A Game Theoretic Approach to Contracts in Multiagent Systems

Guido Boella and Leendert van der Torre

***Abstract*— Contracts are used to create new interaction possibilities among agents, and they therefore play an important role in the game theoretic analysis of agent interaction. We use normative multiagent systems to model both the contracts and the interactions. In particular, we formalize contracts as systems of regulative and constitutive norms within a larger rule-governed setting, and using recursive modelling we develop a game theory where agents make contracts. We show how agents can modify the behavior of normative systems by means of constitutive rules in the contract changing these systems, and we illustrate how agents use the game theory within contract negotiation in organizations.**

***Index Terms*— Contracts, qualitative game theory, multiagent systems, normative systems.**

I. Introduction

Autonomous agents negotiating deals on behalf of human traders are essential in e-commerce and e-trading systems. Many researchers focus their approaches on the game-theoretic analysis of the interaction among negotiation agents to prove macro-level properties like system stability (equilibrium) as well as efficiency (Pareto-optimality). However, the interaction structure in many multiagent systems is not completely fixed in advance to preserve the autonomy of agents. For example, as Dignum *et al.* [1] note, in (virtual) organizations the interaction possibilities can be changed and negotiated. For this reason, several approaches like [1]–[4] introduce the possibility for agents to stipulate contracts. A contract can be defined as a statement of intent regulating behavior among agents and organizations. Contracts have been proposed to make explicit how agents can change the interaction with and within the organization, using the legal effects of contracts involving the creation of obligations, permissions and new possibilities of interaction. From a contractual perspective, organizations can be seen as the possible sets of agreements for satisfying the diverse interests of self interested agents [2].

Hanson and Milosevic [3] distinguish various phases in the negotiation of contracts, like contract schema selection, issue identification, negotiation of values of the issues agreed upon, monitoring the performance of the contract, *et cetera*. As Marsh suggests in the contract negotiation handbook [5], a crucial point in the negotiation is the careful planning of the moves during the bargaining. Accordingly, agents must evaluate the effects of a contract during its negotiation both when proposing and evaluating an offer. In particular, since the compliance of the agents to the contract cannot be taken for granted, an agent must consider whether the other agents will fulfill their commitments or not. To make a prediction about the behavior of an agent it is necessary to consider also the reaction of the normative system [6] acting as a coordinator enforcing control by monitoring and sanctioning violations.

In this paper we develop a formal game-theoretic model for autonomous agents negotiating contracts, which we model as a normative multiagent system, that 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 [7]." It is based on a general conceptual model formalizing contracts as so-called legal institutions, that is, as sets of norms within a larger legal setting. In this way, agents are made competent to determine the course of law within the sphere of the contracts they are allowed to create. To model the interaction among agents and the normative system in this game-theoretic setting, we apply the agent metaphor to normative systems and therefore model the interaction among agents and their normative system analogously to the interaction among real agents.

We define games using recursive modelling. Violation games model an agent deliberating whether to violate a norm or not, where the agent recursively models the normative system to predict whether its behavior counts as a violation and will be sanctioned. The set of possible behaviors includes the possibility of making contracts to change the agent's normative position based on constitutive rules that change the normative system. Negotiation games model an agent deliberating whether it will propose or accept a contract, thereby recursively modelling the other agent to find out whether it will violate the norms in the contract. In this sense, negotiation games add a recursive level of reasoning to violation games. We illustrate how agents can reason in the game theory using an example from international trade.

This paper does not address the problem of equilibrium analysis in the proposed game theory or the negotiation protocol to obtain an agreement. Instead we examine the impact of contracts on the games played by the agents and how the agents reason about the legal effects of contracts.

The layout of this paper is as follows. In Section II we discuss regulative and constitutive rules, legal institutions and how agents modify the behavior of normative systems using contracts. In Section III we discuss the foundations of our model and in Section IV we introduce our formal model applying the agent metaphor to normative systems. In Section V we present recursive modelling, the games which can be played with contracts, and an example illustrating how agents use the game theory for contracting.

Manuscript received XXXXXX XX, 20XX; revised XXXXXX XX, 20XX.

Guido Boella is with Dipartimento di Informatica, Università di Torino, Italy (e-mail: guido@di.unito.it).

Leendert van der Torre is with CWI Amsterdam and Delft University of Technology, The Netherlands (e-mail: leendert@vandertorre.com).

Digital Object Identifier XXX.

II. Contracts as legal institutions

Most formalizations of normative systems identify norms with regulative norms like obligations, prohibitions and permissions. However, Searle [8] argues that there is a distinction between two types of rules.

> "Some rules regulate antecedently existing forms of behaviour. For example, the rules of polite table behaviour regulate eating, but eating exists independently of these rules. Some rules, on the other hand, do not merely regulate an antecedently existing activity called playing chess; they, as it were, create the possibility of or define that activity. The activity of playing chess is constituted by action in accordance with these rules. The institutions of marriage, money, and promising are like the institutions of baseball and chess in that they are systems of such constitutive rules or conventions" ([8], p. 131).

For Searle, regulative and constitutive norms are related via institutional facts like marriage, money and private property. They emerge from an independent ontology of "brute" physical facts through constitutive rules of the form "such and such an X counts as Y in context C" where X is any object satisfying certain conditions and Y is a label that qualifies X as being something of an entirely new sort. E.g., "X counts as a presiding official in a wedding ceremony", "this bit of paper counts as a five euro bill" and "this piece of land counts as somebody's private property". Regulative norms refer to these institutional facts. E.g., consider a society which believes that a field fenced by an agent counts as the fact that the field is the agent's property. The fence is a physical "brute" fact, while being a property is an institutional fact. Regulative norms forbidding trespassing refer to the abstract concept of property rather than to fenced fields.

Moreover, the philosopher of law Ruiter [9] shows, from the legal point of view, that legal effects of actions of the members of a legal system are complex and contracts do not concern only the regulative aspects of a legislation or the constitutive part of it. Rather, contracts are *legal institutions*: "systems of [regulative and constitutive] rules that provide frameworks for social action within larger rule-governed settings" [9]. This systemic view of legal institutions emerged only recently in legal studies, since legal positivism [10] mainly focused on the regulative aspects of law and its justification.

To formalize contracts as legal institutions, we have to extend Searle's model. Searle's analysis of constitutive rules has focused mainly on the attribution of a new functional status to entities like marriages, money, and property. Searle's idea is that constitutive rules "create the possibility or define that activity." In our model the role of constitutive rules is not limited to the creation of an activity and the construction of new abstract legal categories. Constitutive norms specify besides the creation of legal categories also the evolution of the system. The normative system itself specifies by means of constitutive rules how its state can be changed, who can change it, and the limits of the possible changes. In this way, complex normative systems achieve a legal regime that includes rules conferring legal powers on participants: an agent is turned into a "private legislator" (Hart [10]): "he is made competent to determine the course of law within the sphere of his contracts, trusts, wills and other structures [...] which he is enabled to build". Agents become able to design "relatively independent *institutional legal orders* within the comprehensive legal orders" (Ruiter [9]).

The regime of a legal institution can be defined as the set of legal consequences that flow from the existence of the institution. However, the meaning of "legal consequences" differs from what is normally understood by the term. Usually, since obligations have a conditional nature, when the condition of an obligation is satisfied, as a legal consequence the addressee of the obligation is categorically obliged to fulfill it. Legal institutions, like contracts, marriages and properties, refer to a different kind of legal consequences. E.g., the legal rule "in a marriage parents have the reciprocal obligation to take care of and support their children" is not a conditional obligation. It expresses the fact that only when a legal institution of marriage between Amy and Bob is created, the obligation is created that Amy and Bob take care and support their children. The same happens with the legal institution of contracts. When a contract comes into existence it creates obligations for the agents, i.e., new regulative norms which the normative system considers as its own. E.g., the Italian Civil Code art. 1173 (sources of obligations) specifies that obligations are created by contracts and art. 1372 (efficacy of contracts) that a contract has the strength of law (a contract is an agreement among two or more agents to regulate a juridical relationship about valuables *ex* art. 1321).

Therefore, contracts as legal institutions bring with them also constitutive rules creating not only new institutional facts, but also new obligations. In this way, it is possible to specify in a contract new procedures for the interaction among agents, to specify the evolution of the contract and how new obligations are created at a later stage. As Dignum *et al.* [1] notice, a contract specifies the events that alter the status of the contract. It is necessary to specify an interaction structure which indicates the possibilities of an agent and the consequences of its choices. The contract must specify how to proceed if a norm is violated and what the violator is expected to do. E.g., if a payment deadline is not respected, then the agent may be obliged to pay a double fee. Since we model contracts as legal institutions, this rule is not a conditional obligation, but it is an obligation created by an event specified in the contract, in the same way as the contract itself can create obligations. This is possible, because we consider a contract as a legal institution, which may be seen as a normative system inside the main normative system. As a normative system it specifies who has the power to introduce obligations.

To illustrate our model we use in this paper an example of Gordijn and Tan [11] about contracts inside a trade organization. In this example, the contracts are legal institutions within the larger context of legislation of international trade (the UN Convention on International Multimodal Transport of Goods, CIMTG). Gordijn and Tan show how the problem of trust between two agents exchanging goods for money can be solved by means of contracts offered by international trade organizations.

III. Foundations of our model

Inspired by the game-theoretic approach to obligations of Boella and Lesmo [12], we propose a logical multiagent framework for normative reasoning based on the philosophical foundations on strategic interaction in the work of sociologist Goffman [13]. “Strategic interaction” here means, according to Goffman, taking into consideration the actions of other agents:

> “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” [13, p.12].

Goffman sees norms as producing a form of strategic interaction between the agent and the normative system, and gives a game-theoretic interpretation of obligations. 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” [13, 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” [13, p.115]. Clearly, this approach is different from recent logical studies about norms and contracts based on modal logic like so-called deontic logic, see Section VI.

However, there are some problems to use Goffman’s strategic interaction to develop a game theoretic approach to contracts. In particular, classical decision and game theory have been criticized for their assumptions of ideality. Several alternatives have been proposed that take the bounded rationality of decision makers into account. For example, Newell [14] and others develop theories in artificial intelligence using the notion of goal. Agent theory replaces probabilities and utilities by informational (knowledge, belief) and motivational attitudes (goal, desire), and the decision rule by a process of deliberation. Bratman [15] further extends such theories with intentions for sequential decisions and norms for multiagent decision making (see the discussion in Section VI). Moreover, Gmytrasiewitcz and Durfee [16] replace the equilibria analysis in classical game theory by recursive modelling. It 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.

> “In order to solve its own decision-making situation, the agent needs an idea of what the other agents are likely to do. It can arrive at it by representing what it knows about the other agents’ decision-making situations, thus modelling them in terms of their own payoff matrices. The fact that other agents could also be modelling others, including the original agent, leads to a recursive nesting of models.”

Moreover, Goffman does not consider contracts, though his idea that the enforcement power is invested in a social office empowered to make final judgements and to institute payments is even more relevant when we consider besides norms also contracts. Contracts are used by agents to implement the agreements resulting from their negotiations and to change their normative situation. Social order in a multiagent system emerges from the contracts resulting from negotiations about the rights and duties of participants, rather than being given in advance. Moreover, as discussed in Section II, contracts are *legal institutions*, i.e., “systems of [regulative and constitutive] rules that provide frameworks for social action within larger rule-governed settings” [9], In our case, the larger setting is represented by a normative system which establishes the set of possible contracts. For this reason it is necessary to address the problem of contracts being aware of the peculiarities of legal institutions.

We emphasize here two properties of our model. First, legal institutions contain control procedures, which are policies and procedures that help to ensure that management directives are carried out [17], because, intentionally or not, an agent may fail to comply with the contract. For example, as suggested by Milosevic and Dromey [18], the specification given in a contract differs significantly from a computational specification in the expected degree of inconsistency. As Jones and Carmo observe, “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” [7]. We therefore represent norms as soft constraints, which are used in detective control systems where violations can be detected (you can enter a train without a ticket, but you may be checked and sanctioned), instead of hard constraints [19], which are restricted to preventative control systems that are built such that violations are impossible (you cannot enter a metro station without a ticket). Detective control is the result of the action of an agent, and consequently it is subject to errors and can be influenced by the actions of other agents.

Secondly, our model allows to analyze in detail both the decisions of agents and the creation of new regulative and constitutive norms *under existing obligations*. For example, Marsh [5] highlights that also the existing obligations and commitments of the agents play a role in contract negotiation, as well as the procedures they have to follow, in that they constrain both the possibility to propose and accept bids. Consequently, our model cannot only be used for the initial creation of the normative system with its contracts, but also for its evolution.

In our running example of contracts in a trade organization, there are several agents. On the one hand the seller does not want to ship the goods onto the carrier’s vessel without first receiving payment from the buyer. On the other hand the buyer does not want to pay the seller before the goods have been shipped. To solve this deadlock situation banks introduced the letter of credit: an agreement that the bank of the buyer will arrange the payment for the seller as soon as the seller can prove to the bank that he has shipped the goods. The bill of lading is issued by the carrier in return for the goods that he received from the seller. According to Article 10 of the CIMTG the bill of lading as shipment document reliably indicates that the goods have been shipped in international trade procedures. This is not a regulative norm, but a new constitutive rule added by the contract.

IV. The conceptual model of contracts

The conceptual model of the normative multiagent system is visualized in Figure 1. This figure can be interpreted using classical set theory. A box □ stands for a concept and is interpreted as a set. The lines and arrows — and → are interpreted as arbitrary relations over the sets they connect. A special relation is —▷, which is interpreted as the subset relation over the two concepts it connects: the first set is a subset of the second set. Finally, —◇ says that the singleton set o is an element of *SA*. Dashed boxes and lines have the same interpretation, but refer to the normative aspects of the model of multiagent systems. These four elements occur in most conceptual modelling languages, besides other elements. For example, in class diagrams in the unified modelling language (UML), □ is a class interpreted as a set of objects, — and → are associations among classes, —▷ is the "is-a" relation or subsumption relation, and —◇ is "part-of" or aggregation relation.

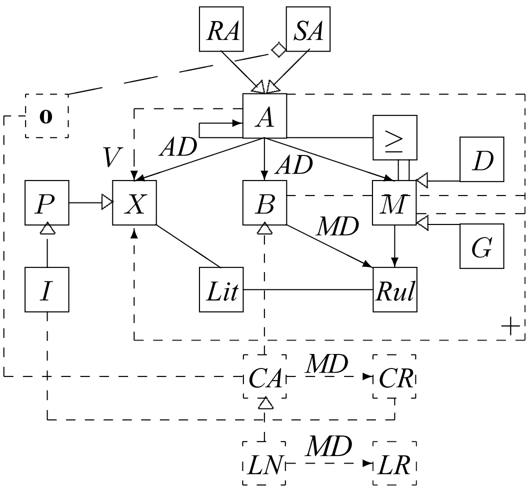

Fig. 1. Conceptual model of normative multiagent systems.

The intuitive reading of these concepts and the logical structure of the relations is detailed in the definitions below. In Section IV-A we explain the conceptual model of multiagent systems, and in Section IV-B we extend it to a normative multiagent systems by introducing violations. In Section IV-C we define how agents can change the normative system by means of constitutive rules by formalizing the constitutive rules as belief rules of the normative system. Finally in Section IV-D we define obligations in terms of desires and goals, and in Section IV-E contracts.

A. Multiagent systems

We first introduce the structural concepts and their relations. A set of propositional variables X describes the different aspects of the world, and we extend it to literals built out of X ($\mathit{Lit}(X)$) to consider also the absence of a state of affairs. Rules built out of the literals ($\mathit{Rul}(X)$) describe the relations among the propositional variables. A rule $l_1 \wedge \ldots \wedge l_n \rightarrow l$ is a pair of a set of literals built from X and a literal built from X. The left hand side is called antecedent or body, and the right hand side is called consequent or head. Rules are used to represent the relations among propositional variables existing in the agent's mental attitudes.

Definition 1: Let X be a set of variables. The set of literals built from X, written as $\mathit{Lit}(X)$, is $X \cup \{\neg x \mid x \in X\}$, and the set of rules built from X, written as $\mathit{Rul}(X)$, is defined by $2^{\mathit{Lit}(X)} \times \mathit{Lit}(X)$, the set of pairs of a set of literals built from X and a literal built from X. A rule is written as $\{l_1, \ldots, l_n\} \rightarrow l$, we also write $l_1 \wedge \ldots \wedge l_n \rightarrow l$ and when $n = 0$ we write $\top \rightarrow l$. Moreover, for $x \in X$ we write $\sim x$ for $\neg x$ and $\sim\neg x$ for x.

In the multiagent system, we consider various sorts of agents. Besides real agents *RA* (either human or artificial) we consider also socially constructed agents *SA* like groups, normative systems and organizations. These latter agents do not exist in the usual sense of the term. Rather, they exist only as they are attributed mental attitudes by real agents. The two sets of agents are disjoint and partition the set of agents A.

The mental attitudes attributed to agents consist of beliefs B, desires D and goals G. A mental description function *MD* associates a rule in $\mathit{Rul}(X)$ with each belief, desire and goal [20]. We introduce priority relations to resolve conflicts among motivations, i.e., desires and goals. A function $\geq$ associates with an agent a transitive and reflexive relation on the powerset of the motivations containing at least the subset relation. Moreover, different mental attitudes are attributed to agents by the agent description relation *AD*. It associates with each agent a set of beliefs, desires and goals. Moreover, *AD* associates also agents with agents, because normative systems and organizations exist only as they are described as agents by other agents. Formally, a socially constructed agent $b \in \mathit{SA}$ exists only as some other agents attribute mental attitudes to it: $\forall b \in \mathit{SA}\ \exists a \in A : b \in \mathit{AD}(a)$.

Multiagent systems also contain concepts concerning informational aspects. First of all, the variables whose truth value are determined by an agent (decision variables) are distinguished from those which are not directly determined by the agent (P, the parameters using Lang *et al.* [21]'s terminology). Besides, "institutional facts" I are states of affairs existing only inside normative systems and organizations. As discussed in Section II, Searle [22] suggests that money, properties, and marriages exist only as part of social reality. Since we model social reality by means of the attribution of mental attitudes to social entities, institutional facts exist at least in the beliefs of these socially constructed agents. We associate to each agent a subset of X by extending further the agent description relation *AD*. Moreover, the set of institutional facts I is a subset of the parameters.

Definition 2 (MAS): A multiagent system is a tuple $\langle \mathit{RA}, \mathit{SA}, X, B, D, G, \mathit{AD}, \mathit{MD}, \geq, I\rangle$, where:

- the real agents *RA*, socially constructed agents *SA*, variables X, beliefs B, desires D, and goals G are six finite disjoint sets. We write $A = \mathit{RA} \cup \mathit{SA}$ for all agents, and $M = D \cup G$ for motivations.
- an agent description $\mathit{AD} : A \rightarrow 2^{A \cup X \cup B \cup D \cup G}$ is a complete function that maps each agent to other agents that exist in its profile, to a set of variables (its decision

variables), and to its beliefs, desires and goals. We assume that there are no cycles in the mapping of agents to agents. For each agent $a \in A$, we write A_a for $A \cap AD(a)$, X_a for $X \cap AD(a)$, *et cetera*. We write $P = X \setminus \cup_{a \in A} X_a$ for the parameters.
- the mental description $MD : B \cup D \cup G \rightarrow Rul(X)$ is a complete function from the sets of beliefs, desires and goals to the set of rules built from X. For $S \subseteq B \cup D \cup G$, we write $MD(S) = \{MD(m) \mid m \in S\}$. Moreover, we write $m\ x \rightarrow y$ for: m such that $MD(m) = x \rightarrow y$.
- a priority relation $\geq: A \rightarrow 2^M \times 2^M$ is a function from agents to a transitive and reflexive relation on the powerset of the motivations containing at least the subset relation. We write $\geq_a$ for $\geq (a)$.
- the institutional facts $I \subseteq P$ are parameters.

Example 1 illustrates a highly simplified version of the running example as a multiagent system. In conceptual models used in practice as well as in the more detailed example later in this section we use meaningful names, but in Example 1 we use single letters to save space. The example describes the transaction of exchanging goods for money in terms of the motivations of the agents and the organization.

Example 1: $\langle RA, SA, X, B, D, G, AD, MD, \geq, I \rangle$ with $RA = \{\mathbf{a}, \mathbf{b}\}$, $SA = \{\mathbf{o}\}$, $P = \{p, q, r, s, t\}$, $I = \{r, s, t\}$ and $X \setminus P$, B, D, G, AD, MD and $\geq$ are implicitly given by the following table:

	a		**b**		**o**	
A	$\mathbf{o}$		$\mathbf{o}$		$-$	
X	x_1		x_2, x_3		x_4, x_5	
B	b_1	$x_2 \rightarrow p$	b_3	$x_3 \rightarrow p$		
	b_2	$x_4 \rightarrow r$	b_4	$x_4 \rightarrow r$	b_5	$x_4 \rightarrow r$
D	d_1	$\top \rightarrow p$	$-$		$-$	
G	$-$		g_1	$\top \rightarrow x_1$	g_2	$p \rightarrow x_1$
$\geq$	$d_1 > g_1 > g_2$		$g_1 > d_1 > g_2$		$g_2 > d_1 = g_1$	

This table should be read as follows. There are three agents, two real agents $\mathbf{a}$ and $\mathbf{b}$, and one socially constructed organization $\mathbf{o}$. Agent $\mathbf{a}$ is a buyer with the desire d_1 that it receives the goods ($MD(d_1) = \top \rightarrow p$), agent $\mathbf{b}$ is a seller with the goal g_1 that agent $\mathbf{a}$ pays for the goods ($MD(g_1) = \top \rightarrow x_1$), and the organization has a goal g_2 that received goods are paid for. The optimal solution is that agent $\mathbf{a}$ decides x_1 and agent $\mathbf{b}$ decides x_2. Agent $\mathbf{a}$ and $\mathbf{b}$ disagree about the fact whether it is the decision x_2 or x_3 which leads to p. All agents agree that decision x_4 leads to institutional fact r. Finally, only a fragment of the priority relation is given, because it is only given for singleton motivations, whereas it is defined over sets of motivations. It says that the agents give highest priority to their own motivations, but they are social in the sense that they incorporate the other agent's motivations.

The table can be extended to deal with more detailed motivations in the obvious way. However, the example already illustrates a drawback of using only multiagent systems to describe the example, because there is no notion of obligation, violation, sanction or contract. We therefore introduce normative multiagent systems and extend the example below.

B. Normative multiagent systems

Boella and Lesmo [12] formalize the relation between multiagent systems and normative systems by attributing mental states to agents as well as to normative systems. Thus, a normative system can metaphorically be considered as a socially constructed agent. The use of the agent metaphor may be seen as an instance of Dennett's *intentional stance* [23] and is inspired by the interpretation of normative *multiagent* systems as dynamic social orders. According to Castelfranchi [24], a social order is a pattern of interactions among interfering agents "such that it allows the satisfaction of the interests of some agent", like a shared goal or a value that is good for (almost) all the agents. For example, the interests may include the avoidance of accidents or a fair society. In this approach the obligations of the agents can be formalized as desires or goals of the normative system. This representation may be paraphrased as "your wish is my command", because the desires or wishes of the normative system are the obligations or commands of the real agents. The agents attribute to the normative system also the ability to autonomously enforce the conformity of the agents to the norms by means of sanctions.

The advantage of the agent metaphor is that game theory can be used to describe the interaction among agents and the normative system in terms of games. For example, an agent considers whether its actions will lead to a reaction of the normative system such as sanctions. An agent can understand when it can evade sanctions by for example ensuring that the normative system does not observe its behavior, or by bribing the system. Moreover, a legislator can play a game with the normative system 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. In Section V we give an example of a game in contracting.

A normative multiagent system contains a normative system $\mathbf{o} \in SA$, which we formalize as a socially constructed agent. Moreover, it contains a norm (violation) description $V : A \times Lit(X) \rightarrow X_{\mathbf{o}} \cup P$, a function from agents and literals to the decision variables of the normative system together with the parameters. We write $V_a(x)$ for the variable representing that $x \in Lit(X)$ is a violation by agent $a \in A$.

Definition 3: A normative multiagent system is a tuple $\langle RA, SA, X, B, D, G, AD, MD, \geq, I, \mathbf{o}, V \rangle$ extending a multiagent system with:

- a socially constructed agent $\mathbf{o} \in SA$ representing the normative system, such that for all real agents $a \in RA$, we have $\mathbf{o} \in A_a$.
- a (partial) function $V : A \times Lit(X) \rightarrow X_{\mathbf{o}} \cup P$ from agents and literals to the decision variables of the normative system and the parameters. We write $V_a(x)$ for $V(a, x)$.

Example 2 (Continued): The socially constructed agent $\mathbf{o}$ in Example 1 is now interpreted as an organization constituting the normative system. Moreover, assume that V is defined as follows. $V_{\mathbf{a}}(\neg x_1) = s$, $V_{\mathbf{b}}(\neg p) = t$, and $V_x(y) =$ *undefined* for all other values of x and y. Thus, if s is the case, then $\neg x_1$ is recognized as a violation of agent $\mathbf{a}$, and if t is the case, then $\neg p$ is recognized as a violation of agent $\mathbf{b}$.

C. Counts-as conditionals and self-modification

Socially constructed agents like normative systems and organizations are able to change themselves to adapt to new situations. Self modifying normative multiagent systems contain additions to the agent description, also known as expansions, which are defined as $+ : A \times (B \cup D \cup G) \rightarrow X$, i.e., as mappings from mental attitudes to propositional variables for each agent. Deletions (also known as contractions) can be defined analogously, and revisions can be defined as a combination of a deletion and an addition. In this paper we do not consider the formalization of deletion and revision, nor any other update to the normative multiagent system, to keep the formal machinery to a minimum.

Since institutional facts I, and among them the additions to the mental state attributed to the normative system, are parameters not directly controlled by the agent, we use belief rules of the normative system to express how they can be made true. Rules concerning beliefs about institutional facts are called constitutive rules and represent the "counts-as" relations (*CA*) introduced by Searle [22] (see Section II). We do not define *CA* in terms of the other concepts, but we include *CA* in the extended normative multiagent systems, to cater for the possibility that there are also belief rules of the normative system not being counts-as conditionals, but still implying institutional facts.

Definition 4 (SNMAS): A self modifying normative multiagent system is represented by a tuple $\langle RA, SA, X, B, D, G, AD, MD, \geq, I, \mathbf{o}, V, +, CA\rangle$ extending a normative multiagent system with:

- Additions are a partial function $+ : A \times (B \cup D \cup G) \rightarrow X$, such that for all $m \in B \cup M$, if $+_{\mathbf{o}}(m)$ is defined, then $+_{\mathbf{o}}(m) \in I$. We write $+_a(m)$ for $+(a, m)$.
- Counts-as conditionals $CA \subseteq B_{\mathbf{o}}$ or constitutive norms are beliefs of the normative system $\mathbf{o}$, such that constitutive rules $CR = MD(CA)$ are rules whose heads are institutional facts in I.
- Constitutive norms with legal effects (*LN*) are the beliefs of the normative system of which the heads of the rules $LR = MD(LN)$ are additions.

The update of a *SNMAS* by a set of literals $L \subseteq Lit(X)$ is $AD'_a = AD_a \cup \{m \mid +_a(m) \in L\}$.

Definition 5 (Counts-as): Given a self modifying normative multiagent system *SNMAS* as $\langle RA, SA, X, B, D, G, AD, MD, \geq, I, \mathbf{o}, V, +, CA\rangle$, we write $SNMAS \models \textit{counts-as}_{\mathbf{o}}(Y, p)$ if $Y \subseteq Lit(X)$, $p \in I$ and $\exists m \in CA$ such that $MD(m) = Y \rightarrow p$.

The running example illustrates counts-as conditionals.

Example 3 (Continued): If $CA = \{b_5\}$, then we have $SNMAS \models \textit{counts-as}_{\mathbf{o}}(x_4, r)$. Moreover, if there is a $b_6 \in B$ not associated with any agent, with $MD(b_6) = \neg p \rightarrow V_{\mathbf{b}}(\neg p)$, and $+_{\mathbf{o}}(b_6) = r$, $+_x(y) = \textit{undefined}$ for other values of x and y, then $b_5 \in LN$ is a constitutive norm with legal effects, and we have $SNMAS \models \textit{counts-as}_{\mathbf{o}}(x_4, +_{\mathbf{o}}(b_6))$. If the normative system decides x_4, then it counts as the addition of the rule $\neg p \rightarrow V_{\mathbf{b}}(\neg p)$, i.e., that if the goods are not received, then it is a violation of agent $\mathbf{b}$.

D. Obligations

Before we can define contracts, we must introduce obligations. We consider obligations of agent **a** only, which are defined in terms of goals and desires of the addressee of the norm **a** and of a normative system **o**. Obligations of agent **b** can be defined analogously.

The definition of obligation contains several clauses. The first clause says that the obligation is in the content of the desires and in the goals of normative system **o** ("your wish is my command"). The second and third clause can be read as "the absence of x is considered as a violation". The association of obligations with violations is inspired by Anderson's reduction of deontic logic to alethic modal logic [25]. The third clause says that the normative system desires that there are no violations. The fourth and fifth clause relate violations to sanctions and assume that normative system **o** is motivated to apply sanctions only as long as there is a violation; otherwise the norm would have no effect. Finally, for the same reason, we assume in the last clause that the agent does not like the sanction.

Definition 6 (Obligation): Let $SNMAS = \langle RA, SA, X, B, D, G, AD, MD, \geq, I, \mathbf{o}, V, +, CA\rangle$ be a self modifying normative multiagent system. Agent $\mathbf{a} \in A$ is *obliged* to decide to do $x \in Lit(X_{\mathbf{a}} \cup P)$ with sanction $s \in Lit(X_{\mathbf{o}} \cup P)$ if $Y \subseteq Lit(X_{\mathbf{a}} \cup P)$ in *SNMAS*, written as $SNMAS \models O_{\mathbf{ao}}(x, s|Y)$, if and only if:

1) $Y \rightarrow x \in MD(D_{\mathbf{o}}) \cap MD(G_{\mathbf{o}})$: if normative system **o** believes Y, then it desires and has as a goal that x.
2) $Y \cup \{\sim x\} \rightarrow V_{\mathbf{a}}(\sim x) \in MD(D_{\mathbf{o}}) \cap MD(G_{\mathbf{o}})$: if normative system **o** believes Y and $\sim x$, then it has the goal and the desire $V_{\mathbf{a}}(\sim x)$: to recognize it as a violation by agent **a**.
3) $\top \rightarrow \neg V_{\mathbf{a}}(\sim x) \in MD(D_{\mathbf{o}})$: normative system **o** desires that there are no violations.
4) $Y \cup \{V_{\mathbf{a}}(\sim x)\} \rightarrow s \in MD(D_{\mathbf{o}}) \cap MD(G_{\mathbf{o}})$: if normative system **o** believes Y and decides $V_{\mathbf{a}}(\sim x)$, then it desires and has as a goal that it sanctions agent **a** with s.
5) $Y \rightarrow \sim s \in MD(D_{\mathbf{o}})$: if normative system **o** believes Y, then it desires not to sanction, $\sim s$. This desire of the normative system expresses that it only sanctions in case of violation.
6) $Y \rightarrow \sim s \in MD(D_{\mathbf{a}})$: if agent **a** believes Y, then it desires $\sim s$, which expresses that it does not like to be sanctioned.

The following example illustrates an obligation of agent **a**. To increase readability of our running example, from now on we use long names for variables. We use *pay* instead of x_1, and we write *shipped* for an institutional fact following from x_2.

Example 4 (Cont.): $SNMAS \models O_{\mathbf{ao}}(\textit{pay}, \textit{san}|\textit{shipped})$, agent **a** is obliged to pay ($\textit{pay} \in X_{\mathbf{a}}$) in case the requested good has been shipped ($\textit{shipped} \in I$) or else it is sanctioned with $\textit{san} \in Lit(X_{\mathbf{o}})$, if:

g_3, d_2 $\quad \textit{shipped} \rightarrow \textit{pay} \in G_{\mathbf{o}}, D_{\mathbf{o}}$
g_4, d_3 $\quad \textit{shipped} \wedge \neg\textit{pay} \rightarrow V_{\mathbf{a}}(\neg\textit{pay}) \in G_{\mathbf{o}}, D_{\mathbf{o}}$
d_4 $\quad \top \rightarrow \neg V_{\mathbf{a}}(\neg\textit{pay}) \in D_{\mathbf{o}}$
g_5, d_5 $\quad \textit{shipped} \wedge V_{\mathbf{a}}(\neg\textit{pay}) \rightarrow \textit{san} \in G_{\mathbf{o}}, D_{\mathbf{o}}$
d_6, d_7 $\quad \textit{shipped} \rightarrow \neg\textit{san} \in D_{\mathbf{o}}, D_{\mathbf{a}}$

E. Contracts

The creation of a contract is represented by an institutional fact $c \in I$. The creation of the contract c is introduced as an intermediary between the agreement and its legal effects, because it allows decoupling the conditions of the creation of the institutional facts from additions of its legal effects. For example, consider a contract c created by the signatures of two agents. Using the decoupling a new way of creating the contract by means of an electronic signature can be specified, maintaining the same rules specifying its legal effects. Since a contract counts as several additions, c works as an abstraction. The contract unifies all its legal effects, rather than connecting the signatures of the agents directly with the additions.

In our model contracts change the norms of a normative system or organization according to what is specified by the normative systems itself, because they are part of the beliefs attributed to the system. A contract is created only if there is some fact – either a brute fact in the world or another institutional fact – counting as c for the normative system **o**. The effect of creating a contract is to modify the mental attitudes of the normative system. Usually, it adds some rules to the beliefs $B_{\mathbf{o}}$, the desires $D_{\mathbf{o}}$, or the goals $G_{\mathbf{o}}$ by an addition $+_{\mathbf{o}}(m)$ where $m \in B \cup M$. The additions are institutional facts: they are made true only if the normative system **o** believes that they are made true by the creation of the contract. E.g., $b \in \mathit{CA} \subseteq B_{\mathbf{o}}$ such that $\mathit{MD}(b) = c \rightarrow +_{\mathbf{o}}(m)$ is a constitutive rule read as "c counts as the addition $+_{\mathbf{o}}(m)$." In summary, a contract is defined as follows.

Definition 7 (Contract): Given a self modifying system $\langle \mathit{RA}, \mathit{SA}, X, B, D, G, \mathit{AD}, \mathit{MD}, \geq, I, \mathbf{o}, V, +, \mathit{CA}\rangle$ called *SNMAS*, we write $\mathit{SNMAS} \models \mathit{contract}(c, E|Y)$ where $E \subseteq \{i \in I \mid \exists m : i = +(m)\}$ and $Y \subseteq \mathit{Lit}(X)$ if:

1) $c \in I$ is an institutional fact representing that the contract has been created.
2) $\mathit{SNMAS} \models \mathit{counts\text{-}as}_{\mathbf{o}}(Y, c)$: Y counts as the creation of the contract.
3) $\mathit{SNMAS} \models \mathit{counts\text{-}as}_{\mathbf{o}}(c, +_{\mathbf{o}}(m))$ for each $+_{\mathbf{o}}(m)$ in E: the creation of the contract counts as the set of additions E.

Our running example about contracts in trade organizations illustrates how a contract can create the obligation to pay for shipped goods. The contract is created by a signature of agent **a**, and has as legal effects three goals and desires.

Example 5: [Continued] Assume that agent **a** signing the contract, represented by $\mathit{sign} \in X_{\mathbf{a}}$, is a sufficient condition for the contract $c \in I$ to be created. Moreover, assume that *SNMAS* contains d_4, d_6 and d_7 of Example 4, representing that agent **a** does not like to be sanctioned and organization **o** does not like violations and sanctions. Finally, assume that the creation c of the contract achieves the following effects on the mental attitudes of the organization **o**. It adds the goal and desire that shipped goods are paid for (g_3, d_2), the goal and desire to consider the lack of payment for shipped goods as a violation (g_4, d_3), and the goal and desire to sanction violations (g_5, d_5).

Under these assumptions, the creation of the obligation to pay for shipped goods as part of the contract created by the signature is represented by $\mathit{SNMAS} \models \mathit{contract}(c, E|\mathit{sign})$, when the following holds.

$E = \{+_{\mathbf{o}}(g_3), +_{\mathbf{o}}(d_2), +_{\mathbf{o}}(g_4), +_{\mathbf{o}}(d_3), +_{\mathbf{o}}(g_5), +_{\mathbf{o}}(d_5)\}$,
$\mathit{SNMAS} \models \mathit{counts\text{-}as}_{\mathbf{o}}(\mathit{sign}, c)$ (b_7 $\mathit{sign} \rightarrow c \in B_{\mathbf{o}}$) and,
$\mathit{SNMAS} \models \mathit{counts\text{-}as}_{\mathbf{o}}(c, +_{\mathbf{o}}(m))$ for each $+_{\mathbf{o}}(m)$ in E:

$$\begin{aligned} \{ \quad & b_8\ c \rightarrow +_{\mathbf{o}}(g_3), \quad b_9\ c \rightarrow +_{\mathbf{o}}(d_2), \\ & b_{10}\ c \rightarrow +_{\mathbf{o}}(g_4), \quad b_{11}\ c \rightarrow +_{\mathbf{o}}(d_3), \\ & b_{12}\ c \rightarrow +_{\mathbf{o}}(g_5), \quad b_{13}\ c \rightarrow +_{\mathbf{o}}(d_5)\ \} \subseteq B_{\mathbf{o}} \end{aligned}$$

The following example illustrates that the contract can create a counts-as conditional specifying an institutional fact to be used in the interaction. The contract specifies that a document called bill-of-lading counts as the institutional fact that the goods have been shipped.

Example 6 (Continued): Assume that the contract specifies that the fact that the good has been shipped is an institutional fact $\mathit{shipped} \in I$, which holds if there is some document like the so-called bill of lading ($\mathit{bill} \in P$) issued by a third party [11]. With this addition of the contract we have $\mathit{SNMAS} \models \mathit{contract}(c, E \cup \{+_{\mathbf{o}}(b_{14}\}|\mathit{sign})$: the constitutive rule b_{15} $c \rightarrow +_{\mathbf{o}}(b_{14}) \in B_{\mathbf{o}}$ creates another constitutive rule $\mathit{MD}(b_{14}) = \mathit{bill} \rightarrow \mathit{shipped}$, which is added to the beliefs of the organization **o** by the addition $+_{\mathbf{o}}(b_{14})$ as a consequence of the contract c.

We could also add the other agents to the example, such as the trusted third party in the above example, or the shipper and the banks. Moreover, due to the uniform representation of facts and the creations of norms, these constitutive norms can be nested to formalize arbitrarily complex creations. As Searle [22] observes, this nesting of counts-as conditionals leads to the complexity of social reality in which we live. For example, Example 5 and 6 illustrates that constitutive rules created by contracts can eventually introduce new obligations and new constitutive rules. In this way a contract can specify how new obligations may arise during the interaction of the agents. Likewise, contract frames can be defined as contracts describing the creation of other contracts.

Example 7 (Continued): Assume that if an agent does not pay the fee for a shipped good, it is obliged to pay a double sum of money (pay_2): $O_{\mathbf{ao}}(\mathit{pay}_2, \mathit{san}_2 \mid \top)$ [1]. This obligation is not a preexisting conditional obligation: it is created as a legal consequence of an event, the sanction *san* for not having paid the fee. The sanction *san*, in this case, rather than being a direct punishment for agent **a**, counts as the action of creating a second obligation. Note that this obligation does not exist until the normative system recognizes a violation and applies the sanction *san*. This part of the contract is thus represented by the constitutive rules creating further constitutive rules about the creation of goals and desires (where $\mathit{san}_2 \in X_{\mathbf{o}}$ is a sanction both feared by agent **a** and not desired by organization **o**). E.g., some of the clauses of the obligation are as follows.

b_{16} $c \rightarrow +_{\mathbf{o}}(b_{19}\ \mathit{san} \rightarrow +_{\mathbf{o}}(g_6\ \top \rightarrow \mathit{pay}_2))$
b_{17} $c \rightarrow +_{\mathbf{o}}(b_{20}\ \mathit{san} \rightarrow +_{\mathbf{o}}(g_7\ \neg\mathit{pay}_2 \rightarrow V_{\mathbf{a}}(\neg\mathit{pay}_2)))$,
b_{18} $c \rightarrow +_{\mathbf{o}}(b_{21}\ \mathit{san} \rightarrow +_{\mathbf{o}}(g_8\ V_{\mathbf{a}}(\neg\mathit{pay}_2) \rightarrow \mathit{san}_2))$

Due to space limitations, in the following we restrict ourselves to the creation of the obligation in Example 5 and the creation of the counts-as conditional in Example 6.

V. Games with contracts

In this section we define violation games in which an agent recursively models the normative system. Violation games are the simplest and most frequent kind of games and involve only two agents. Negotiation games in which an agent models another agent recursively modelling the normative system, can be defined in our conceptual model of normative multiagent systems too. More complex games like negotiation games extend violation games with additional levels of recursion.

A. Recursive modeling

The game theory is based on an agent decision model that uses belief rules to calculate the effects of decisions, and desire and goal rules to evaluate decisions. This is also where the updates of the rules are incorporated. Since we only consider rule additions and no other updates of the normative multiagent system, we assume that all relevant beliefs, desires and goals together with their priorities are already present in the system, such that we only have to adapt the agent description *AD*.

To define the effects of decisions, we first address the question how rules are applied such that the normative system is modified. Even without counts-as conditionals, there are several ways to define the application of rules. This is studied, for example, by Makinson and van der Torre using so-called input/output logics [26], where the input of a set of propositional formulas produces as output a set of propositional formulas using a set of rules expressed as pairs of propositional formulas. Here we use so-called reusable output, where the input is not necessarily part of the output, rules can be applied one after the other, and without so-called reasoning by cases. The following definition extends reusable output logic to incorporate a way to deal with additions. Instead of working with a fixed set of rules, at every step of the calculation of out^i the set of rules R^i can be updated too. The output of applying a set of rules R on a set of literals S with additions E is written as $\mathit{out}(S,R,E)$.

Definition 8 (Applying rules): The set of rule additions defined on X and I, written as $\mathit{Add}(I,X) = I \times \mathit{Rul}(X)$, is the set of all pairs of institutional facts and rules. We define $\mathit{out} : 2^{\mathit{Lit}(X)} \times 2^{\mathit{Rul}(X)} \times 2^{\mathit{Add}(I,X)} \rightarrow 2^{\mathit{Lit}(X)}$ as a function from a set of literals, a set of rules and a set of rule additions to a set of literals. $\mathit{out}(S,R,E)$ is the closure of a set of literals S under the rules R updated by rule additions E.

- $R^0(S,R,E) = R$
- $\mathit{out}^0(S,R,E) = \emptyset$
- $R^{k+1}(S,R,E) = R^k(S,R,E) \cup$ $\{r \in \mathit{Rul}(X) | (i,r) \in E$ and $i \in S \cup \mathit{out}^k(S,R,E)\}$
- $\mathit{out}^{k+1}(S,R,E) = \mathit{out}^k(S,R,E) \cup$ $\{l | L \rightarrow l \in R^k(S,R,E)$ and $L \subseteq S \cup \mathit{out}^k(S,R,E)\}$
- $\mathit{out}(S,R,E) = \cup_o^\infty \mathit{out}^k(S,R,E)$

Decisions of agents are sets of literals built from decision variables which do not imply a contradiction. The input/output logic is used in the game theory to define the closure of a decision under a set of belief rules. When calculating the effects of decisions, the rule additions are the inverse of the addition function $+$ mapped to rule descriptions. We describe the new belief rules of agent $\mathbf{a}$ as additions $E^B_{\mathbf{a}}$. We exclude the cases in which agent **a** makes a decision which normative system **o** does not hold possible, since we do not incorporate how normative system **o** revises its beliefs in case of such a surprise.

Definition 9 (Decisions): Let the set of belief additions of agent $\mathbf{a}$ be $E^B_{\mathbf{a}} = \{(i, \mathit{MD}(b)) \mid b \in B, +_{\mathbf{a}}(b) = i\}$, the belief extension of agent $\mathbf{a}$ of a set of literals $S \subseteq \mathit{Lit}(X)$ be $S \cup \mathit{out}(S, B_{\mathbf{a}}, E^B_{\mathbf{a}})$, and analogously for normative system **o**. The set of decision profiles Δ is the set of sets $\delta = \delta_{\mathbf{a}} \cup \delta_{\mathbf{o}}$ with $\delta_{\mathbf{a}} \subseteq \mathit{Lit}(X_{\mathbf{a}})$ and $\delta_{\mathbf{o}} \subseteq \mathit{Lit}(X_{\mathbf{o}})$ such that the belief extensions of agent **a** and of normative system **o** do not contain a variable and its negation.

The logic is illustrated in an extension of Example 3.

Example 8 (Continued): The effects of the beliefs of a are $\mathit{out}(B_{\mathbf{a}}, \{x_4, \neg p\}, E^B_{\mathbf{a}}) = \{r = +_{\mathbf{o}}(b_6), s = V_{\mathbf{b}}(\neg p)\}$: if the normative system decides x_4 and the goods are not received, then it is a violation.

We assume that agent **a** only considers its own motivations (and is in this sense selfish), which can be relaxed in the obvious way, and that normative system **o** considers its own motivations including the new rules. The unfulfilled motivations of decision profile $\delta = \delta_{\mathbf{a}} \cup \delta_{\mathbf{o}}$ are the motivations whose body is part of the closure of the decision under the belief rules but whose head is not. Given a decision $\delta_{\mathbf{a}}$, a decision $\delta_{\mathbf{o}}$ is optimal for normative system **o** if it minimizes the unfulfilled motivational attitudes in $D_{\mathbf{o}}$ and $G_{\mathbf{o}}$ according to the $\geq_{\mathbf{o}}$ relation. The decision of agent **a** is more complex. When agent **a** takes its decision $\delta_{\mathbf{a}}$ it minimizes its unfulfilled motivational attitudes. But when it considers these attitudes, it must not only consider its decision $\delta_{\mathbf{a}}$ and the consequences of this decision. It must consider also the decision $\delta_{\mathbf{o}}$ of the normative system **o** and its consequences, for example that it is sanctioned by normative system **o**. So agent **a** recursively considers which decision normative system **o** will take depending on its different decisions $\delta_{\mathbf{a}}$.

Definition 10 (Recursive modelling): Let $S_{\mathbf{a}}(\delta)$ and $S_{\mathbf{o}}(\delta)$ be the belief extensions of decision profile δ for agent **a** and **o** respectively, and unfulfilled motivations of agent **a** and normative system **o** defined as follows.

$$\begin{aligned} U(\delta,\mathbf{a}) = \quad & \{m \in M_{\mathbf{a}} \mid MD(m) = l_1 \wedge \ldots \wedge l_n \rightarrow l, \\ & \{l_1,\ldots,l_n\} \subseteq S_{\mathbf{a}}(\delta) \text{ and } l \notin S_{\mathbf{a}}(\delta)\} \\ U(\delta,\mathbf{o}) = \quad & \{m \in M_{\mathbf{o}} \cup \{m' \in M \mid +_{\mathbf{o}}(m') \in S_{\mathbf{o}}(\delta)\} \mid \\ & MD(m) = l_1 \wedge \ldots \wedge l_n \rightarrow l \text{ and} \\ & \{l_1,\ldots,l_n\} \subseteq S_{\mathbf{o}}(\delta) \text{ and } l \notin S_{\mathbf{o}}(\delta)\ \} \end{aligned}$$

- A decision profile $\delta_{\mathbf{a}} \cup \delta_{\mathbf{o}}$ is *optimal* for normative system **o** if and only if there is no decision of the normative system $\delta'_{\mathbf{o}}$ such that $U(\delta_{\mathbf{a}} \cup \delta_{\mathbf{o}}, \mathbf{o}) >_{\mathbf{o}} U(\delta_{\mathbf{a}} \cup \delta'_{\mathbf{o}}, \mathbf{o})$.
- A decision profile $\delta_{\mathbf{a}} \cup \delta_{\mathbf{o}}$ is optimal for agent **a** and normative system **o** if and only if it is optimal for normative system **o** and there is no decision $\delta'_{\mathbf{a}}$ such that for all decision profiles $\delta_{\mathbf{a}} \cup \delta''_{\mathbf{o}}$ and $\delta'_{\mathbf{a}} \cup \delta'_{\mathbf{o}}$ optimal for normative system **o** we have $U(\delta_{\mathbf{a}} \cup \delta''_{\mathbf{o}}, \mathbf{a}) >_{\mathbf{a}} U(\delta'_{\mathbf{a}} \cup \delta'_{\mathbf{o}}, \mathbf{a})$.

Negotiation games involving all three agents **a**, **b** and **o** can be defined analogously. For example, to find out whether agent **a** will fulfill its obligations, agent **b** has to consider his beliefs about agent **a**'s beliefs about normative system **o**, extending the violation game with another level of recursion.

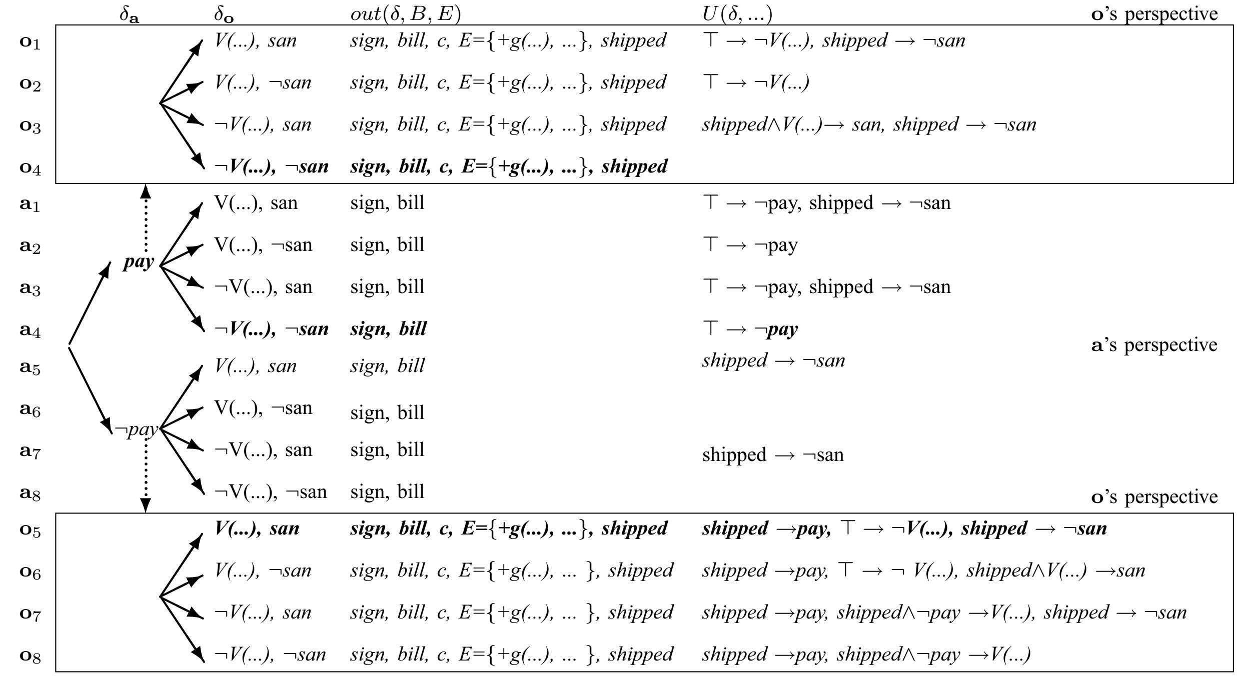
δa
δo
out(δ, B, E)
U(δ, ...)
o's perspective
o1 V(...), san sign, bill, c, E={+g(...), ...}, shipped ⊤ → ¬V(...), shipped → ¬san
o2 V(...), ¬san sign, bill, c, E={+g(...), ...}, shipped ⊤ → ¬V(...)
o3 ¬V(...), san sign, bill, c, E={+g(...), ...}, shipped shipped∧V(...)→ san, shipped → ¬san
o4 ¬V(...), ¬san sign, bill, c, E={+g(...), ...}, shipped
a1 V(...), san sign, bill ⊤ → ¬pay, shipped → ¬san
a2 V(...), ¬san sign, bill ⊤ → ¬pay
pay
a3 ¬V(...), san sign, bill ⊤ → ¬pay, shipped → ¬san
a4 ¬V(...), ¬san sign, bill ⊤ → ¬pay
a's perspective
a5 V(...), san sign, bill shipped → ¬san
a6 V(...), ¬san sign, bill
¬pay
a7 ¬V(...), san sign, bill shipped → ¬san
a8 ¬V(...), ¬san sign, bill
o's perspective
o5 V(...), san sign, bill, c, E={+g(...), ...}, shipped shipped →pay, ⊤ → ¬V(...), shipped → ¬san
o6 V(...), ¬san sign, bill, c, E={+g(...), ... }, shipped shipped →pay, ⊤ → ¬ V(...), shipped∧V(...) →san
o7 ¬V(...), san sign, bill, c, E={+g(...), ... }, shipped shipped →pay, shipped∧¬pay →V(...), shipped → ¬san
o8 ¬V(...), ¬san sign, bill, c, E={+g(...), ... }, shipped shipped →pay, shipped∧¬pay →V(...)

Fig. 2. An example of the game between agent **a** and organization **o**.

B. Example

The violation game concerns a trade contract that creates the obligation to pay for shipped goods, enforced by a sanction, and the constitutive rule saying that the bill of lading counts as the institutional fact that the goods have been shipped, as discussed in Example 5 and 6. The extensive game tree visualized in Figure 2 is based on the rules discussed in the examples thus far, together with additional rules and priorities among rules discussed below. Both agent **a** and organization **o** believe that agent **a** has already signed the contract and the bill of lading $bill \in P$ has been issued.

$\{b_{22}\ \top \rightarrow sign, b_{23}\ \top \rightarrow bill\} \subseteq B_{\mathbf{a}}$

$\{b_{24}\ \top \rightarrow sign, b_{25}\ \top \rightarrow bill\} \subseteq B_{\mathbf{o}}$

For agent **a**, not being sanctioned has a higher priority than not paying. For organization **o**, in case of conflict counting behavior as a violation has a higher priority than not doing so, and sanctioning violations has a higher priority than not doing so.

$\{shipped \rightarrow \neg san\} >_{\mathbf{a}} \{\top \rightarrow \neg pay\}$

$\{shipped \wedge \neg pay \rightarrow V_{\mathbf{a}}(\neg pay), \} >_{\mathbf{o}} \{\top \rightarrow \neg V_{\mathbf{a}}(\neg pay)\}$

$\{shipped \wedge V_{\mathbf{a}}(\neg pay) \rightarrow san\} >_{\mathbf{o}} \{shipped \rightarrow \neg san\}$

...

Figure 2 should be read as follows. The branches represent decisions of the agents. Real agent $\mathbf{a} \in RA$ has two alternatives, to pay or not to pay. Organization $\mathbf{o} \in SA$ has four alternatives, to count agent **a**'s behavior as a violation or not, and to sanction it or not. A complete path from left to right thus represents a decision profile. The leaves of the tree represent the belief extensions of the agents for the related decision profile, calculated using the input/output logic. As discussed in Examples 5 and 6, the believed consequences of the normative system contain the legal effects of the contract, represented by a set of additions. It also includes the consequences of the added constitutive norm, leading to the occurrence of *shipped* in some nodes. The rightmost column represents the unfulfilled desires and goals, calculated by considering for each relevant rule whether the body of the rule is implied by the believed effects of the decision, without the head being implied. For the normative system also the added rules of the contract are considered, occurring as the unfulfilled desires and goals in the right most column of the figure. The viewpoint of organization **o** is visualized in rectangular boxes on the bottom and the top, and the viewpoint of agent **a** is in between.

The violation games adopt the viewpoint of agent **a** deciding whether to pay its fee or not, for which it recursively models the organization $\mathbf{o} \in A_{\mathbf{a}}$. For each node we visualize the optimal decision in boldface. Consider first the viewpoint of the organization visualized in rectangular boxes. The top four lines indicate that if agent **a** pays, then it does not count it as a violation and does not sanction the agent (organization **o** prefers profile o_4 to profiles $o_1 - o_3$), and the bottom four lines indicate that if agent **a** does not pay, then it sanctions the agent (organization **o** prefers profile o_5 to profiles $o_6 - o_8$). Consequently agent **a** can choose among only profile a_4 and profile a_5, of which it chooses profile a_4. Note that the belief extensions and unfulfilled motivations of $a_1 - a_3$ and $a_6 - a_8$ do not have to be calculated to obtain this result.

The example illustrates why the definition of obligation contains six clauses. Clauses 2-6 are all necessary to derive that selfish agent **a** will fulfill the obligation. Clause 1 of the definition is not necessary in this example, but it is used to derive the same result for respectful agents internalizing the goals of the normative system as their own goals. There are many ways in which the example can be modified such that no longer can be derived that agent **a** will pay, including many subtle variants of fraud and deception.

VI. Related work

In this paper we use a qualitative game-theoretic approach different from the classical quantitative one based on utility functions, probability distributions, and equilibria analysis. Starting with the pioneering work of Herbert Simon and others in the 50s, many alternative conceptualizations and formalizations of decision making have been proposed in artificial intelligence and agent theory as a response to the ideality assumptions of classical decision and game theory. For example, utility functions have been replaced by goals and desires, probability functions by beliefs, and decision rules by a process of deliberation. See Dastani *et al.* [27] for a comparison between classical decision and game theory and alternative qualitative theories developed in artificial intelligence and agent theory.

More complex games have been defined for normative multiagent systems introduced in this paper. Due to space limitations, in this paper we only consider two stage games. In [28] we define games with multiple agents and multiple stages, but without contracts or self-modification. Based on a hierarchy, agents model the behavior of the next agent in the hierarchy. Moreover, in [29] we define norm creation games from the viewpoint of the normative system recursively modelling an agent to find out whether it will violate the new norm, which just like negotiation games extend violation games with another level of recursion.

Moreover, normative multiagent systems have been studied which are in some aspects more general than the one used in this paper. For example, in [28] we define multiple normative systems playing the role of local and global authorities. The central question concerning the interaction among normative systems is how the global authority uses global policies to control the local authorities with their local policies. The local authorities are again not forced but motivated, for which we consider a kind of nested obligations and permissions.

There are several other issues in normative multiagent systems relevant for the game-theoretic approach to contracts. In [30] we study how sanctions are negotiated among agents in the context of negotiating the distribution of obligations. Instead of defining obligations as a set of explicitly given goals and desires, we can also use an input/output logic to state that the mental attitudes logically follow from the mental states of the agents. Moreover, permissions, prohibitions and reward-based obligations can be defined in terms of motivational attitudes in a similar way. A first version of our formal model of self-modifying normative multiagent systems can be found in [31], and a preliminary formalization of our game theoretic approach to contracts can be found in [32].

There are several other proposals to formalize contracts in the area of multiagent systems. Inspired by Sandholm and Lesser [33], Teague and Sonenberg [34] discuss the impact on reputation of levelled commitment contracts, i.e., contracts where each party can decommit by paying a predeterminate penalty. While reputation is beyond the scope of this paper, our model of contracts can specify not only decommitment penalties, but also explicit procedures for the agents to withdraw from the contract by means of constitutive rules.

Pacheco and Carmo [4] and Dignum *et al.* [1] define the clauses of a contract as conditional obligations, whereas we use constitutive rules creating obligations when the contract is created or when another relevant event happens. Dignum *et al.* propose the language *LCR* for modelling contracts. They define contracts as tuples composed of agents, contract clauses, stages and interactional structure. They also give a definition of obligations in terms of violations, but they do not take a subjective perspective and they do not consider the decision problem of an agent subject to obligations.

Daskalopulu and Maibaum [35] model contracts as processes having states representing the legal relations among agents. They introduce obligations as consequences of the unfulfillment of other obligations. However, they do not consider the role of constitutive rules in contracts and the fact that violations are recognized only as an effect of the activity of the normative system.

Vasconcelos *et al.* [36], [37] and others call normative systems regulating the normative position of a multiagent system e-institutions. E-institutions establish interaction conventions and social consequences of actions, and enforce the satisfaction of commitments. These concepts could receive a foundation in terms of legal institutions as we propose in this paper.

Besides deontic logic, other formalisms have been proposed to model contracts, for example, Petri nets [38] and state transition graphs [2].

In the context of e-business Hanson and Milosevic [3] propose a framework for dealing with several aspects of contracts, from negotiation to validation, from monitoring the execution to recovering from violations. As discussed in Section I, we agree with Milosevic and Dromey [18] that a specification of an organization given by contract differs significantly from a computational specification in that the compliance cannot be taken for granted. Even if Milosevic and Dromey [18] are aware that contracts are not exhausted by obligations and permissions, they define them only in terms of deontic concepts. Our view of contracts as legal institutions emphasizes the role of constitutive norms besides regulative ones. Hanson and Milosevic [3] define contract negotiation as a process of making and adjusting offers among potential agents willing to be involved in an economic transaction - until an agreement is reached or the process is terminated without an agreement. We adopt a similar perspective on contract negotiation and we focus our work on the problem of evaluating offers according to the commitments made by the agents and the procedures created by constitutive norms.

Besides works on modelling contracts, a large literature is devoted to the problem of the automated negotiation of contracts, and, in particular, to provide protocols of interaction which are self enforcing and have desirable properties, like, for example, the maximization of the sum of the outcomes of the agents [33], [39], [40]. This topic is beyond the scope of this paper focusing on reasoning about contracts rather than on searching the space of possible contracts. Many works assume that the execution of the contract is carried out without violations. This assumption is too strong, as noticed by Milosevic and Dromey [18].

VII. Conclusions and future work

The role of norms and contracts in the interaction of agents is a major aspect of e-commerce and e-trading systems. Agents must reason about the fulfillment of norms, the possible violations and what to do to repair such violations. However, classical game and decision theory presume a fixed set of interaction possibilities. Since contracts can be used to change the interaction possibilities, norms and contracts pose an important challenge to the game theoretic analysis of agent interactions. In this paper we explain and formalize how agents can modify the behavior of normative systems via contracts.

Our approach to modify the behavior of normative systems by means of contracts is based on constitutive rules changing the normative system, using the distinction between regulative and constitutive norms developed by Searle [8], and the concept of legal institutions as developed by Ruiter [9]. Legal effects of actions of the members of a legal system are complex and contracts do not concern only the regulative aspects of a legislation, i.e., the rules of behavior specified by obligations, or the constitutive part of it, i.e., the rules introducing institutional facts such as bidding in an auction, but contracts are systems of regulative and constitutive rules that provide frameworks for social action *within larger rule-governed settings*. Therefore contracts as legal institutions bring with them constitutive norms creating not only institutional facts, but creating also new regulative and constitutive norms. In this way, it is possible to specify in a contract new procedures for the interaction among agents.

The game theoretic approach to contracts we develop in this paper builds on the notion of strategic interaction developed by Goffman. Inspired by Goffman we adopt the idea that agents can play games with normative systems just as they play games with other agents, thus modeling the normative system as a socially constructed agent, and inspired by Searle we represent the regulative norms and constitutive norms as motivational and informational attitudes of this socially constructed agent. This leads to a relatively simple model of a complex social phenomena, which can be further developed in agent theory to develop e-commerce and e-trading systems. Moreover, we introduce a qualitative game theory based on recursive modelling.

Our formalization distinguishes between the structure of normative multiagent systems and games played in it. Starting from Boella and Lesmo [12]'s observation that a normative system behaves like an agent monitoring and sanctioning violations, we develop an ontology of social reality, including not only normative systems but also groups and organizations. This ontology is based on the idea that social entities can be modelled as agents which are attributed mental attitudes. The metaphor allows us to define regulative rules as goals of the normative system and constitutive rules as its beliefs. We introduce self-modifying normative multiagent systems to define constitutive norms with legal effects, which update the beliefs and motivations of the normative system. In these normative multiagent systems, obligations are defined using six clauses, distinguishing among the motivational and regulative constituents of obligations. The creation of contracts is represented by an institutional fact that works as an abstraction between the agreement and its legal effects. Contracts seen as legal institutions are based on a tight integration of constitutive and regulative rules.

Moreover, we develop a qualitative game theory where agents are allowed to make contracts to change their normative positions, and we explain and show how agents can use the game theory. By using recursive modelling and applying the agent metaphor to normative systems we define violation games among an agent and the normative system in which the agent predicts the behavior of normative systems, and we use an example from international trade to illustrate how agents can use the game theory. Negotiation games involving for example all three agents **a**, **b** and **o** can be defined analogously. For example, assume that agent **b** also has to sign the contract before it comes into force, and it is deliberating whether to sign it or not. One of the crucial factors in signing is whether agent **b** believes that agent **a** will fulfill its obligations, or violate them. Agent **b** therefore recursively models agent **a**. However, to find out whether agent **a** will fulfill its obligations, agent **b** has to consider his beliefs about agent **a**'s beliefs about organization **o**, as detailed in the violation game discussed in Section V-B. The negotiation game therefore extends the violation game with another level of recursion.

The qualitative game theory is not restricted to selfish agents minimizing their sanctions, based on clauses 2-6 of the definition of obligation, but it can also be used for respectful agents internalizing the goals of the normative system as their own goals, based on clause 1. The example also suggests many ways in which the example can be modified such that no longer can be derived that agent **a** will pay, including many subtle variants of fraud and deception.

In this paper we use normative system and organization interchangeably, though there are several additional issues to be studied in organizations. For example, as Milosevic and Dromey [18] suggest, contracts are strictly related to roles. Contracts are used for assigning roles: they create the obligations of the holder of a role starting from the description of the role. The notion of role allows also to structure the normative system in various types of agents, like those with the task of monitoring violations and applying sanctions. Analogously new roles can be added to normative systems regulating the contract negotiation process to act as mediators or coordinators of contracts. Legal institutions, like contracts, can even create new normative systems, for example, contracts creating organizations (such as the legislation on societies). It is thus necessary that such contracts are able to specify obligations and permissions about obligations and permissions created by an organizations.

The use of normative multiagent systems is not restricted to contract negotiation. Many theories and applications of multiagent systems such as electronic commerce, virtual communities, theories of fraud and deception, of trust dynamics and reputation, secure knowledge management, *et cetera*, can fruitfully employ the notion of a normative system regulating an agent society. Each of these applications comes with its own characteristic properties, which are subject of further research.

References

[1] V. Dignum, J.-J. Meyer, and H. Weigand, "Towards an organizational-oriented model for agent societies using contracts," in *Procs. of AAMAS'02*. ACM Press, 2002, pp. 694–695.

[2] C. Dellarocas and M. Klein, "Contractual agent societies: Negotiated shared context and social control in open multiagent systems," in *Social Order in MAS*, R. Conte and C. Dellarocas, Eds. Kluwer, 2001, pp. 113–134.

[3] J. Hanson and Z. Milosevic, "Conversation-oriented protocols for contract negotiations," in *Procs. of EDOC'03*, Brisbane (AU), 2003, pp. 40–49.

[4] O. Pacheco and J. Carmo, "A role based model of normative specification of organized collective agency and agents interaction," *Autonomous Agents and Multiagent Systems*, vol. 6, pp. 145–184, 2003.

[5] P. Marsh, *Contract Negotiation Handbook*. Aldershot: Gower, 1984.

[6] C. Alchourrón and E. Bulygin, *Normative Systems*. Wien: Springer-Verlag, 1971.

[7] A. Jones and J. Carmo, "Deontic logic and contrary-to-duties," in *Handbook of Philosophical Logic*, D. Gabbay and F. Guenthner, Eds. Kluwer, 2001, pp. 203–279.

[8] J. Searle, *Speech Acts: an Essay in the Philosophy of Language*. Cambridge (UK): Cambridge University Press, 1969.

[9] D. Ruiter, "A basic classification of legal institutions," *Ratio Juris*, vol. 10(4), pp. 357–371, 1997.

[10] H. Hart, *The Concept of Law*. Oxford: Clarendon Press, 1961.

[11] J. Gordijn and Y.-H. Tan, "A design methodology for trust and value exchanges in business models," in *Procs. of BLED Conference*, 2003, pp. 423–432.

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

[13] E. Goffman, *Strategic Interaction*. Oxford: Basil Blackwell, 1970.

[14] A. Newell, "The knowledge level," *Artificial Intelligence*, vol. 18, pp. 87–127, 1982.

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

[16] P. J. Gmytrasiewicz and E. H. Durfee, "Formalization of recursive modeling," in *Procs. of ICMAS'95*, 1995, pp. 125–132.

[17] V. Kartseva, J. Gordijn, and Y.-H. Tan, "Designing control mechanisms for network organisations," in *Procs. of ICEC'04*. IEEE Press, 2004.

[18] Z. Milosevic and G. Dromey, "On expressing and monitoring behaviour in contracts," in *Procs. of EDOC'02*, Lausanne (CH), 2002, pp. 3–14.

[19] Y. Shoham and M. Tennenholtz, "On the emergence of social conventions: Modeling, analysis and simulations," *Artificial Intelligence*, vol. 94, no. 1–2, pp. 139–166, 1997.

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

[21] J. Lang, L. van der Torre, and E. Weydert, "Utilitarian desires," *Autonomous Agents and Multiagent Systems*, pp. 329–363, 2002.

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

[23] D. C. Dennett, *The Intentional Stance*. Cambridge, MA: The MIT Press, 1987.

[24] C. Castelfranchi, "Engineering social order," in *Procs. of ESAW'00*. Berlin: Springer Verlag, 2000, pp. 1–18.

[25] A. Anderson, "The logic of norms," *Logic et analyse*, vol. 2, 1958.

[26] D. Makinson and L. van der Torre, "Input-output logics," *Journal of Philosophical Logic*, vol. 29, pp. 383–408, 2000.

[27] M. Dastani, J. Hulstijn, and L. van der Torre, "How to decide what to do?" *European Journal of Operational Research*, vol. 160(3), pp. 762–784, 2005.

[28] G. Boella and L. van der Torre, "Local policies for the control of virtual communities," in *Procs. of IEEE/WIC WI'03*. IEEE Press, 2003, pp. 161–167.

[29] ——, "Rational norm creation: Attributing mental attitudes to normative systems, part 2," in *Procs. of ICAIL'03*. Edinburgh: ACM Press, 2003, pp. 81–82.

[30] ——, "The distribution of obligations by negotiation among autonomous agents," in *Procs. of ECAI'04*. IOS Press, 2004, pp. 13–17.

[31] ——, "Regulative and constitutive norms in normative multiagent systems," in *Procs. of KR'04*. AAAI Press, 2004, pp. 255–265.

[32] ——, "Contracts as legal institutions in organizations of autonomous agents," in *Procs. of AAMAS'04*. ACM Press, 2004, pp. 948–955.

[33] T. Sandholm and V. Lesser, "Issues in automated negotiation and electronic commerce: Extending the contract net framework," in *Procs. of ICMAS'95*, San Francisco (CA), 1995, pp. 328–335.

[34] V. Teague and L. Sonenberg, "Investigating commitment flexibility in multiagent contracts," in *Game Theory and Decision Theory in Agent-Based Systems*, S. Parsons, P. Gymtrasiewicz, and M. Wooldridge, Eds. Kluwer, 2002, pp. 267–292.

[35] A. Daskalopulu and T. Maibaum, "Towards electronic contract performance," in *Procs. of Legal Information Systems Applications*, 2001, pp. 771–777.

[36] M. Esteva, J. Rodriguez-Aguilar, C. Sierra, and W. Vasconcelos, "Verifying norm consistency in electronic institutions," in *Procs. of Workshop on Agent Organizations at AAAI'04*, San Jose (CA), 2004.

[37] W. Vasconcelos, J. Sabater, C. Sierra, and J. Querol, "Skeleton-based agent development for electronic institutions," in *Procs. of AAMAS 2002*. ACM press, 2002, pp. 696–703.

[38] R. Lee, "Documentary Petri nets: A modeling representation for electronic trade procedures," in *Business Process Management, LNCS 1806*. Berlin: Springer Verlag, 2000, pp. 359–375.

[39] D. Reeves, M. Wellman, and B. Grosof, "Automated negotiation from declarative contract descriptions," *Computational Intelligence*, vol. 18(4), pp. 482–500, 2002.

[40] J. S. Rosenschein and G. Zlotkin, *Rules of Encounter. Designing Conventions for Automated Negotiation among Computers*. Cambridge, MA: MIT Press, 1994.

Guido Boella received the PhD degree at the University of Torino in 2000.
He is currently a researcher at the Department of Computer Science and the Center for Cognitive Science of the University of Torino. His research interests include multiagent systems, in particular, normative systems, institutions and roles using qualitative decision theory. He organized the first workshops on normative multiagent systems, on coordination and organization, and the AAAI Fall Symposium on roles.

Leendert van der Torre received the PhD degree from Erasmus University Rotterdam in 1997.
He is currently researcher at CWI Amsterdam and lecturer at Delft University of Technology. He has developed the so-called input/output logics and the BOID agent architecture. He is coordinator of the BOID project and of the CWI node of the ArchiMate project. His current research interests include deontic logic, qualitative game theory and coordination in normative multiagent systems.