Contracts in Multiagent Systems:
the Legal Institution Perspective

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

[1]Dipartimento di Informatica - Università di Torino - Italy, guido@di.unito.it
[2]CWI - Amsterdam - The Netherlands, torre@cwi.nl

Abstract. In this paper we address the problem of how the autonomy of agents in an organization can be enhanced by means of contracts. Contracts are modelled as legal institutions: systems of legal rules which allow to change the regulative and constitutive rules of an organization. The methodology we use is to attribute to organizations mental attitudes, beliefs, desires and goals, and to take into account their behavior by using recursive modelling.

1 Introduction

One of the main challenges in multi-agent societies is the coordination of the autonomous agents. Coordination can be achieved finding a trade off between a bottom-up view and a top-down view of the problem. In the former, the MAS is an aggregation of autonomous agents interacting with each other, where their emergent behavior is not necessarily the desired one. In the latter, the system's objectives are achieved without requiring specific agent's internal design, by means of organizational design together with roles and norms as incentives for cooperation.

As Dignum *et al.* [12] note, however, the interaction structure of the organization should be not be completely fixed in advance. The autonomy of the agents should be preserved even if within limits. For this reason, some approaches like [10, 12, 19] introduce the possibility for agents to stipulate contracts. A contract can be defined as a statement of intent that regulates behavior among organizations and individuals. Contracts have been proposed as means to make explicit the way agents can change the interaction with and within the society: they create obligations, permissions and new possibilities of interactions. From a contractual perspective, organizations can be seen as the possible sets of agreements for satisfying the diverse interests of self interested individuals [10].

Social order, thus, emerges from the negotiation of contracts about the rights and duties of participants, rather than being given in advance. But the organization itself specifies the possible contracts and enforces the obligations created by them as they were issued by the organization itself. As Ruiter [20] shows, however, from the legal point of view, legal effects of actions of the members of a legal system (as an organization is) are a difficult problem. Contracts do not concern only the regulative aspects of a legislation (i.e., the rules of behavior specified in obligations), or the constitutive part of it (i.e., the rules introducing institutional facts such bidding in an auction). Rather, contracts are *legal institutions*: "systems of [regulative and constitutive] rules that provide

frameworks for social action within larger rule-governed settings" [20]; in our case the larger setting is represented by organizations.

This systemic view of legal institutions emerged only recently in legal studies [20], since legal positivism [15] mainly focused on the regulative aspects of law and its justification. For this reason is necessary to address the problem of contracts being aware of the peculiarities of legal institutions.

The research question of this paper is: how can be legal institutions, like contracts, be formalized? and, as subquestions, how can agents modify the behavior of the organization via contracts? Which games can agents play when they are allowed to make contracts?

As methodology we use the agent metaphor: we attribute to organizations mental attitudes, beliefs, desires and goals, and we take into account their behavior by using recursive modelling [13]. We apply to organizations the methodology we adopted for social entities like groups, virtual communities [4] and normative multiagent systems [3, 7, 6].

In the next section we discuss constitutive rules and how legal institutions are created. In Section 3 we discuss the conceptual model, with the definition of obligations and contracts. In Section 4 we present the games which can be played with contracts, together with a detailed example. Related work and summary close the paper.

2 Legal institutions

Normative multiagent systems, like organizations, are "sets 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" (Jones and Carmo [16]).

In [3] we formalize the relation between multiagent systems and normative systems by attributing mental states to agents as well as to normative systems, as proposed by Boella and Lesmo [2]. The agent metaphor may be seen as an instance of Dennett's *intentional stance* [11] and is inspired by the interpretation of normative *multiagent* systems as dynamic social orders. According to Castelfranchi [8], a social order is a pattern of interactions among interfering agents "such that it allows the satisfaction of the interests of some agent". These interests can be a shared goal, a value that is good for everybody or for most of the members. But the agents attribute to the normative system, besides goals, also the ability to autonomously enforce the conformity of the agents to the norms by means of sanctions. In this approach the 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" because the desires or wishes of the normative agent are the obligations or commands of the other agents.

Most formalizations of normative systems, however, including [3], identify norms with obligations, prohibitions and permissions. This is not sufficient in complex normative multiagent systems for the following three reasons. First of all, these norms, called regulative norms, specify all the conditions when they are applicable. It would be more economic that regulative norms could factor out particular cases and could refer to

more abstract concepts only. Hence, the normative system should include mechanisms to introduce new legal categories of abstract entities for classifying possible states of affairs. Second, the dynamics of the social order is due to the fact that the normative system evolves over time by introducing new norms and abrogating outdated ones. So the normative system itself must specify how it can be changed by introducing new regulative norms, new legal categories and by whom the changes can be done. Third, the dynamics of a normative system includes the possibility that not only new norms are introduced by the legislators, but also that ordinary agents create new obligations and permissions concerning specific agents. This feature is fundamental to preserve the autonomy of agents inside an organization. In particular, it allows modelling contracts which introduce new normative relations among agents, like the duty to pay a fee for a service.

We therefore introduce a formal framework for the construction of normative multi-agent systems, based on Searle's notion of the construction of social reality. Searle [21] argues that there is a distinction between two types of rules, a distinction which also holds for formal rules like those composing normative systems:

> 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 ([21], p. 131).

For Searle, institutional facts like marriage, money and private property 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".

Like we formalize obligations in terms of desires and goals, in the next section, we formalize the constitutive rules as belief rules of the normative agent. E.g., consider a society which believes that a field fenced by an agent counts as the fact that the field is property of that agent. The fence is a physical "brute" fact, while being a property is an institutional fact attributed to the beliefs of the normative system. Regulative norms which forbid trespassing refer to the abstract concept of property rather than to fenced fields. As the system evolves, new cases are added to the notion of property, without changing the regulative norms about property. E.g., if a field is inherited, then it is property of the heir.

Searle's analysis of constitutive rules has focused mainly on the attribution of a new functional status to entities, as in the examples above: marriages, money, property. Searle's idea is that constitutive rules "create the possibility or define that activity". We believe, however, that 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 both the creation of legal categories and the evolution of the system: the normative system itself specifies by means of constitutive rules (included in its belief rules) how its beliefs, desires and goals can be changed, who can change them, and the limits of the possible changes. In this way, complex normative systems achieve a legal regime that includes rules conferring legal powers on certain participants: an agent is turned on a "private legislator" (Hart, [15]): "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 [20]).

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 conditions of an obligation are 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 rule: it expresses the fact that when a legal institution of marriage comes into existence (say between Amy and Bob) only then the obligation that the spouses (Amy and Bob) take care and support their children is created. The same happens with the legal institution of contracts: when a contract comes into existence it creates obligations for the parties, i.e., new regulative rules 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 parties to regulate a juridical relationship about valuables *ex* art. 1321).

Moreover, contracts as legal institutions bring with them not only new regulative rules, but also constitutive ones which create new institutional facts and also new obligations; in this way, it is possible to specify in a contract new procedures for the interaction between parties, for specifying the evolution of the contract and how new obligations are created later. As Dignum *et al.* [12] notice, a contract specifies the events that alter the status of the contract. It is necessary to specify an interaction structure which indicates the possibility 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 some payment deadline is not respected, the agent is obliged to pay a double fee. Since we model contracts as legal institutions, we are now aware that this rule is not a conditional obligation: it is an obligation created by some 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, i.e., a normative system inside the main normative system: as a normative system it specifies who has the power to introduce obligations.

3 The conceptual model

In order to provide a formalization of contracts as legal institutions in organizations we first delineate the conceptual model we adopt.

First of all, the structural concepts and their relations. We have to describe the different aspects of the world and the relationships among them. We therefore introduce a set of propositional variables X and we extend it to consider also negative states of affairs: $L(X) = X \cup \{\neg x \mid x \in X\}$. Moreover, for $x \in X$ we write $\sim x$ for $\neg x$ and $\sim (\neg x)$ for x. The relations between the propositional variables are given by means of conditional rules written as $R(X) = 2^{L(X)} \times L(X)$: the set of pairs of a set of literals built from X and a literal built from X, written as $l_1 \wedge \ldots \wedge l_n \rightarrow l$, and, when $n = 0$, $\top \rightarrow l$. The rules will be used to represent the relations among propositional variables existing in beliefs, desires and goal of the agents.

Then there are the different sorts of agents A we consider. Besides real agents RA (either human or artificial) we consider as agents in the model also socially constructed agents like groups, normative systems and organizations SA. These latter agents do not exist in the usual sense of the term. Rather, they exist only as they are attributed mental attitudes by other agents (either real or not). By mental attitudes we mean beliefs B, desires D and goals G.

Coming to the relations existing between these structural concepts, mental attitudes, even if they do not coincide with, are represented by rules: $MD : B \cup D \cup G \rightarrow R(X)$. When there is no risk of confusion we will abuse the notation by identifying rules and mental states. To resolve conflicts among motivations we introduce a priority relation by means of $\geq : A \rightarrow 2^M \times 2^M$ 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 $\geq (a)$. Moreover, different mental attitudes are attributed to all the different sorts of agents by the agent description relation $AD : A \rightarrow 2^{B \cup D \cup G \cup A}$. We write $B_a = AD(a) \cap B$, $A_a = AD(a) \cap A$ for $a \in A$, *etc.*

Also agents are in the target of the AD relation for the following reason: groups, normative systems, and organizations agents exist only as profiles attributed by other agents. So groups, normative systems and organizations exist only as they are described as agents by other agents, according to the agent description relation. The AD relation induces an exists-in-profile relation specifying that an agent $b \in SA$ exists only as some other agents attribute to it mental attitudes: $\{a \mid b \in A_a\}$.

Finally, the different sorts of agents are disjoint and are all subsets of the set of agents A: $RA \cup SA \subseteq A$.

We introduce now concepts concerning informational aspects. First of all, the set of variables whose truth value is determined by an agent (decision variables) [17] are distinguished from those which are not P (the parameters).

Besides, we need to represent also the so called "institutional facts" I. They are states of affairs which exist only inside normative systems and organizations. As discussed in the previous section, Searle [22] suggests, money, properties, 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 are just in the beliefs of these agents. Similarly, we need to represent that social entities like normative systems and organizations are able to change themselves. The actions determining the changes are

called creation actions C. Finally, we introduce contracts CT: they are agreements between agents in normative systems or organizations which have legal consequences; they are defined in Section 3.2.

As concerns the relations among these concepts, we have that parameters P are a subset of the propositional variables X. The complement of X and P represents the decision variables controlled by the different agents. Hence we have to associate to each agent a subset of $X \setminus P$ by extending again the agent description relation $AD : A \rightarrow 2^{B \cup D \cup G \cup A \cup (X \setminus P)}$.

Moreover, the institutional facts I are a subset of the parameters P: $I \subseteq P$. And the creation actions C are a subset of the institutional facts $C \subseteq I$: they do not exist outside the mind of agents and they have effects on the mental attitudes of agents only as far as the agents believe they have.

Since social entities depend on the attribution of mental attitudes, we represent their modification by means of the modification of their mental attitudes expressed as rules. So the creation action relation $CR : \{b, d, g\} \times A \times R(X) \rightarrow C$ is a mapping from rules (for beliefs, desires and goals) to propositional variables, where $CR(b, a, r)$ stands for the creation of $m \in B_a$, $CR(d, a, r)$ stands for the creation of $m \in D_a$, and $CR(g, a, r)$ stands for the creation of $m \in G_a$, such that the mental attitude is described by the rule $r \in R(X)$: $r = MD(m)$. For space reasons, in this paper we consider only the creation of new rules and not their deletion from the mental attitudes of an agent.

Since institutional facts I and the creation actions C exist only in the beliefs of a normative system or an organization, we need a way to express how these beliefs can be made true. As we discussed in the previous section, the relations among propositional variables are expressed as rules. In this case we have rules concerning beliefs about institutional facts: they are called constitutive rules and represent the "counts as" relations introduced by Searle [22] (see previous section). We thus identify the subset CN of the belief rules which express the relation between propositional variables and institutional facts: rules $C \cup \{x\} \rightarrow y \in R(X)$ expressing the fact that a literal $x \in L(X)$ in context $C \subseteq Lit(X)$ counts as the institutional fact $y \in L(I)$.

Finally, we have to model the effect of the creation actions on the mental attitudes of agents. For this reason we introduce an update relation UP from creation actions and mental attitudes to set of rules representing the new mental attitudes $UP : \{B, D, G\} \times C \rightarrow 2^{\{B, D, G\}}$. Since a decision of an agent can make true some creation actions, the consequences of these actions must be considered in the subsequent reasoning of the agent. The update function can be used to define the history of the multiagent system representing its evolution. Note that in this paper we do not consider the problem of the belief (and goal) revision. While this problem is sometimes addressed when dealing with normative systems, we consider here only the problem of introducing we rules and not of deciding which rules are necessary to get a certain revision.

We can now define a multiagent system as $MAS = \langle RA, SA, X, P, B, D, G, C, AD, MD, \geq, I, CT \rangle$.

We need to introduce normative multiagent systems to model organizations which are able to issue and enforce obligations: let the normative agent $\mathbf{o} \in SA$ be an agent representing the organization. Let the norms $\{n_1, \ldots, n_m\} = N$ be a set. Let the norm description $V : N \rightarrow X_{\mathbf{o}} \cup P$ be a complete function from the norms to the decision

variables of the normative agent together with the parameters: we write $V(n, a)$ for the decision variable which represents that there is a violation of norm n by agent $a \in A$. With these elements we define sanction based obligations in the next section. The tuple $\langle RA, SA, X, P, B, D, G, C, AD, MD, \geq, I, CT, \mathbf{o}, N, V \rangle$ is a normative multiagent system $NMAS$.

As concerns the behavior of agents, in Section 4, we introduce the games that can be played between two agents $\mathbf{a}$ and $\mathbf{o}$. Before games, we have to introduce two further notions: consequences of beliefs and decisions of agents.

To incorporate the consequences of belief rules, we introduce a simple logic of rules called out: it takes the transitive closure of a set of rules, called reusable input/output logic in [18]; $out(E, S)$ be the closure of $S \subseteq L(X)$ under the rules E:

- $out^0(E, S) = S$
- $out^{i+1}(E, S) = out^i(E, S) \cup \{l \mid L \to l \in E, L \subseteq out^i(E, S)\}$ for $i \geq 0$
- $out(E, S) = \cup_0^\infty (E, S)$

We can now introduce decisions of agents; they must be consistent with the consequences of beliefs according to the two agents $\mathbf{a}$ ($out(B_\mathbf{a}, \delta)$) and $\mathbf{o}$ ($out(B_\mathbf{o}, \delta_\mathbf{o} \cup out(B_\mathbf{o}, \delta_\mathbf{a}))$). The set of decisions Δ is the set of subsets $\delta = \delta_\mathbf{a} \cup \delta_\mathbf{o} \subseteq L(X)$ such that their closures under the beliefs $out(B_\mathbf{a}, \delta)$ and $out(B_\mathbf{o}, \delta_\mathbf{o} \cup out(B_\mathbf{o}, \delta_\mathbf{a}))$ do not contain a variable and its negation.

3.1 Obligations

Since contracts affect the obligations of an agent, we must first summarize their definition given in [3]. Obligations are defined in terms of goals of the addressee of the norm $\mathbf{a}$ and of the organization $\mathbf{o}$. The definition of obligation contains several clauses. The first one is the central clause of our definition and defines obligations of agents as goals of the normative agent, following the 'Your wish is my command' strategy [3]. The first clause says that the obligation is implied by the desires of agent $\mathbf{o}$, implied by the goals of agent $\mathbf{o}$.

The second and third clauses can be read as "the absence of p is considered as a violation". The association of obligations with violations is inspired to Anderson [1]'s reduction of deontic logic to alethic logic. The third clause says that the agent desires that there are no violations.

The fourth and fifth clauses relate violations to sanctions. The fourth clause assumes that agent $\mathbf{o}$ is motivated not to count behavior as a violation and apply sanctions as long as their is no 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 1 (Obligation). *Let* $NMAS = \langle RA, SA, X, P,$
$B, D, G, C, AD, MD, \geq, I, CT, \mathbf{o}, N, V \rangle$ *be a normative multiagent system.*

Agent $\mathbf{a} \in A$ *is obliged to decide to do* $x \in L(X_{\mathbf{a}} \cup P)$ *with sanction* $s \in L(X_{\mathbf{o}} \cup P)$
if $Y \subseteq L(X_{\mathbf{a}} \cup P)$ *in* $NMAS$, *written as* $NMAS \models O_{\mathbf{ao}}(x, s | Y)$, *if and only if:*

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

Permissions and prohibitions can be defined in terms of motivational attitudes, too [4].

As discussed in [3], sanctions or rewards are not the only possible motivations to stick to obligations, but they are necessary to cope for the worst cases.

3.2 Contracts

Contracts are part of the beliefs attributed to the organization $\mathbf{o}$: this fact makes it possible that they change the beliefs of the organization according to what specified by the organization itself.

A contract $ct \in CT$ is created (a fact represented by the institutional fact $c \in I$) only if the organization believes that some other fact has as a consequence that c is true. More precisely, if there is some fact which counts as c for the organization $\mathbf{o}$. This fact can be a brute fact in the world or another institutional fact. E.g., since contracts are created by agreements, the contract c is created by the signatures of two agents, two decision variables e and f: a constitutive norm in the belief rules of agent $\mathbf{o}$ ($e \wedge f \rightarrow c \in B_{\mathbf{o}}$). One reason why the creation of the contract c is introduced as an intermediary between the agreement and its legal effects is that, as many other institutional facts, it allows decoupling the conditions of the creation of the institutional facts from its legal effects. In this way, e.g., it is possible to specify new ways of creating the contract (for instance using an electronic signature) maintaining the same rules specifying its legal effects.

The effect a contract achieves is to modify the mental attitudes of the normative system. Usually, it adds more than one rule to the beliefs $B_{\mathbf{o}}$, the desires $D_{\mathbf{o}}$, or the goals $G_{\mathbf{o}}$ by making true some creation actions in C. Again, the creation actions are institutional facts: they are made true only if the organization $\mathbf{o}$ believes that they are made true by the creation of the contract: e.g., $c \rightarrow t \in B_{\mathbf{o}}$, is another constitutive rule, read as $c \in I$ counts as the creation action $t \in C$. Since a contract counts as several creation actions $t \in C$, the conclusion is that c works as an abstraction: rather than connecting the signatures of the agents with the creation actions, the contract unifies all the legal effects.

Finally, we consider which mental attitudes are changed. A contract modifies the mental attitudes of the normative system: since both regulative rules like obligations and constitutive ones like those composing contracts are themselves defined in terms of mental attitudes of the normative system, a contract can have legal effects.

By making true some creation actions, a contract is able to create regulative norms as the obligations of an agent $\mathbf{a}$ to pay ($pay \in X_{\mathbf{a}}$) in case the requested good has been shipped to him; $O_{\mathbf{ao}}(pay, s \mid shipped)$ is defined by the normative goal and desire that shipped goods are payed: $shipped \rightarrow pay \in D_{\mathbf{o}} \cap G_{\mathbf{o}}$; the goal and desire to consider the lack of payment for shipped goods as a violation: $shipped \wedge \neg pay \rightarrow V(\neg pay, \mathbf{a}) \in D_{\mathbf{o}} \cap G_{\mathbf{o}}$. And finally, the goal and desire to sanction violations: $V(\neg pay, \mathbf{a}) \rightarrow s \in D_{\mathbf{o}} \cap G_{\mathbf{o}}$; avoiding the sanction $\top \rightarrow \neg s$ is a desire of agents $\mathbf{a}$ and $\mathbf{o}$, thus, it is a precondition of the obligation [5].

The creation of the contract achieves these effects on the mental attitudes of the organization $\mathbf{o}$ since it counts as a series of creation actions: that the goals and desires defining the obligation are added. Since the counts as relation is described by constitutive rules in the beliefs of agent $\mathbf{o}$ we have (as concerns goals):

$$\{$$
$$c \rightarrow CR(g, \mathbf{o}, shipped \rightarrow pay),$$
$$c \rightarrow CR(g, \mathbf{o}, shipped \wedge \neg pay \rightarrow V(\neg pay, \mathbf{a})),$$
$$c \rightarrow CR(g, \mathbf{o}, V(\neg pay, \mathbf{a}) \rightarrow s)$$
$$\} \subseteq B_{\mathbf{o}}.$$

Also constitutive rules can be created by contracts: they are defined by belief rules of the normative system $\mathbf{o}$, so they are created by a creation action $CR(b, \mathbf{o}, t) \in C$.

First of all, the contract may specify some institutional fact which should be used in the interaction. E.g., the shipment of the exchanged good is an institutional fact $shipped \in I$; the fact that the good has been shipped is not a brute fact of the world (the buyer cannot check it), rather it is a fact which holds if there is some document like the so called bill of lading ($bill \in P$) issued by a third party [14]: $bill \rightarrow shipped$ is the rule t added to the beliefs of the organization $\mathbf{o}$ by the creation action $CR(b, \mathbf{o}, t) \in C$; the creation action is a consequence of the contract c: the constitutive rule $c \rightarrow CR(b, \mathbf{o}, bill \rightarrow shipped) \in B_{\mathbf{o}}$ creates another constitutive rule.

Second, constitutive rules created by contracts will introduce new obligations and new constitutive rules. In this way a contract can specify how new obligations may arise during the interaction of the parties. We return on [12]'s example: if an agent does not pay the fee for a shipped good, it is obliged to pay a double sum of money ($pay2$): $O_{\mathbf{ao}}(pay2, s' \mid shipped \wedge \neg pay)$. This obligation is not a preexisting conditional obligation: it is created as a legal consequence of an event, the sanction s for not having payed the fee. The sanction s, in this case, rather than being a direct punishment for agent $\mathbf{a}$, 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 s. This part of the contract is thus represented by the constitutive rules which create further constitutive rules about goals (where $s' \in X_\mathbf{o}$ is a sanction both feared by agent $\mathbf{a}$ and not desired by agent $\mathbf{o}$): e.g.,

$\{$

$\quad c \rightarrow CR(b, \mathbf{o}, s \rightarrow CR(g, \mathbf{o}, shipped \wedge \neg pay \rightarrow pay2)),$
$\quad c \rightarrow CR(b, \mathbf{o}, s \rightarrow CR(g, \mathbf{o}, \neg pay2 \rightarrow V(\neg pay2, \mathbf{a})))$

$\} \subseteq B_\mathbf{o}$

In summary, a contract is defined as:

Definition 2 (Contract). *A contract $ct \in CT$ is a triple:*

1. *An institutional fact $c \in I$ representing that the contract $ct \in CT$ has been created.*
2. *A constitutive rule which makes true the institutional fact c: $Y \rightarrow c \in B_\mathbf{o}$ where $Y \subseteq L(X)$*
3. *A set of constitutive rules having as antecedent the creation c of the contract ct and as consequent creation actions modifying the mental attitudes of the organization $\mathbf{o}$: $c \rightarrow CR(E, \mathbf{o}, r) \in B_\mathbf{o}$ where $E = \{b, d, g\}$.*

4 Games

The advantage of the attribution of mental attitudes to organizations is that standard techniques developed in decision and game theory can be applied to reasoning on contracts. Here we consider a simple form of games of two stages only where an agent $\mathbf{a}$ takes the normative agent $\mathbf{o}$ into account by playing games with it.
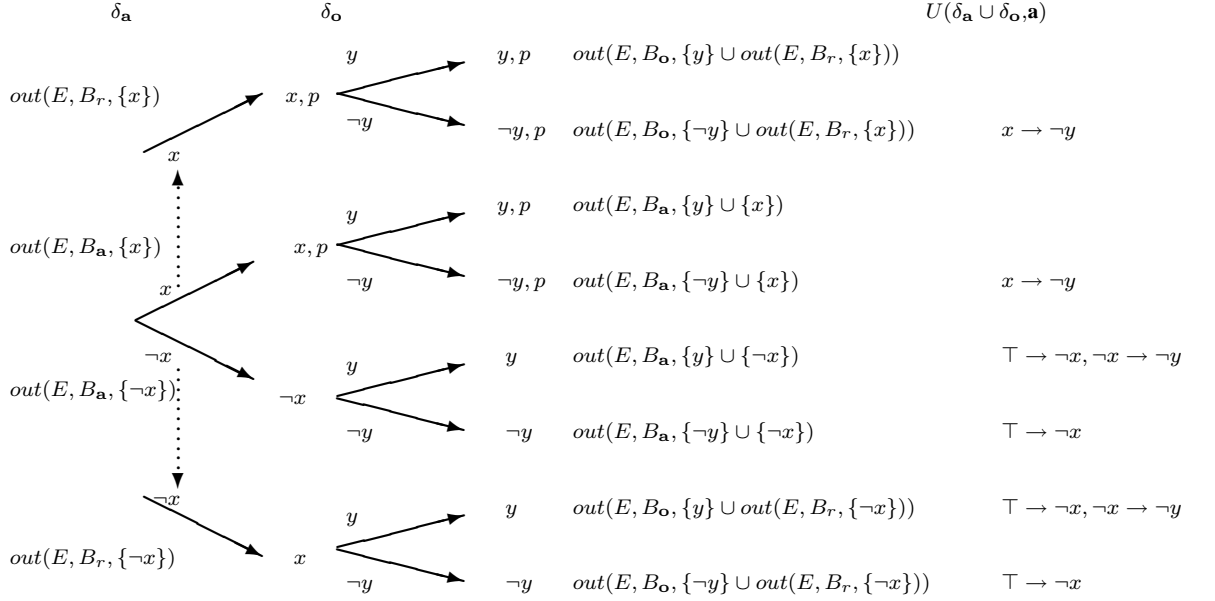
When agent $\mathbf{a}$ takes its decision $\delta_\mathbf{a}$ it has to minimize its unfulfilled motivational attitudes. But when it considers these attitudes, it must not only consider its decision $\delta_\mathbf{a}$ and the consequences of this decision; it must consider also the decision $\delta_\mathbf{o}$ of the organization $\mathbf{o}$ and its consequences, for example that it is sanctioned by agent $\mathbf{o}$. So agent $\mathbf{a}$ recursively considers which decision agent $\mathbf{o}$ will take depending on its different decisions $\delta_\mathbf{a}$. Note that here we assume that $\mathbf{o}$ is aware of agent $\mathbf{a}$'s decision: hence, agent $\mathbf{o}$ takes its decision considering the legal effects of agent $\mathbf{a}$'s decision on its beliefs and motivations using the update function UP and the creation actions made true by the decision $\delta_\mathbf{a}$.

Given a decision $\delta_\mathbf{a}$, a decision $\delta_\mathbf{o}$ is optimal for agent $\mathbf{o}$ 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 $\mathbf{a}$ is more complex: for each decision $\delta_\mathbf{a}$ it must consider which is the optimal decision $\delta_\mathbf{o}$ for agent $\mathbf{o}$.

Definition 3 (Recursive modelling). *Let:*

– *the unfulfilled motivations of decision δ according to agent $\mathbf{a} \in A$ be the set of motivations whose body is part of the closure of the decision under the belief rules but whose head is not.*
  $U(\delta, \mathbf{a}) = \{m \in M \mid MD(m) = l_1 \wedge \ldots \wedge l_n \rightarrow l, \{l_1, \ldots, l_n\} \subseteq out(B_\mathbf{a}, \delta) \text{ and } l \notin out(B_\mathbf{a}, \delta)\}.$

$\delta_{\mathbf{a}}$	$\delta_{\mathbf{o}}$			$U(\delta_{\mathbf{a}} \cup \delta_{\mathbf{o}}, \mathbf{a})$
		y	y, p $\quad out(E, B_{\mathbf{o}}, \{y\} \cup out(E, B_r, \{x\}))$	
$out(E, B_r, \{x\})$	x, p			
		$\neg y$	$\neg y, p$ $\quad out(E, B_{\mathbf{o}}, \{\neg y\} \cup out(E, B_r, \{x\}))$	$x \rightarrow \neg y$
x				
		y	y, p $\quad out(E, B_{\mathbf{a}}, \{y\} \cup \{x\})$	
$out(E, B_{\mathbf{a}}, \{x\})$	x, p			
x		$\neg y$	$\neg y, p$ $\quad out(E, B_{\mathbf{a}}, \{\neg y\} \cup \{x\})$	$x \rightarrow \neg y$
		y	y $\quad out(E, B_{\mathbf{a}}, \{y\} \cup \{\neg x\})$	$\top \rightarrow \neg x, \neg x \rightarrow \neg y$
$\neg x$	$\neg x$			
$out(E, B_{\mathbf{a}}, \{\neg x\})$		$\neg y$	$\neg y$ $\quad out(E, B_{\mathbf{a}}, \{\neg y\} \cup \{\neg x\})$	$\top \rightarrow \neg x$
$\neg x$				
		y	y $\quad out(E, B_{\mathbf{o}}, \{y\} \cup out(E, B_r, \{\neg x\}))$	$\top \rightarrow \neg x, \neg x \rightarrow \neg y$
$out(E, B_r, \{\neg x\})$	x			
		$\neg y$	$\neg y$ $\quad out(E, B_{\mathbf{o}}, \{\neg y\} \cup out(E, B_r, \{\neg x\}))$	$\top \rightarrow \neg x$

Fig. 1. The game between agent **a** and agent **o**.

- *the unfulfilled motivations of decision δ according to agent **o** be the set of motivations whose body is part of the closure of the decision under the belief rules and whose head is not, but considering the consequences of agent **a**'s decision on agent **o**'s beliefs and motivations. We write $O = out(B_{\mathbf{o}}, \delta_{\mathbf{a}}) \cap C$ for the set of creation actions which follow from $\delta_{\mathbf{a}}$:*

  $U(\delta, \mathbf{o}) = \{m \in UP(D_{\mathbf{o}} \cup G_{\mathbf{o}}, O) \mid MD(m) = l_1 \wedge \ldots \wedge l_n \rightarrow l, \{l_1, \ldots, l_n\} \subseteq out(UP(B_{\mathbf{o}}, O), \delta_{\mathbf{o}} \cup out(B_{\mathbf{o}}, \delta_{\mathbf{a}}))$ and $l \notin out(UP(B_{\mathbf{o}}, O), \delta_{\mathbf{o}} \cup out(B_{\mathbf{o}}, \delta_{\mathbf{a}}))\}.$

- *A decision δ (where $\delta = \delta_{\mathbf{a}} \cup \delta_{\mathbf{o}}$) is* optimal *for agent **o** if and only if there is no decision $\delta'_{\mathbf{o}}$ such that $U(\delta, \mathbf{o}) >_{\mathbf{o}} U(\delta_{\mathbf{a}} \cup \delta'_{\mathbf{o}}, \mathbf{o})$. A decision δ is optimal for agent **a** and agent **o** if and only if it is optimal for agent **o** and there is no decision $\delta'_{\mathbf{a}}$ such that for all decisions $\delta' = \delta'_{\mathbf{a}} \cup \delta'_{\mathbf{o}}$ and $\delta_{\mathbf{a}} \cup \delta''_{\mathbf{o}}$ optimal for agent **o** we have that $U(\delta', \mathbf{a}) >_{\mathbf{a}} U(\delta_{\mathbf{a}} \cup \delta''_{\mathbf{o}}, \mathbf{a})$.*

4.1 Example

We now return to the example about trade contracts. For space reasons, we formalize it as concerns only the obligation $O_{\mathbf{ao}}(pay, s \mid shipped)$ and the constitutive rule saying that the bill of lading counts as the good has been shipped.

We have two agents: the agent $\mathbf{a} \in RA$ and the organization $\mathbf{o} \in SA$. Agent **a** attributes mental attitudes to the organization $\mathbf{o}$ ($\mathbf{o} \in A_{\mathbf{a}}$).

The agent **a** can perform the actions of signing a contract and paying ($\{sign, pay\} \subseteq X_{\mathbf{a}}$), it believes that it has already signed the contract and the bill of lading $bill \in P$

has been issued $\{\top \to sign, \top \to bill\} \subseteq B_\mathbf{a}$, it desires not to give its money away ($\top \to \neg pay, \in D_\mathbf{a}$) and not to be sanctioned by agent $\mathbf{o}$ ($\top \to \neg s \in D_\mathbf{a}$).

The organization $\mathbf{o}$ does not desire to consider a violator ($V(\neg pay, \mathbf{a}) \in X_\mathbf{o}$) and to sanction agent $\mathbf{a}$ ($s \in X_o$) without motivation: $\{\top \to \neg V(\neg pay, \mathbf{a}), \top \to \neg s\} \subseteq D_\mathbf{o}$. It believes that if agent $\mathbf{a}$ signs ($sign$) the contract, this counts as the creation ($c \in I$) of the contract ($ct \in CT$): $sign \to c \in B_\mathbf{o}$. It believes that the contract has been signed and the bill of lading ($bill \in P$) has been issued $\{\top \to sign, \top \to bill\} \subseteq B_\mathbf{o}$ (as agent $\mathbf{a}$ does) and also the constitutive norms concerning the effects of the contract.

The first effect is that the new obligation to pay when the good is shipped is introduced: $O_\mathbf{ao}(pay, s \mid shipped)$. The obligation is defined by a set of desires and goals: the normative goal and desire that shipped goods are payed: $shipped \to pay \in D_\mathbf{o} \cap G_\mathbf{o}$; the goal and desire to consider the lack of payment for shipped goods as a violation: $shipped \land \neg pay \to V(\neg pay, \mathbf{a}) \in D_\mathbf{o} \cap G_\mathbf{o}$. And the goal and desire to sanction violations: $V(\neg pay, \mathbf{a}) \to s \in D_\mathbf{o} \cap G_\mathbf{o}$; note that the desire $\top \to \neg s$ of agents $\mathbf{a}$ and $\mathbf{o}$ are requested by the definition of obligation. The contract achieves these effects on the mental attitudes of the organization $\mathbf{o}$ since it counts as a series of creation actions: that the goals and desires defining the obligation are added. Since the counts as relation is described by constitutive norms, i.e., belief rules of agent $\mathbf{o}$, we have:

$$\{$$

$$c \to CR(g, \mathbf{o}, shipped {\to} pay),$$

$$c \to CR(d, \mathbf{o}, shipped {\to} pay),$$

$$c \to CR(g, \mathbf{o}, shipped \land \neg pay {\to} V(\neg pay, \mathbf{a})),$$

$$c \to CR(g, \mathbf{o}, V(\neg pay, \mathbf{a}) {\to} s),$$

$$c \to CR(d, \mathbf{o}, shipped \land \neg pay {\to} V(\neg pay, \mathbf{a})),$$

$$c \to CR(d, \mathbf{o}, V(\neg pay, \mathbf{a}) {\to} s)$$

$$\} \subseteq B_\mathbf{o}$$

The second effect is that the bill of lading ($bill$) is considered as the proof that the good has been shipped; the contract creates a constitutive rule in the beliefs of the normative system $\mathbf{o}$: $c \to CR(b, \mathbf{o}, bill \to shipped) \in B_\mathbf{o}$.

We adopt the perspective of agent $\mathbf{a}$ who has to decide whether to pay its fee or not. To take a decision agent $\mathbf{a}$ must recursively model the organization $\mathbf{o}$'s decision. Agent $\mathbf{a}$ takes the decision whose consequences minimize its unfulfilled motivational attitudes given the decision of the organization and its consequences. Moreover, the decision of the organization $\mathbf{o}$ is assumed to be taken from the point of view which considers the legal effects in the consequences $out(B_\mathbf{o}, \delta_\mathbf{a})$ of agent $\mathbf{a}$'s decision. Agent $\mathbf{a}$ has already signed the contract, so its signature counts as the creation of the contract with its legal effects, $c \in out(B_\mathbf{o}, \delta_\mathbf{a})$:

$O = out(B_{\mathbf{o}}, \delta_{\mathbf{a}}) \cap C =$
$\{$
  $CR(g, \mathbf{o}, shipped \rightarrow pay),$
  $CR(d, \mathbf{o}, shipped \rightarrow pay),$
  $CR(g, \mathbf{o}, shipped \land \neg pay \rightarrow V(\neg pay, \mathbf{a})),$
  $CR(g, \mathbf{o}, V(\neg pay, \mathbf{a}) \rightarrow s),$
  $CR(d, \mathbf{o}, shipped \land \neg pay \rightarrow V(\neg pay, \mathbf{a})),$
  $CR(d, \mathbf{o}, V(\neg pay, \mathbf{a}) \rightarrow s),$
  $CR(b, \mathbf{o}, bill \rightarrow shipped)$
$\}$

The updated beliefs and motivations are:

$UP(B_{\mathbf{o}}, O) \setminus B_{\mathbf{o}} = \{bill \rightarrow shipped\}$

$UP(D_{\mathbf{o}}, O) \setminus D_{\mathbf{o}} = UP(G_{\mathbf{o}}, O) \setminus G_{\mathbf{o}} = \{shipped \rightarrow pay, shipped \land \neg pay \rightarrow V(\neg pay, \mathbf{a}), V(\neg pay, \mathbf{a}) \rightarrow s\}$

The organization $\mathbf{o}$ has to decide whether agent $\mathbf{a}$'s behavior respects the obligation or not; in the latter case agent $\mathbf{o}$ considers this as a violation and sanctions it.

The creation of the new constitutive rule has a further consequence, that the good has been shipped since the bill of lading counts as such:

$shipped \in out(UP(B_{\mathbf{o}}, O), \{\neg V(\neg pay, \mathbf{a}), \neg s\} \cup out(B_{\mathbf{a}}, \{\neg pay\})$

Thus, the new obligation $O_{\mathbf{ao}}(pay, s \mid shipped)$ has its condition satisfied. If the agent decides not pay it violates its duty. Agent $\mathbf{o}$'s unfulfilled mental attitudes are:

$U(\{\neg V(\neg pay, \mathbf{a}), \neg s\} \cup \{\neg pay\}, \mathbf{o}) \cap (D_{\mathbf{o}} \cup G_{\mathbf{o}}) =$
  $\{shipped \rightarrow pay, shipped \land \neg pay \rightarrow V(\neg pay, \mathbf{a})\}$

We assume that fulfilling the set of motivations $\{shipped \rightarrow pay, shipped \land \neg pay \rightarrow V(\neg pay, \mathbf{a})\}$ is preferred, according to the ordering $\geq_{\mathbf{o}}$ on motivations, with respect to fulfilling $\{shipped \rightarrow pay, \top \rightarrow \neg V(\neg pay, \mathbf{a}), \top \rightarrow \neg s\}$: sanctioning violations worths its cost.

So the optimal decision for the organization is to consider $\mathbf{a}$'s behavior as a violation and to sanction it $\delta_{\mathbf{o}} = \{V(\neg pay, \mathbf{a}), s\}$, as the unfulfilled motivations are:

$U(\{V(\neg pay, \mathbf{a}), s\} \cup \{\neg pay\}, \mathbf{o}) \cap (D_{\mathbf{o}} \cup G_{\mathbf{o}}) =$
  $\{shipped \rightarrow pay, \top \rightarrow \neg V(\neg pay, \mathbf{a}), \top \rightarrow \neg s\}$

Instead, given the decision to pay the fee $\delta_{\mathbf{a}} = \{pay\}$, the optimal decision of agent $\mathbf{o}$ is not to consider as a violation the behavior of agent $\mathbf{a}$ and not to sanction it. Given $\delta_{\mathbf{o}} = \{\neg V(\neg pay, \mathbf{a}), \neg s\}$ the unfulfilled mental attitudes are:

$U(\{\neg V(\neg pay, \mathbf{a}), \neg s\} \cup \{pay\}, \mathbf{o}) \cap (D_{\mathbf{o}} \cup G_{\mathbf{o}}) = \emptyset$

How does agent $\mathbf{a}$ take a decision?

- if $\delta_{\mathbf{a}} = \{\neg pay\}$, then $\delta_{\mathbf{o}} = \{V(\neg pay, \mathbf{a}), s\}$:
  $U(\{V(\neg pay, \mathbf{a}), s\} \cup \{\neg pay\}, \mathbf{a}) \cap (D_{\mathbf{a}} \cup G_{\mathbf{a}}) = \{\top \rightarrow \neg s\}$
- if $\delta_{\mathbf{a}} = \{pay\}$, then $\delta_{\mathbf{o}} = \{\neg V(\neg pay, \mathbf{a}), \neg s\}$:
  $U(\{\neg V(\neg pay, \mathbf{a}), \neg s\} \cup \{pay\}, \mathbf{a}) \cap (D_{\mathbf{a}} \cup G_{\mathbf{a}}) = \{\top \rightarrow \neg pay\}$

If paying is preferred to being sanctioned $\{\top \rightarrow \neg s\} >_{\mathbf{a}} \{\top \rightarrow \neg pay\}$, agent $\mathbf{a}$ decides for $\delta_{\mathbf{a}} = \{\neg pay\}$.

5 Related work and summary

In this paper we address the problem of defining contracts as legal institutions. Using the methodology of attributing mental attitudes to social entities like organizations, we show that contracts have as precondition an agreement which counts as the creation of the contract and as legal consequences the creation of new mental attitudes. These attitudes define new obligations as well as new constitutive rules. We also show that the new constitutive rules can be used to prescribe the subsequent behavior expected by the parties involved in the contract.

What distinguishes our approach from other models of counts as relations is that we can connect goals, and obligations defined as goals, to institutional facts inside the overall frame of the attribution of the status of agent to the normative system: institutional facts are beliefs of the normative agent as any other belief.

Teague and Sonenberg [23] 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 also the procedures for the parties' decommitment.

Dignum *et al.* [12] propose the language *LCR* for modelling contracts. They define contracts as tuples composed of agents, contract clauses, stages and interactional structure. With respect to their work we do not define the clauses of a contract as conditional obligations (as also Pacheco and Carmo [19] do). Rather we use constitutive rules which create obligations when the contract is created or when some relevant event happens. Finally, as they propose, we give a definition of obligations in terms of violations but we take a subjective perspective and consider the decision problem of an agent subject to obligations.

Daskalopulu and Maibaum [9] model contracts as processes having as states legal relations among the parties. They introduce obligations which are 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 agent.

Future work concerns extending the games with multiple stages, so to track the evolution of the contract. We also propose contracts for modelling roles using the agent metaphor. Contracts are used for assigning roles: they create the obligations of the holder of a role starting from the description of the role. Moreover, roles and functional areas in an organization can be created as new legal institutions. Finally, our approach can be used to model security issues in virtual communities of agents. Contractual access control has been proposed in recent developments to cope with the issue of delegating powers to modify rights and permissions.

References

1. A. R. Anderson. The logic of norms. *Logic et analyse*, 2, 1958.
2. G. Boella and L. Lesmo. A game theoretic approach to norms. *Cognitive Science Quarterly*, 2(3-4):492–512, 2002.
3. G. Boella and L. van der Torre. Attributing mental attitudes to normative systems. In *Procs. of AAMAS'03*, pages 942–943. ACM Press, 2003.

4. G. Boella and L. van der Torre. Local policies for the control of virtual communities. In *Procs. of IEEE/WIC Web Intelligence Conference*, pages 161–167. IEEE Press, 2003.

5. G. Boella and L. van der Torre. Permissions and obligations in hierarchical normative systems. In *Procs. of ICAIL'03*, pages 109–118, Edinburgh, 2003. ACM Press.

6. G. Boella and L. van der Torre. Contracts as legal institutions in organizations of autonomous agents. In *Procs. of AAMAS'04*, New York, 2004.

7. G. Boella and L. van der Torre. Regulative and constitutive norms in normative multiagent systems. In *Procs. of 9th International Conference on the Principles of Knowledge Representation and Reasoning*, Whistler (CA), 2004.

8. C. Castelfranchi. Engineering social order. In *Procs. of ESAW'00*, pages 1–18. Springer Verlag, 2000.

9. A. Daskalopulu and T.S.E. Maibaum. Towards electronic contract performance. In *Procs. of Legal Information Systems Applications*, pages 771–777, 2001.

10. C. Dellarocas. Negotiated shared context and social control in open multi-agent systems. In R. Conte and C. Dellarocas, editors, *Social Order in MAS*. Kluwer, 2001.

11. D. C. Dennett. *The Intentional Stance*. The MIT Press, Cambridge, MA, 1987.

12. 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.

13. P. J. Gmytrasiewicz and E. H. Durfee. Formalization of recursive modeling. In *Procs. of first ICMAS'95*, 1995.

14. J. Gordijn and Y.-H. Tan. A design methodology for trust and value exchanges in business models. In *Procs. of 16th BLED Conference*, pages 423–432, 2003.

15. H. L. A. Hart. *The Concept of Law*. Clarendon Press, Oxford, 1961.

16. A. Jones and J. Carmo. Deontic logic and contrary-to-duties. In D. Gabbay, editor, *Handbook of Philosophical Logic*, pages 203–279. Kluwer, 2001.

17. J. Lang, L. van der Torre, and E. Weydert. Utilitarian desires. *Autonomous Agents and Multi-agent Systems*, pages 329–363, 2002.

18. D. Makinson and L. van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.

19. O. Pacheco and J. Carmo. A role based model of normative specification of organized collective agency and agents interaction. *Autonomous Agents and Multi-agent Systems*, 6:145–184, 2003.

20. D.W.P. Ruiter. A basic classification of legal institutions. *Ratio Juris*, 10(4):357–371, 1997.

21. J.R. Searle. *Speech Acts: an Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, England, 1969.

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

23. V. Teague and L. Sonenberg. Investigating commitment flexibility in multi-agent contracts. In *Procs. of decision theoretic and game theoretic agents workshop at ICMAS'00*, Boston, 2000.