system such as being sanctioned. In their model, the agent can evade sanctions by

for example ensuring that the normative agent does not observe their behavior, or by bribing the system. The advantage of the approach is that standard techniques developed in decision and game theory can be applied to normative reasoning. Moreover, a legislator can play a game with the normative agent and another agent to see whether a new norm it introduces will be complied with, and which kind of sanctions it has to associate with the norm to achieve the desired behavior.

The normative systems as agents perspective together with the attribution of mental attitudes to these normative systems is not only, as Tuomela [111] proposed, a powerful metaphor leading to useful techniques, but it can also be explained from a conceptual point of view in several ways. For example, according to Wooldridge and Jennings [115] the conditions for calling a system an agent are its autonomy to control its actions and internal state, its social ability to interact with other agents, its reactivity to the changes it observes in the environment, and its pro-activeness due to goal directed behavior and taking the initiative. All these conditions are met by normative systems. First, it is autonomous in counting behavior as violations and applying sanctions. Second, it interacts with these other agents by influencing their behavior, and the other agents interact with it by for example creating norms. Third, it is reactive since counting behavior as a violation of norms depends on its observations of such behavior. Fourth, it is proactive since this action is taken without any explicit triggering action of the agent, who would of course instead try to evade being violators and being sanctioned.

Moreover, attribution of mental attitudes to normative systems can be explained by the interpretation of normative multiagent systems as dynamic social orders. According to Castelfranchi [74], a social order is a pattern of interactions among interfering agents "such that it allows the satisfaction of the interests of some agent A". These interests can be a shared goal, a value that is good for everybody or for most of the members; for example, the interest may be to avoid accidents. The use of goals of a multiagent system can be explained by the notion of social delegation [77]. *Social delegation* describes the behavior of a social group or institution where some of the agents, on behalf of the other ones, have to achieve some goal which is part of the plans of all members of the group or institution. In this interpretation, the use of sanctions can be explained by the notion of *social control*, "an incessant local (micro) activity of its units" [74], aimed at restoring the regularities prescribed by norms, because a dynamic social order requires a continuous activity for ensuring that the normative system's goals are achieved. In case of sanction-based obligations, this ability is required since the application of sanctions in response to violations cannot be taken for granted.

For example, consider a peer-to-peer system like Napster. Each individual agent does not want that all other agents start to download files from their computer system. From the individual desires that other agents do not all use only their side, we get to a shared or group desire that downloads are evenly distributed over the network. In a normative system, this shared desire may

lead to a norm which says that one should not use computers which already are used intensively. The agents accept this norm, because they rationally see that this increases their own benefits. This example of social delegation is of course analogous to the development of moral principles based on the Kantian imperative: do not do to others what you could not accept yourself. Since, in some cases, for the agents it is rational to violate the norm, sanctions are added to enforce their respect. In our peer-to-peer example, those agents who do not share enough files get a reduced download speed rate.

As a more abstract example involving sanctions, consider the widely discussed prisoner's dilemma. In this example, both agents are better of if they cooperate, but rationality tells them to defect (since this is the only Nash equilibrium). Since there is a shared desire that both agents cooperate, in a normative system a norm will be created that counts defection as a violation that will be sanctioned. The agents will accept this norm and the related sanction, because rationality tells them that they will be better off. In this example, the sanction should be high enough to deter the agents from defecting, for which we can use the game described above.

4 Results

This section is structured along the complexity of the interaction between the normative system and the agents, as proposed in [23]. In the simplest setting there is only a single agent and a single normative system it is subject too. Then we consider the role of agents in a normative system (creating or enforcing norms, like legislators, policeman, judges, etc. in legal systems). We then also consider interaction among agents in contracting settings (where the norms of contract live in legal setting of contract frames or legal institutions) and, finally, we consider the most complex case of multiple normative systems (country law vs European law, virtual communities, etc.). Each subsection extends the theory developed in the previous subsection.

4.1 Violation games: interacting with normative systems, the obligation mechanism, with applications in trust, fraud and deception.

In [46] we start with considering the normative system from the (subjective) viewpoint of an agent. We also test our model in Tennenholtz artificial social systems as enforceable social laws [43].

Problems: Normative systems can be divided in preventative and detective ones. In the former, a system is built such that violations are impossible (you cannot enter metro station without a ticket), in the latter violations can be detected (you can enter train without a ticket but you may be checked and sanctioned). When we accept that norms can be violated, then we go beyond classical cryptographic security into the area of risk analysis, fraud, deception, trust and reputation, and so on. In all these areas, we have to find a way to

asses risk, anticipate behavior of other agents, and so on. Numbers are usually not available, so we develop a qualitative game theory for agents in normative systems.

Technique: Introduction of the notion of obligation and its relation with sanctions. Combination of normative reasoning and game theory. Analysis of the way an agent can violate obligations without being sanctioned by the normative system. Analysis of contrary-to-duty reasoning.

Case-studies. Transaction trust in normative multiagent systems [19], and energy markets [20].

4.2 Institutionalized games: counts-as mechanism, with applications in distributed systems, grid, p2p, virtual communities

We study permission and authorization in policies for virtual communities of agents [49], and local versus global policies and centralized versus decentralized control in virtual communities of agents [41].

4.3 Negotiation games: MAS interaction in a normative system, norm creation action mechanism, with applications in electronic commerce and contracting

In [58] we consider social phenomena in normative multiagent systems like directed and group obligations, how group obligations can be distributed and how norms can emerge. We consider also argument games for interactive access control [21], and negotiation of the distribution of obligations with sanctions among autonomous agents [38,56].

Application area: Contracting among agents in e-commerce applications.

Problem: In an open multi-agent system agents should be able to constrain the behavior of other agents with whom they are interacting, for example when making economic transactions. Thus it is necessary to have a way to enable agents to create new obligations which are enforced by the normative system. Contracts define the creation of new obligations and permissions by agents involved in a negotiation. The possibility to create new norms is defined by the legal system itself, using constitutive norms.

Technique: Introducing constitutive norms and their relation to regulative norms like obligations and permissions. Constitutive or counts-as norms have first been studied in speech act theory, and theories of construction of social reality. They define so-called institutional facts, and we extend their definition so they also define the way in which normative systems can change.

4.4 Norm creation games: MAS structure of a normative system, permission mechanism, with applications in legal theory.

We consider the agents from the viewpoint of the normative system: it not only views the agents as subjects which it obliges, prohibits and permits, sanctions

and controls, but it also views the agents as subjects that play roles in the system, for example which can modify the normative system [35]. We consider also game specification in normative multiagent systems based on the Trias Politica [39], and the evolution of artificial Social Systems [44],

Application area: Legal systems. Introducing norms in multi-agent systems requires a correct understanding of how norms work in human society. It is necessary to build formal models which match the complexities of legal systems.

Problems: Open multi-agent systems cannot be regulated off-line since new unforeseen situations emerge and new norms must be added runtime to deal with specific situations without having to modify the entire set of norms. Legal systems cannot be composed only of obligations, since it becomes too complex to specify new exceptional cases which are not subject to the obligations. So permissions are introduced as a way to specify exceptions. Moreover, the relation between permissions and obligations issued by different levels of hierarchical normative systems is considered.

Technique: Introducing permissions and their relation to obligations. Introducing hierarchical normative systems and priority relations among norms. Permissions as exceptions are defined in the input/output logic framework [63], and the notion of authorization is studied [48].

4.5 Control games: interaction among normative systems, nested norms mechanism, with applications in security and secure knowledge management systems

In [62] we consider systems that contain multiple normative systems. We consider also the delegation of control to autonomous agents (called defender agents) [33].

Application area: Regulating virtual communities of peer agents by means of policies, like in Secure knowledge management.

Problem: In scenarios like grid architectures or peer to peer systems every node can be seen as an autonomous agent. As such it can impose freely norms on the use of its resources. When the agent joins a virtual community, it must offer its resources at disposal of the other members according to the global policies of the community. Global policies about local policies, however, have not been studied yet, since they require nesting of norms.

Technique: When there are several normative systems, then global norms may put some restrictions on the possibility of issuing local norms. For example, European law limits the way European countries can create and enforce norms. The problem is how to describe global policies on local policies. Our game theoretic analysis illustrates that a naive approach does not work, and we therefore have to formalize an idea suggested by von Wright called delegation of will.

4.6 Related topics

Our work on the game-theoretic approach to normative multiagent systems has led us to study some related topics, where we have obtained the following results.

- Model of emergence of norms called the social delegation cycle [64], including a study of the use of power in norm negotiation [66].
- A foundational ontology of organizations and roles [57], an agent oriented ontology of social reality [36], and a model of organizations as socially constructed agents in the agent oriented paradigm [47]. We study the attribution of mental attitudes to roles [37] and we define groups as agents with mental attitudes [40]. We also consider the representation of organizations in artificial social systems [61].
  We develop an extension of the programming language Java called powerJava [10]. We consider power in powerJava [8] and interaction among objects [12]. We bridge agent theory and object orientation by importing social roles in object oriented languages [7] and by agent-like communication among objects [11].
- Based on various social viewpoints on multiagent systems [28] we consider definitions of coalitions [30,31] based on admissible agreements among goal-directed agents [25]. We consider reduction of coalition structures via contracts' representation [29] and an abstraction from power to coalition structures [27].
- We consider the role of roles in agent communication languages [24], and propose role based semantics as a synthesis between mental attitudes and social commitments [15,16,25]. We distinguish propositional and action commitment in agent communication [17]. We consider FIPA communicative acts also in defeasible logic [18].

5 Interdisciplinary aspects

Normative multiagent systems are an example of the use of sociological theories in multiagent systems, and more generally of the relation between agent theory and the social sciences such as sociology, philosophy, economics, and legal science. Other examples are the use of the social concept of "power" in programming language powerJava, see also [9,13]. As Castelfranchi [75] argues , not only can social and cognitive concepts like norms and beliefs should be used in agent theory, but agent theory should be used to enrich social sciences, making it more computational. For example, in the area of coordination and organization [53], organizational concepts may be used in coordination languages, or these languages may be used for human organizations. However, our theory is a computer science approach. It may be used outside computer science, but we restrict ourselves to computer science. For an example of how our theory may be used in cognitive science see [45]. For some new challenges for deontic logic due to our theory, see [22,34].

Our game-theoretic approach to normative systems combines deontic logic, agent architecture and artificial social systems. Deontic logic in computer science is an area of knowledge representation and reasoning within artificial intelligence, and related to philosophical logic. Agent architecture is an area of artificial intelligence related to software engineering and psychology. Artificial social systems

are a kind of game theory related to economics, linguistics and organizational theory.

The need for social science theories and concepts like norms in multiagent systems is now well established. For example, Wooldridge's weak notion of agency is based on flexible autonomous action [115], and social ability as the interaction with other agents and co-operation is one of the three meanings of flexibility; the other two are reactivity as interaction with the environment, and pro-activeness as taking the initiative. In this definition autonomy refers to non-social aspects, such as operating without the direct intervention of humans or others, and have some kind of control over their actions and internal state. For some other arguments for the need for social theory in multiagent systems, see, for example, [69,79,114]. For a more complete discussion on the need of social theory in general, and norms in particular, see the AgentLink roadmap [93].

Social concepts like norms are important for multiagent systems, because multiagent system research and sociology share the interest in the relation between micro-level agent behaviour and macro-level system effects. In sociology this is the (in)famous micro-macro link [3] that focuses on the relation between individual agent behaviour and characteristics at the level of the social system. In multiagent system research, this boils down to the question "How to ensure efficiency at the level of the multiagent system whilst respecting individual autonomy?". According to Verhagen [113] three possible solutions to this problem comprise of the use of central control which gravely jeopardizes the agent's autonomy, internalized control like the use of social laws [106], and structural coordination [103] including learning norms.

6 Summary

Normative multi-agent systems study general and domain independent properties of norms. It builds on results obtained in deontic logic, the logic of obligations and permissions, for the representation of norms as rules, the application of such rules, contrary-to-duty reasoning and the relation to permissions. However, it goes beyond logical relations among obligations and permissions by explaining the relation among social norms and obligations, relating regulative norms to constitutive norms, explaining the evolution of normative systems, and much more. Our work on the game-theoretic approach to normative multi-agent systems has the following four objectives:

1. How to combine the three existing representations of normative multiagent systems? A representation of a normative multiagent systems that builds on and extends the deontic logic approach for the explicit representation of norms, the agent architecture approach for software engineering of agents, and the game-theoretic artificial social systems approach for model of interaction.
2. Why do we need norms in computer science? A logical framework for qualitative risk analysis, building on ideas in security and risk management.

3. When do we need which kind of norms? A classification of situations in which one needs which elements of normative systems (e.g., obligations, permissions, counts-as conditionals).
4. How can we use norms in computer science applications? Examples and classification of which kind of normative multiagent systems are to be used for which kind of applications in computer science.

In our approach we use input/output logic for deontic logic, BOID architecture for the agent architecture, and both game theory and recursive modeling for artificial social systems. The following table summarizes our results:

Behavior	violation games	institutional games	negotiation games	norm creation games	control games
Structure	Interacting with NS		MAS interaction in NS	MAS structure of NS	Interaction among NSs
Mechanism	Obligation	Counts-as	Norm creation actions	Permission	nested rules
Theory	BDI, deontic logic				
Application	Trust Fraud, deception grid, p2p	Distributed systems Virtual communities	e-commerce e-contracting	security, SKM, policies	

In our work, we use game theory as basis of NMAS. However, vice versa, it is an open problem whether our model / approach also has something to say about the role of norms in game theory.

References

1. Alchourrón, C., "Philosophical foundations of deontic logic and the logic of defeasible conditionals", in: Meyer, J.-J. and Wieringa, R. (eds.), *Deontic Logic in Computer Science: Normative System Specification*, John Wiley & Sons, 1993, 43–84.
2. Alchourrón, C. and Bulygin, E., *Normative Systems*, Wien: Springer, 1971.
3. Alexander, J. C., Giesen, B., Munch, R. and Smelser, N. J., *The micro-macro link*, Berkeley: University of California Press, 1987.
4. Anderson, A., "A reduction of deontic logic to Alethic modal logic", *Mind*, **67**, 1958, 100–103.
5. Anderson, A., "Some nasty problems in the formalization of ethics", *Noûs*, **1**, 1967, 345–360.
6. Aqvist, L., "Deontic Logic", in: Gabbay, D. and Guenthner, F. (eds.), *Handbook of Philosophical Logic: Volume II: Extensions of Classical Logic*, Dordrecht: Reidel, 1984, 605–714.
7. Baldoni, M., Boella, G. and van der Torre, L., "Bridging Agent Theory and Object Orientation: Importing Social Roles in Object Oriented Languages", in: *Programming Multi-Agent Systems, Third International Workshop, ProMAS 2005*, vol. 3862 of *LNCS*, Berlin: Springer, 2006, 57–75.
8. Baldoni, M., Boella, G. and van der Torre, L., "I fondamenti ontologici dei linguaggi di programmazione orientati agli oggetti: i casi delle relazioni e dei ruoli", *Networks, rivista di filosofia dell'intelligenza artificiale e scienze cognitive*, **6**, 2006.
9. Baldoni, M., Boella, G. and van der Torre, L., "Modelling the Interaction Between Objects: Roles as Affordances.", in: *Knowledge Science, Engineering and Management, First International Conference, KSEM 2006*, vol. 4092 of *LNCS*, Springer, 2006, 42–54.

10. Baldoni, M., Boella, G. and van der Torre, L., "Roles as a Coordination Construct: Introducing powerJava", *Electronic Notes in Theoretical Computer Science (ENTCS) Procs. of the First International Workshop on Methods and Tools for Coordinating Concurrent, Distributed and Mobile Systems (MTCoord 2005)*, **150**(1), 2006, 9–29.
11. Baldoni, M., Boella, G. and van der Torre, L., "Bridging Agent Theory and Object Orientation: Interaction among Objects", in: *Programming Multi-Agent Systems, Fourth International Workshop, ProMAS 2006*, vol. 4411 of *LNCS*, Springer, 2007, 151–166.
12. Baldoni, M., Boella, G. and van der Torre, L., "Interaction between Objects in powerJava", *Journal of Object Technology*, **6**(2), 2007, 7–12.
13. Baldoni, M., Boella, G. and van der Torre, L., "Relationships meet their roles in object oriented programming", in: *Procs. of the 2nd International Symposium on Fundamentals of Software Engineering 2007 Theory and Practice (FSEN '07)*, 2007.
14. Beccaria, C., *Dei delitti e delle pene*, Livorno, 1764.
15. Boella, G., Damiano, R., Hulstijn, J. and van der Torre, L., "ACL Semantics between Social Commitments and Mental Attitudes", in: *International Workshops on Agent Communication, AC 2005 and AC 2006*, vol. 3859 of *LNAI*, Berlin: Springer, 2006, 30–44.
16. Boella, G., Damiano, R., Hulstijn, J. and van der Torre, L., "Role-Based Semantics for Agent Communication: Embedding of the Mental Attitudes and Social Commitments Semantics", in: *Procs. of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)*, New York (NJ): ACM, 2006, 688–690.
17. Boella, G., Damiano, R., Hulstijn, J. and van der Torre, L., "Distinguishing Propositional and Action Commitment in Agent Communication", in: *Procs. of the 7th Workshop on Computational Models of Natural Argument (CMNA'07)*, 2007.
18. Boella, G., Hulstijn, J., Governatori, G., Riveret, R., Rotolo, A. and van der Torre, L., "FIPA Communicative Acts in Defeasible Logic", in: *Procs. of the 7th International Workshop on Nonmontonic Reasoning, Action and Change (NRAC'07)*, 2007.
19. Boella, G., Hulstijn, J., Tan, Y. and van der Torre, L., "Transaction trust in normative multiagent systems", in: *Procs. of Trust in Agent Societies Workshop at AAMAS'05*, 2005.
20. Boella, G., Hulstijn, J., Tan, Y. and van der Torre, L., "Modeling Control Mechanisms with Normative Multiagent Systems: The Case of the Renewables Obligation", in: *Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems AAMAS 2005 International Workshops on Agents, Norms, and Institutions for Regulated Multiagent Systems, ANIREM 2005 and on Organizations in Multi-Agent Systems, OOOP 2005*, vol. 3913 of *LNAI*, Berlin: Springer, 2006, 114–126.
21. Boella, G., Hulstijn, J. and van der Torre, L., "Argumentation for Access Control.", in: *AI*IA 2005: Advances in Artificial Intelligence, 9th Congress of the Italian Association for Artificial Intelligence*, vol. 3673 of *LNCS*, Berlin: Springer, 2005, 86–97.
22. Boella, G., Hulstijn, J. and van der Torre, L., "Interaction in Normative Multi-Agent Systems.", *Electronic Notes in Theoretical Computer Science, Proceedings of the Workshop on the Foundations of Interactive Computation (FInCo 2005)*, **141**(5), 2005, 135–162.

23. Boella, G., Hulstijn, J. and van der Torre, L., "Virtual organizations as normative multiagent systems", in: *Procs. of the 38th Hawaii International Conference on System Sciences (HICSS-38 2005)*, 2005.
24. Boella, G., Hulstijn, J. and van der Torre, L., "The Roles of Roles in Agent Communication Languages", in: *Procs. of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'06)*, IEEE, 2006, 381–384.
25. Boella, G., Hulstjin, J. and van der Torre, L., "A Synthesis Between Mental Attitudes and Social Commitments in Agent Communication Languages", in: *Procs. of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'05)*, IEEE, 2005, 358–364.
26. Boella, G. and Lesmo, L., "A game theoretic approach to norms", *Cognitive Science Quarterly*, **2(3-4)**, 2002, 492–512.
27. Boella, G., Sauro, L. and van der Torre, L., "An Abstraction from Power to Coalition Structures", in: *Procs. of the 16th European Conference on Artificial Intelligence (ECAI'04)*, Amsterdam: IOS, 2004, 965–966.
28. Boella, G., Sauro, L. and van der Torre, L., "Social Viewpoints on Multiagent Systems", in: *Procs. of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'04)*, New York (NJ): ACM, 2004, 1358–1359.
29. Boella, G., Sauro, L. and van der Torre, L., "Reducing Coalition Structures via contracts' representation", in: *Procs. of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05)*, New York (NJ): ACM, 2005, 1187–1188.
30. Boella, G., Sauro, L. and van der Torre, L., "Strengthening Admissible Coalitions", in: *Procs. of the 17th European Conference on Artificial Intelligence (ECAI'06)*, Amsterdam: IOS, 2006, 195–199.
31. Boella, G., Sauro, L. and van der Torre, L. W. N., "From Social Power to Social Importance", *Web Intelligence and Agent Systems Journal (WIAS)*, 2007.
32. Boella, G. and van der Torre, L., "Attributing mental attitudes to normative systems", in: *Procs. of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)*, New York (NJ): ACM Press, 2003, 942–943.
33. Boella, G. and van der Torre, L., "Norm Governed Multiagent Systems: The delegation of control to autonomous agents", in: *Proceedings of the 2003 IEEE/WIC International Conference on Intelligent Agent Technology (IAT'03)*, IEEE, 2003, 329– 335.
34. Boella, G. and van der Torre, L., "Obligations as Social Constructs", in: *AI*IA 2003 - Advances in Artificial Intelligence, 8th Congress of the Italian Association for Artificial Intelligence*, vol. 2829 of *LNAI*, Berlin: Springer, 2003, 27–38.
35. Boella, G. and van der Torre, L., "Rational norm creation: Attributing mental attitudes to normative systems, part 2", in: *Procs. of the 8th International Conference on Artificial Intelligence and Law (ICAIL'03)*, New York (NJ): ACM Press, 2003, 81–82.
36. Boella, G. and van der Torre, L., "An agent oriented ontology of social reality", in: *Procs. of Formal Ontologies in Information Systems (FOIS'04)*, Amsterdam: IOS, 2004, 199–209.
37. Boella, G. and van der Torre, L., "Attributing Mental Attitudes to Roles: The Agent Metaphor Applied to Organizational Design", in: *Procs. of the 6th International Conference on Electronic Commerce (ICEC'04)*, New York (NJ): ACM, 2004, 130–137.

38. Boella, G. and van der Torre, L., "The Distribution of Obligations by Negotiation among Autonomous Agents", in: *Procs. of the 16th European Conference on Artificial Intelligence (ECAI'04)*, Amsterdam: IOS, 2004, 13–17.
39. Boella, G. and van der Torre, L., "Game Specification in Normative Multiagent System: the Trias Politica", in: *Procs. of IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'04)*, IEEE, 2004, 504–508.
40. Boella, G. and van der Torre, L., "Groups as Agents with Mental Attitudes", in: *Procs. of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'04)*, New York (NJ): ACM, 2004, 964–971.
41. Boella, G. and van der Torre, L., "Local vs Global Policies and Centralized vs Decentralized Control in Virtual Communities of Agents", in: *Procs. of IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)*, IEEE, 2004, 690–693.
42. Boella, G. and van der Torre, L., "Regulative and Constitutive Norms in Normative Multiagent Systems", in: *Procs. of the 10th International Conference on the Principles of Knowledge Representation and Reasoning KR'04*, Menlo Park (CA): AAAI, 2004, 255–265.
43. Boella, G. and van der Torre, L., "Enforceable social laws", in: *Procs. of 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05)*, New York (NJ): ACM, 2005, 682–689.
44. Boella, G. and van der Torre, L., "The Evolution of Artificial Social Systems", in: *Procs. of the 19th International Joint Conference on Artificial Intelligence (IJCAI'05)*, Professional Book Center, 2005, 1655–1556.
45. Boella, G. and van der Torre, L., "From the Theory of Mind to the Construction of Social Reality", in: *Procs. of the 27th Annual Conference of the Cognitive Science Society (CogSci'05)*, Mahwah (NJ): Lawrence Erlbaum, 2005, 298–303.
46. Boella, G. and van der Torre, L., "Normative multiagent systems and trust dynamics", in: *Trusting Agents for Trusting Electronic Societies, Theory and Applications in HCI and E-Commerce*, vol. 3577 of *LNAI*, Berlin: Springer, 2005, 1–17.
47. Boella, G. and van der Torre, L., "Organizations as Socially Constructed Agents in the Agent Oriented Paradigm", in: *Engineering Societies in the Agents World V, 5th International Workshop (ESAW'04)*, vol. 3451 of *LNAI*, Berlin: Springer, 2005, 1–13.
48. Boella, G. and van der Torre, L., "Permission and Authorization in Normative Multiagent Systems", in: *Procs. of the 10th International Conference on Artificial Intelligence and Law (ICAIL'05)*, New York (NJ): ACM, 2005, 236–237.
49. Boella, G. and van der Torre, L., "Permission and Authorization in Policies for Virtual Communities of Agents", in: *Agents and Peer-to-Peer Computing, Third International Workshop (AP2PC'04)*, vol. 3601 of *LNCS*, Berlin: Springer, 2005, 86–97.
50. Boella, G. and van der Torre, L., "Role-based rights in artificial social systems", in: *Procs. of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'05)*, IEEE, 2005, 516–519.
51. Boella, G. and van der Torre, L., "An architecture of a normative system", in: *Procs. of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)*, New York (NJ): ACM, 2006, 229–231.
52. Boella, G. and van der Torre, L., "Constitutive Norms in the Design of Normative Multiagent Systems", in: *Computational Logic in Multi-Agent Systems, 6th International Workshop (CLIMA VI)*, vol. 3900 of *LNCS*, Berlin: Springer, 2006, 303–319.

53. Boella, G. and van der Torre, L., "Coordination and Organization: Definitions, Examples and Future Research Directions.", *Electronic Notes in Theoretical Computer Science (ENTCS) Procs. of the First International Workshop on Coordination and Organisation (CoOrg 2005)*, **150**(3), 2006, 3–20.
54. Boella, G. and van der Torre, L., "Count-As Conditionals, Classification and Context.", in: *Procs. of the 17th European Conference on Artificial Intelligence (ECAI'06)*, Amsterdam: IOS, 2006, 719–720.
55. Boella, G. and van der Torre, L., "Delegation of Power in Normative Multiagent Systems", in: *Deontic Logic and Artificial Normative Systems, 8th International Workshop on Deontic Logic in Computer Science (ΔEON'06)*, vol. 4048 of *LNCS*, Berlin: Springer, 2006, 36–52.
56. Boella, G. and van der Torre, L., "Fair Distribution of Collective Obligations.", in: *Procs. of the 17th European Conference on Artificial Intelligence(ECAI'06)*, Amsterdam: IOS, 2006, 721–722.
57. Boella, G. and van der Torre, L., "A Foundational Ontology of Organizations and Roles", in: *Declarative Agent Languages and Technologies IV, 4th International Workshop (DALT'06)*, vol. 4327 of *LNCS*, 2006, 78–88.
58. Boella, G. and van der Torre, L., "A Game Theoretic Approach to Contracts in Multiagent Systems", *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, **36**(1), 2006, 68–79.
59. Boella, G. and van der Torre, L., "Game-Theoretic Foundations for Norms", in: *Procs. of Artificial Intelligence Studies*, vol. 3(26), 2006, 39–51.
60. Boella, G. and van der Torre, L., "A Logical Architecture of a Normative System", in: *Deontic Logic and Artificial Normative Systems, 8th International Workshop on Deontic Logic in Computer Science (ΔEON'06)*, vol. 4048 of *LNCS*, Berlin: Springer, 2006, 24–35.
61. Boella, G. and van der Torre, L., "Organizations in Artificial Social Systems", in: *Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems AAMAS 2005 International Workshops on Agents, Norms, and Institutions for Regulated Multiagent Systems, ANIREM 2005 and on Organizations in Multi-Agent Systems, OOOP 2005*, vol. 3913 of *LNAI*, Berlin: Springer, 2006, 198–210.
62. Boella, G. and van der Torre, L., "Security Policies for Sharing Knowledge in Virtual Communities", *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, **36**(3), 2006, 439–450.
63. Boella, G. and van der Torre, L., "Institutions with a Hierarchy of Authorities in Distributed Dynamic Environments", *Artificial Intelligence and Law Journal (AILaw)*, 2007.
64. Boella, G. and van der Torre, L., "Norm Negotiation in Multiagent Systems", *International Journal of cooperative Information Systems (IJCIS) Special Issue: Emergent Agent Societies*, **16**(2), 2007, 97–122.
65. Boella, G. and van der Torre, L., "The Ontological Properties of Social Roles in Multi-agent Systems: Definitional Dependence, Powers and Roles Playing Roles", *Artificial Intelligence and Law Journal (AILaw)*, 2007.
66. Boella, G. and van der Torre, L., "Power in Norm Negotiation", in: *Procs. of the 1st KES Symposium on Agent and Multi-Agent Systems Technologies and Applications (KES-AMSTA'07)*, LNCS, Berlin: Springer, 2007.
67. Boella, G. and van der Torre, L., "Substantive and Procedural Norms in Normative Multiagent Systems", *Journal of Applied Logic*, 2008.
68. Boella, G., van der Torre, L. and Verhagen, H., "Introduction to normative multiagent systems", *Computation and Mathematical Organizational Theory, Special issue on Normative Multiagent Systems*, **12**(2-3), 2006, 71–79.

69. Bond, A. and Gasser, L., "An Analysis of Problems and Research in DAI", in: *Readings in Distributed Artificial Intelligence*, San Mateo (CA): Morgan Kaufmann, 1988, 3–35.
70. Brafman, R. and Tennenholtz, M., "On Partially Controlled Multi-Agent Systems.", *Journal of Artificial Intelligence Research (JAIR)*, **4**, 1996, 477–507.
71. Bratman, M., *Intentions, plans, and practical reason*, Harvard (Massachusetts): Harvard University Press, 1987.
72. Broersen, J., Dastani, M., Hulstijn, J. and van der Torre, L., "Goal generation in the BOID architecture", *Cognitive Science Quarterly*, **2(3-4)**, 2002, 428–447.
73. Broersen, J., Dastani, M., Hulstijn, J. and van der Torre, L., "Goal generation in the BOID architecture", *Cognitive Science Quarterly*, **2(3-4)**, 2002, 428–447.
74. Castelfranchi, C., "Engineering social order", in: *Engineering Societies in the Agent World, First International Workshop (ESAW'00)*, vol. 1972 of *LNAI*, Berlin: Springer, 2000, 1–18.
75. Castelfranchi, C., "The micro-macro constitution of power", *Protosociology*, **18**, 2003, 208–269.
76. Conte, R., Castelfranchi, C. and Dignum, F., "Autonomous norm-acceptance", in: *Intelligent Agents V (ATAL'98)*, vol. 1555 of *LNCS*, Berlin: Springer, 1999, 99–112.
77. Ferber, J., Gutknecht, O., Jonker, C., Müller, J. P. and Treur, J., "Organization Models and Behavioural Requirements Specification for Multi-Agent Systems", in: *Procs. of Modelling Autonomous Agents in a Multi-Agent World - 10th European Workshop on Multi-Agent Systems (MAAMAW'01)*, 2002.
78. Ferber, J., Gutknecht, O. and Michel, F., "From Agents to Organizations: an organizational view of multiagent systems", in: *Agent-Oriented Software Engineering IV, 4th International Workshop (AOSE'03)*, vol. 2935 of *LNCS*, Berlin: Springer, 2003, 214–230.
79. Gilbert, N. and Conte, R., *Artificial Societies: The Computer Simulation of Social Life*, London: UCL Press, 1995.
80. Gmytrasiewicz, P. J. and Durfee, E. H., "Formalization of recursive modeling", in: *Procs. of the 1st International Conference on Multiagent Systems (ICMAS'95)*, Cambridge (MA): AAAI/MIT Press, 1995, 125–132.
81. Gneezy, U. and Rustichini, A., "A Fine is a Price", *The Journal of Legal Studies*, **29**(1), 2000, 1–18.
82. Goffman, E., *Strategic Interaction*, Oxford: Basil Blackwell, 1970.
83. Gordijn, J. and Tan, Y.-H., "A Design Methodology for Trust and Value Exchanges in Business Models", in: *Procs. of BLED Conference*, 2003, 423–432.
84. Grossi, D., Dignum, F. and Meyer, J., "Contextual Terminologies", in: *Computational Logic in Multi-Agent Systems, 6th International Workshop (CLIMA VI)*, vol. 3900 of *LNCS*, Berlin: Springer, 2006, 284–302.
85. Grossi, D., Meyer, J.-J. and Dignum, F., "Counts-as: Classification or Constitution? An Answer Using Modal Logic", in: *Deontic Logic and Artificial Normative Systems, 8th International Workshop on Deontic Logic in Computer Science (ΔEON'06)*, vol. 4048 of *LNCS*, Berlin: Springer, 2006, 115–130.
86. Hart, H., *The Concept of Law*, Oxford: Clarendon Press, 1961.
87. Hollis, M., *Trust within reason*, Cambridge: Cambridge University Press, 1998.
88. Jennings, N. R., "On Agent-Based Software Engineering", *Artificial Intelligence*, **117(2)**, 2000, 277–296.
89. Jones, A. and Carmo, J., "Deontic logic and Contrary-to-Duties", in: Gabbay, D. and Guenthner, F. (eds.), *Handbook of Philosophical Logic*, vol. 3, Dordrecht (NL): Kluwer, 2001, 203–279.

90. Jones, A. and Sergot, M., "A Formal Characterisation of Institutionalised Power", *Journal of IGPL*, **3**, 1996, 427–443.
91. Lee, R., "Documentary Petri Nets: A Modeling Representation for Electronic Trade Procedures", in: *Business Process Management, Models, Techniques, and Empirical Studies*, vol. 1806 of *LNCS*, Berlin: Springer, 2000, 359–375.
92. Levitt, S. D. and Dubner, S. J., *Freakonomics : A Rogue Economist Explores the Hidden Side of Everything*, New York: William Morrow, 2005.
93. Luck, M., McBurney, P. and Preist, C., *Agent Technology: Enabling Next Generation Computing (A Roadmap for Agent Based Computing)*, AgentLink, 2003.
94. Makinson, D. and van der Torre, L., "Input-output logics", *Journal of Philosophical Logic*, **29**(4), 2000, 383–408.
95. Makinson, D. and van der Torre, L., "Constraints for input-output logics", *Journal of Philosophical Logic*, **30(2)**, 2001, 155–185.
96. Makinson, D. and van der Torre, L., "Permissions from an input-output perspective", *Journal of Philosophical Logic*, **32(4)**, 2003, 391–416.
97. McCallum, M., Norman, T. and Vasconcelos, W., "A Formal Model of Organisations for Engineering Multi-Agent Systems", in: *Procs. of Coordination in Emergent Agent Societies Workshop (CEAS'04)*, 2004.
98. Merriam-Webster, *Dictionary of Law*, Merriam-Webster, 1996.
99. Merriam-Webster, *On Line Dictionary*, Merriam-Webster, 2007.
100. Meyer, J.-J. and Wieringa, R., *Deontic Logic in Computer Science: Normative System Specification*, Chichester, England: John Wiley & Sons, 1993.
101. Meyer, J. J. C., "A Different Approach to Deontic Logic: Deontic Logic Viewed as a Variant of Dynamic Logic", *Notre Dame Journal of Formal Logic*, **29(1)**, 1988, 109–136.
102. Newell, A., "The knowledge level", *Artificial Intelligence*, **18**, 1982, 87–127.
103. Ossowski, S., *Co-Ordination in Artificial Agent Societies: Social Structures and Its Implications for Autonomous Problem-Solving Agents*, Berlin: Springer, 1999.
104. Ruiter, D., "A basic classification of legal institutions", *Ratio Juris*, **10(4)**, 1997, 357–371.
105. Searle, J., *The Construction of Social Reality*, New York: The Free Press, 1995.
106. Shoham, Y. and Tennenholtz, M., "On Social Laws for Artificial Agent Societies: Off-Line Design", *Artificial Intelligence*, **73**(1-2), 1995, 231–252.
107. Shoham, Y. and Tennenholtz, M., "On the Emergence of Social Conventions: Modeling, Analysis and Simulations", *Artificial Intelligence*, **94**(1–2), 1997, 139–166.
108. Tennenholtz, M., "On Stable Social Laws and Qualitative Equilibria", *Artificial Intelligence*, **102(1)**, 1998, 1–20.
109. van der Torre, L., "Contextual Deontic Logic: Normative Agents, Violations and Independence", *Annals of Mathematics and Artificial Intelligence*, **37(1-2)**, 2003, 33–63.
110. van der Torre, L. and Tan, Y., "Contrary-To-Duty Reasoning with Preference-based Dyadic Obligations", *Annals of Mathematics and Artificial Intelligence*, **27**(1-4), 1999, 49–78.
111. Tuomela, R., *Cooperation: A Philosophical Study*, Dordrecht: Kluwer, 2000.
112. Verhagen, H., "On the learning of norms", in: *Procs. of MultiAgent System Engineering, 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'99)*, vol. 1647 of *LNCS*, Berlin: Springer, 1999.
113. Verhagen, H., *Norm Autonomous Agents*, Ph.D. thesis, Stockholm University, 2000.

114. Verhagen, H. and Smit, R., "Multiagent systems as simulation tools for social theory testing", in: *Procs. of International Conference on Computer Simulation and the Social Sciences (ISSC&SS'97)*, 1997.
115. Wooldridge, M. J. and Jennings, N. R., "Intelligent Agents: Theory and Practice", *Knowledge Engineering Review*, **10**(2), 1995, 115–152.
116. von Wright, G., "Deontic logic - as I see it", in: McNamara, P. and Prakken, H. (eds.), *Norms, Logics and Information Systems. New Studies on Deontic Logic and Computer Science*, IOS, 1999, 15–25.
117. von Wright, G. H., "Deontic logic", *Mind*, **60**, 1951, 1–15.
118. von Wright, G. H., *An Essay in Modal Logic*, Amsterdam: North-Holland, 1951.
119. Zambonelli, F., Jennings, N. and Wooldridge, M., "Developing Multiagent Systems: The Gaia Methodology", *IEEE Transactions of Software Engineering and Methodology*, **12(3)**, 2003, 317–370.