Attributing Mental Attitudes to Roles: The Agent Metaphor Applied to e-Trade Organizations

Guido Boella
Dipartimento di Informatica
Università di Torino - Italy
e-mail: guido@di.unito.it

Leendert van der Torre
SEN-3 - CWI
Amsterdam - The Netherlands
e-mail: torre@cwi.nl

ABSTRACT

In this paper we address the problem of defining roles in organizations like e-trade ones. The methodology we use is to model roles according to the agent metaphor: we attribute to roles mental attitudes, like beliefs, desires and goals, we relate them to the agent's required expertise and responsibilities, and we model role behavior in game theoretic terms. Analogously, the organization is modelled as an agent which acts as a normative system: it imposes obligations to roles and to the agents playing the roles.

1. INTRODUCTION

Multiagent systems are often proposed as a solution for the organizational design of open systems [2, 14, 17] among which those for e-trade. A key notion in the structure of an organization is that of role. Roles allow to specify the activities delegated by the organization to achieve its overall goal while abstracting from the individuals which will eventually play them. The description of a role is usually given in terms of normative descriptions [28], expectations [24], standardised patterns of behavior [17], social commitments [13, 18], goals and planning rules [14]. The normative description specifies the obligations an agent who plays the role (called the actor) should obey. Goals are his intrinsic motivations. Roles, thus, seem to be strictly related to the notion of agent: they are described using notions like actions, goals and obligations.

The research questions we address in this paper are: how can the roles in an organization like the ones for e-trade be considered and described as agents? How are obligations associated to actors and roles? Finally, how do the actors take a decision under the obligations related to their roles?

The methodology we use in this paper is to use a logical multiagent framework based on a qualitative game theory to model social entities as agents; using again the agent metaphor, the organization is an agent to which mental attitudes are attributed, too. Once the organization and roles are described in terms of mental attitudes it is possible to consider the decisions they take in a certain situation. Thus, we apply the methodology we use for describing and reasoning about other social entities like groups [9], virtual communities [5], contracts [7] and normative multiagent systems [4, 11].

The attribution of mental attitudes to roles is partially supported also by other authors; e.g., Sichman and Conte [30] propose "to use the same mental attitudes (goals, plans, actions and resources) to characterize the agent endogenous mental attitudes (which we call personal mental attitudes) and the ones which he is endowed with by playing the role (which we call role mental attitudes)". Dastani *et al.* [14] attribute to roles not only goals, but, as we propose, also beliefs; the beliefs of roles have the form planning rules: these rules specify how a goal can be achieved in a certain context. We show in the next section why and how it is possible to attribute to roles also other forms of beliefs and how they are used in the evaluation of the behavior of an actor.

A consequence of the agent metaphor is that in the organizational design roles are modelled as a kind of abstract agent, together with its capabilities and beliefs (the role's expertise) and its goals (its responsibilities). The attribution of motivations and of an autonomous behavior to roles implies that the autonomy of the actors of roles must be regulated. The control of autonomy requires obligations associated with sanctions. As we discuss in the paper obligations addressed to roles must be distinguished from those addressed to actors. For dealing with obligations we model the organization as a normative system defined as in [4].

In this paper, we do not consider the problem of the creation of the obligations related to roles, e.g., by means of a trade contract, and we do not discuss here the problem of the interaction among different roles, see, e.g., [16, 17].

In the next section we discuss the use of the agent metaphor for roles. In Section 3 we introduce the conceptual model. In Section 4 we summarize our model of obligations. In Section 5 we present the games which can be played with roles, together with a detailed example in a e-trade scenario. Related work, discussion and summary close the paper.

2. ROLES ARE AGENTS

Why and how do we adopt the agent metaphor? In the life-cycle of an organization we can distinguish three phases: 1) organizational design, 2) role assignment, 3) organizational governance. The attribution of mental attitudes to roles is useful in all these three phases.

If we design an organization in terms of roles by attributing mental attitudes to them, we can design it like we would design an organization in terms of agents. However, roles are not agents with their individual aspects which may be problematic or complex, but abstract agents that can be defined by the organization. Moreover, in this way, the organization does not depend on the agents operating in it: they can be replaced by other ones playing the same roles. Roles, as agents, are able to take autonomous and proactive decisions. Thus roles can be delegated tasks which are reached with

a greater flexibility: it would be not sufficient to describe roles as machines. On the other hand, the autonomy implies that actors can depart from the goals of the roles. So, roles are often defined in terms of or associated with obligations to balance its autonomy.

If we consider the problem of assigning roles to actors and we attribute mental attitudes to roles, then the problem becomes finding a match between the beliefs representing expertise of the role and the beliefs of the actor. Moreover, as also Dastani *et al.* [14] discuss, it is possible to understand whether an agent is suited for a role from the fact that he already possesses the goals required by the role, or these goals are compatible with his private goals, or the incentives (i.e., the sanctions associated with his obligations) are enough to resolve the conflicts between the agent's private goals and the goals of the role.

Finally, if we consider the governance of organizations, then the attribution of mental attitudes to roles makes it easier to verify whether an agent is acting according to the overall plan of the organization: if he is acting according to the role's beliefs and achieving the role's goals, then he is fulfilling his responsibilities.

Which beliefs and goals are attributed to a role? Not all the beliefs of the organization are beliefs of a role due to the role assignment problem: the less beliefs a role has, the easier it becomes to assign agents to this role. Similarly for goals - not all the goals of an organization are the goals of a role due to the governance problem: to distribute responsibilities over the agents, and to control the running system. Consequently, the organizational design aims at defining roles with minimal beliefs and goals. In reality, however, the picture is more complicated than this, because not only the beliefs and goals of an organization may not be goals of the role, also vice versa a goal of a role may not be a goal of the organization. These more complex phenomena are the topic of the following subsections.

2.1 Beliefs of roles

Why are the actor's beliefs and those of the organization not sufficient to model roles? Beliefs should be attributed to roles if they can be different from the beliefs of the agent playing the role or from those of the organization the role belongs to.

First, Tuomela in his analysis of beliefs attributed to groups [32] suggests that the actor (holder of a position in his terminology) can have beliefs different from the role's:

> "Positional beliefs (or positional views) [...] are beliefs or views that a position-holder has qua a position-holder or has internalized and accepted as a basis of his performances of [...] social tasks. I will assume that the correct social and normative circumstances C will implicitly partially define positional beliefs. Accordingly, I will explicitly speak of positional beliefs as distinguished from mere personal beliefs. A position-holder can naturally have his own personal beliefs at the same time with his positional beliefs, some or even all of which may contradict his personal ones [...]. To take a fictitious example, we may think that the Flat Earth Society secretary has the positional rule-based belief that the earth is flat. Her having that belief is based on the general presupposition underlying the existence and characteristic or constitutive rules of that society, viz. the assumption that the earth is flat. But her personal belief might be different. [...] Generally speaking, positional beliefs rely on the presuppositions embodied in the statutes, by-laws, and other constitutive rules as well as (possibly) on various informal rules and proper social norms characterizing a collective."

Differently from Tuomela, we do not claim that positional beliefs should be beliefs of the actor, rather, the actor should act as if he had such beliefs. Since agents are autonomous their beliefs and motivations cannot be changed by the fact of playing a role.

A clear example where roles have beliefs which may be different from those of the actors, as suggested by Gordijn and Tan [20], can be found in the legislation of international trade (the UN Convention on International Multimodal Transport of Goods):

> "Suppose we have a seller in Hong Kong and a buyer in the Netherlands. [...] On the one hand the seller does not want to ship the goods onto the carriers 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; which is 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 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. Note that this article has a normative element. Whether the Bill of Lading is evidence does not depend so much on whether a person is psychologically convinced by it, but the law simply stipulates that everybody involved in a letter of credit procedure should consider this document as sufficient evidence."

We use this example of a trading organization which defines the roles of buyer and seller as a running example in the rest of the paper assuming the perspective of the buyer. The role of buyer has a belief attributed by the trading organization: if he believes that the bill of lading has been issued, then he believes that the good has been shipped. Moreover the role of buyer has a goal: when the good has been shipped, he wants to pay the fee for the good. However, the agent who plays the role may not believe that the good has been shipped, and he could even claim that he has not to proceed with the payment.

Moreover, it is necessary to attribute beliefs to roles since they can be different from the beliefs of the organization the role belongs to. E.g., as suggested by Goffman [19], in foreign affairs, ambassadors are not entitled to know effectively which are the beliefs and goals of the nation they represent. Their government can use the ambassadors to mislead another nation by obliging them to propose some deal or manage a negotiation in a way which is different from real beliefs of the government. For instance, the government can communicate to the ambassador a false reservation price for a negotiation. It is neither necessary nor even useful that the person playing the role knows which are the real beliefs of the organization. Because being unaware of the real state of affairs the agent can play his role in a more spontaneous way, or because in this way the organization avoids the risk of a security violation, or even because the beliefs of the organizations are too complex for the actor, so that he would be overwhelmed by information. Moreover, the organization may not want to the actor to discover its real beliefs, since it is violating the law.

2.2 Obligations of roles and of actors

Besides beliefs, also motivational attitudes like desires and goals are attributed to roles. It is clear that the motivations of the role can be different from those of the actor (e.g., an employee can fulfill his role for his salary, not for the sake of the organization); but, as for beliefs, the goals attributed to the role can be different from those of the organization, e.g., a role can be attributed the goal of proposing some deal to another agent just because his organization wants to mislead this agent about its true intentions.

In order to motivate an actor to fulfill the goals attributed to his role and to act as if he possessed the role's beliefs, obligations are associated to roles. In our example, the agent playing the role of the buyer has to pay when the good is shipped. Under this regard, however, we partially depart from most current proposals and we do not associate obligations with the definition of roles. Obligations are addressed to the actors playing the roles: they provide them additional motivations. Obligations provide motivations only since they are associated by sanctions or rewards. When we consider the motivational role of obligations, it appears that the obligations we are discussing are addressed only to actors and not to roles. As we discuss in [6], a precondition of obligations is that the addressee fears the sanction (or desires the reward). So it is not possible that the same obligation is addressed both to the role and to the actor: the actor does not necessarily have the same desires and goals as the role. Indeed, the role of obligations is to make the agent adopt the desires and goals of the role.

Moreover, since the obligations are addressed to the actor and not to the role, these obligations must be created not at the moment of the definition of the role in the organization, but at the moment an agent starts playing a role; e.g., when he becomes enrolled in an organization by signing an employment contract. The obligations are created starting from the goals specified in the definition of a role. If in our example the role has the goal to pay when the good has been shipped, then when the actor enters the role, the obligation that agent pays when the good is shipped is created. Sanctions or rewards associated to the new obligations are chosen according to the policies of the organization to which the role belongs. E.g., an employee can be fired for not fulfilling an obligation or payed for having done his work or rewarded with some additional benefit.

Finally, since organizations and roles are modelled as agents, nothing prevents that they are subject to obligations too. As Pacheco and Carmo [28] notice, an organization is considered as a legal person by the law. However, the role obligations are different from those addressed to the actor. Consider the rules of soccer. A player (a role) is sanctioned with an expulsion if he touches the ball with his hands: the sanction is against the role, who is attributed the desire to play and score goals; but the sanction is not necessarily against the actor (he can even do the fault on purpose to stop playing since he is tired). In contrast, an obligation addressed to the actor is not to use illegal drugs: the sanction (e.g., a fine) is against the agent and not the role (which has no money). However, the obligation holds only as long as he remains in the role of football player.

Analogously to obligations, roles can be subject to prohibitions, permissions, rights, e.g., to use some resources.

2.3 The behavior of roles and actors

The roles are modelled as autonomous agents who act to achieve their goals. Thus, one advantage of the agent metaphor is that we can model the interaction between actors, organizations and roles with the usual game-theoretic machinery. How is the behavior of a role evaluated? Roles act via the actions of the actors, who are subject to obligations. So what is evaluated by the organization is the behavior of the actors. In Section 5, we introduce games between actors and organizations. When an actor takes a decision, he has to consider what is considered obligatory, not only from his own point of view but also from the point of view of the role he is playing.

Consider the international trade example. The actor knows that the trade organization considers him obliged to pay the fee in case the bill of lading has been issued, since this means that the good has been shipped. So, when the actor decides what to do, he has to consider the fact that his actions are not evaluated according to his own beliefs: he could believe that the good has not been shipped, despite the bill of lading, so, from his point of view, he is not obliged to pay. But according to the beliefs attributed to the role, he is obliged to pay (the good has been shipped since the bill of lading has been issued). Hence, the actor has to act as if he had the role's beliefs, or else he violates his duties; when he takes a decision, he figures out which beliefs the organization attributes to his role.

Note that the obligations are directed towards the actor's actions and not towards his beliefs, even if there is a deontic component in the assertion that the bill of lading should be considered as evidence of the shipment. The reason is that an autonomous agent cannot be compelled to change his mental attitudes: he can be persuaded or motivated, but not coerced. Moreover, his mental states are not accessible. What can be punished is only his behavior: if he does not act in the same way as the role would have done, he is punished. This is the rationale underlying the fact that in the recursive modelling the organization considers the actor behavior from the point of view of the role.

3. THE CONCEPTUAL MODEL

In order to provide a formalization of roles inside an organization 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 $l_1 \wedge \ldots \wedge l_n \rightarrow l$ or, when $n = 0$, $\top \rightarrow l$. The rules are used to represent the relations among propositional variables existing in beliefs, desires and goal of the agents. It is well known that desires are different from goals, and we can adopt distinct logical properties for them. For example, goals can be adopted from other agents, whereas desires cannot. In this paper we do not make any additional assumptions on desires and goals, and we thus do not formally characterize the distinction between desires and goals, because it is beyond the scope of this paper.

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, organizations SA and roles RO. These agents do not exist in the usual sense. 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 are represented by rules, even if they do not coincide with them: $MD : B \cup D \cup G \rightarrow R(X)$. When there is no risk of confusion we 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 ($M = D \cup G$) 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: normative systems, organizations and roles exist only as profiles attributed by other agents. So groups, normative systems, organizations and roles 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 $b \in SA \cup RO$ exists only as some other agents attribute to it mental attitudes: $\{a \in A \mid b \in A_a\}$. The set $RO \cap A_o$ defines the role structure of the organization $o \in SA$.

Roles are described as agents but they are also associated with agents playing the role, $PL : RO \rightarrow RA$. An agent can play more than one role. Finally, the different sorts of agents are disjoint and are all subsets of the set of agent A: $RA \cup SA \cup RO \subseteq A$.

We introduce now concepts concerning informational aspects. First of all, the set of variables whose truth value is determined by an agent (decision variables) [23] are distinguished from those which are not (the parameters P). 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 Searle [29] suggests, money, debts, 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 attributed to these agents [11].

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 (their capabilities). Hence we have to associate to each agent a subset of $X \setminus P$ by extending again the agent description relation $AD : A \rightarrow 2^{B \cup D \cup G \cup A \cup (X \setminus P)}$. We write $X_a = AD(a) \cap X$.

Moreover, the institutional facts I are a subset of the parameters P: $I \subseteq P$. Since the institutional facts I 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 above, the relations among propositional variables are expressed as rules. In this case we have rules concerning beliefs about institutional facts. We thus identify a subset of the rules expressing the relation among propositional variables and institutional facts of an organization $o \in SA$: a belief rule $C \cup \{x\} \rightarrow y \in B_o$, called a constitutive rule, expresses the fact that a literal $x \in L(X)$ in context $C \subseteq Lit(X)$ has the institutional fact $y \in L(I)$ as a consequence; using Searle [29]'s terminology, x "counts as" y in context C for institution a.

Before introducing obligations in the next section, we define a multiagent system as $MAS = \langle RA, SA, RO, X, P, B, D, G, AD, MD, \geq, I \rangle$ and a normative system to model organizations. Let the normative agent $\mathbf{o} \in SA$ be an agent describing 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 $a \in A$. Finally, let the goal distribution $GD : A \rightarrow 2^{G_{\mathbf{o}}}$ be a function from the agents to the powerset of the goals of the normative agent, such that if $L \rightarrow l \in MD(GD(a))$, then $l \in L(X_a \cup P)$ for $\mathbf{a} \in A$.

The tuple $\langle RA, SA, RO, X, P, D, G, AD, MD, PL, \geq, I, \mathbf{o}, N, V, GD \rangle$ is called the normative multiagent system $NMAS$

4. OBLIGATIONS

Obligations are defined in terms of desires and goals of the addressee **a** and of the organization **o** in the normative multiagent system $NMAS$. The definition of obligation to do x in context Y with sanction s denoted by $O_{\mathbf{ao}}(x, s \mid Y)$ 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 [4]. The first clause says that the obligation is implied by the desires of agent **o**, implied by the goals of agent **o**, and it has been distributed by agent **o** to the agent. The latter two steps are represented by $GD(\mathbf{a})$.

The second and third clauses can be read as "the absence of x 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 the normative system is motivated not to count behavior as a violation and apply sanctions as long as their is no violation, because otherwise the obligation would have no effect. Finally, for the same reason, in the last clause that the addressee **a** does not like the sanction, so that the sanction can work as as a motivation.

DEFINITION 1 (OBLIGATION). *Let $NMAS = \langle RA, SA, RO, X, P, B, D, AD, MD, PL, \geq, I, \mathbf{o}, N, V, GD \rangle$ 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. *$Y \rightarrow x \in D_{\mathbf{o}} \cap GD(\mathbf{a})$: if agent **o** believes Y then it desires and has as a goal that x, and this goal has been distributed to agent **a**.*

2. *$Y \cup \{\sim x\} \rightarrow V(\sim x, \mathbf{a}) \in D_{\mathbf{o}} \cap G_{\mathbf{o}}$: if agent **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 **a**.*

3. *$\top \rightarrow \neg V(\sim x, \mathbf{a}) \in D_{\mathbf{o}}$: agent **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 **o** believes Y and decides $V(\sim x, \mathbf{a})$ then it desires and has as a goal that it sanctions agent **a**.*

5. *$Y \rightarrow \sim s \in D_{\mathbf{o}}$: if agent **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 **a** believes Y, then it desires $\sim s$: it does not like to be sanctioned.*

As discussed in [4], sanctions or rewards are not the only possible motivations to stick to obligations, but they are necessary to cope for the worst cases.

5. GAMES BETWEEN ACTORS AND ORGANIZATIONS

In this section we formalize the games played between the agent **a** playing the role **r** and the organization **o**.

First of all, to incorporate the 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 [25]:

DEFINITION 2 (CONSEQUENCES). *$out(E, S)$ be the closure of $S \subseteq L(X)$ under the rules E:*

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

When agent **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 **o** and its consequences $out(B_{\mathbf{a}}, \delta_{\mathbf{a}} \cup \delta_{\mathbf{o}})$, for example, that it is sanctioned by agent **o**. So agent **a** recursively considers which decision **o** will take depending on its different decisions $\delta_{\mathbf{a}}$. Moreover, when agent **a** considers how the organization **o** evaluates the consequences of its decision $\delta_{\mathbf{a}}$, it must use the beliefs $B_{\mathbf{r}}$ of the role **r** it is playing, rather than its own beliefs $B_{\mathbf{a}}$ ($out(B_{\mathbf{o}}, \delta_{\mathbf{o}} \cup out(B_{\mathbf{r}}, \delta_{\mathbf{a}}))$). As we discussed in Section 2, the performance of agent **a** is judged according to beliefs of the role **r**.

We can now introduce decisions consistent with the consequences of beliefs according to the two agents **a** and **o**.

DEFINITION 3 (DECISIONS). *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 $out(B_{\mathbf{o}}, \delta_{\mathbf{o}} \cup out(B_{\mathbf{r}}, \delta_{\mathbf{a}}))$ do not contain a variable and its negation.*

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

DEFINITION 4 (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.*
  $U(\delta, \mathbf{a}) = \{m \in M \mid MD(m) = l_1 \wedge \ldots \wedge l_n \rightarrow l, \{l_1, \ldots, l_n\} \subseteq out(B_{\mathbf{a}}, \delta)$ *and* $l \notin out(B_{\mathbf{a}}, \delta)\}$.
- *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 but whose head is not, but considering the consequences of the decision $\delta_{\mathbf{a}}$ from the role* **r***'s point of view:*
  $U(\delta, \mathbf{o}) = \{m \in M \mid MD(m) = l_1 \wedge \ldots \wedge l_n \rightarrow l, \{l_1, \ldots, l_n\} \subseteq out(B_{\mathbf{o}}, \delta_{\mathbf{o}} \cup out(B_{\mathbf{r}}, \delta_{\mathbf{a}}))$ *and* $l \notin out(B_{\mathbf{o}}, \delta_{\mathbf{o}} \cup out(B_{\mathbf{r}}, \delta_{\mathbf{a}}))\}$.
- *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 $U(\delta', \mathbf{a}) >_{\mathbf{a}} U(\delta_{\mathbf{a}} \cup \delta''_{\mathbf{o}}, \mathbf{a})$.*

5.1 Example

We return to the example about international trade and we formalize it (see Figure 1). We have three agents in $NMAS$: the actor $\mathbf{a} \in RA$, the organization $\mathbf{o} \in SA$ and the role $\mathbf{r} \in RO$. Agent **a** attributes mental attitudes to the organization **o** ($\mathbf{o} \in A_{\mathbf{a}}$) and the organization attributes mental attitudes to the role **r** ($\mathbf{r} \in A_{\mathbf{o}}$).

The agent **a** can give a cheque ($cheque \in X_{\mathbf{a}}$), it desires not to give its money away ($\top \rightarrow \neg cheque \in D_{\mathbf{a}}$) and not to be sanctioned by agent **o** ($\top \rightarrow \neg s \in D_{\mathbf{a}}$).

The organization **o** believes that if agent **a** gives a cheque this counts as paying ($pay \in I$): $cheque \rightarrow pay \in B_{\mathbf{o}}$. Moreover, it does not desire to consider a violator ($V(\neg pay, \mathbf{a}) \in X_{\mathbf{o}}$) and to sanction agent **a** by doing $s \in X_o$ without motivation: $\{\top \rightarrow \neg V(\neg pay, \mathbf{a}), \top \rightarrow \neg s\} \subseteq D_{\mathbf{o}}$.

The role of buyer **r** is attributed by the organization the belief that the bill of lading ($bill \in P$) has been issued ($\top \rightarrow bill \in B_{\mathbf{r}}$) and that it means that the good has been shipped ($shipped \in I$ and $bill \rightarrow shipped \in B_{\mathbf{r}}$). Moreover it is attributed a goal: when the good has been shipped, then the role pays the fee for the good ($shipped \rightarrow pay \in G_{\mathbf{r}}$).

However, the agent **a** may not believe that the good has been shipped: $\{\top \rightarrow bill, \top \rightarrow \neg shipped\} \in B_a$.

The agent **a** enters the role **r** by signing a contract. The legal effects of the contract are, first, that the role playing relation PL is extended with the pair $(\mathbf{r}, \mathbf{a}) \in PL$; second, that the obligation to pay when the good is shipped is introduced: $O_{\mathbf{ao}}(pay, s \mid shipped)$. Note that the obligation is not directed towards the role **r**, but towards the agent **a** who is playing the role: agent **a** fears the sanction s, while the role is indifferent to it. The obligation is defined by a set of desires and goals: the normative goal distributed to agent **a** and desire that shipped goods are payed: $shipped \rightarrow pay \in D_{\mathbf{o}} \cap GD(\mathbf{a})$. 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: $V(\neg pay, \mathbf{a}) \rightarrow s \in D_{\mathbf{o}} \cap G_{\mathbf{o}}$; note that avoiding the sanction $\top \rightarrow \neg s$ is a desire of agents **a** and **o** as requested by the definition of obligation.

We adopt the perspective of agent **a** who has to decide whether to give a cheque for paying its fee or not. Personally, it does not believe that the good has been shipped, the bill of lading notwithstanding. The consequences of its decisions are: $out(B_{\mathbf{a}}, \{cheque\}) = \{bill, cheque, \neg shipped\}$ and $out(B_{\mathbf{a}}, \{\neg cheque\}) = \{bill, \neg cheque, \neg shipped\}$. Its unfulfilled motivational attitudes: $U(\{\neg cheque\}, \mathbf{a}) = \emptyset$, and $U(\{cheque\}, \mathbf{a}) = \{\top \rightarrow \neg cheque\}$ otherwise.

However, this reasoning is not sufficient to take a decision: agent **a** must recursively model the organization **o**'s decision. Agent **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 **o** is assumed to be taken considering the consequences of **a**'s behavior from the point of view of the role **r** it is playing and not from agent **a**'s own point of view.

The organization **o** has to decide whether **a**'s behavior respects the obligation or not; in the latter case agent **o** considers agent **a**'s behavior as a violation and sanctions it. So, given the decision $\delta_{\mathbf{a}} = \{\neg cheque\}$, the consequences for the organization of its decision $\delta_{\mathbf{o}} = \{\neg V(\neg pay, \mathbf{a}), \neg s\}$ and its unfulfilled mental attitudes are:

- from agent **a**'s point of view:
  $out(B_{\mathbf{o}}, \{\neg V(\neg pay, \mathbf{a}), \neg s\} \cup out(B_{\mathbf{a}}, \{\neg cheque\})) = \{bill, \neg cheque, \neg pay, \neg shipped, \neg V(\neg pay, \mathbf{a}), \neg s\}$
  $U(\{\neg V(\neg pay, \mathbf{a}), \neg s\} \cup \{\neg cheque\}, \mathbf{o}) \cap (D_{\mathbf{o}} \cup G_{\mathbf{o}}) = \emptyset$

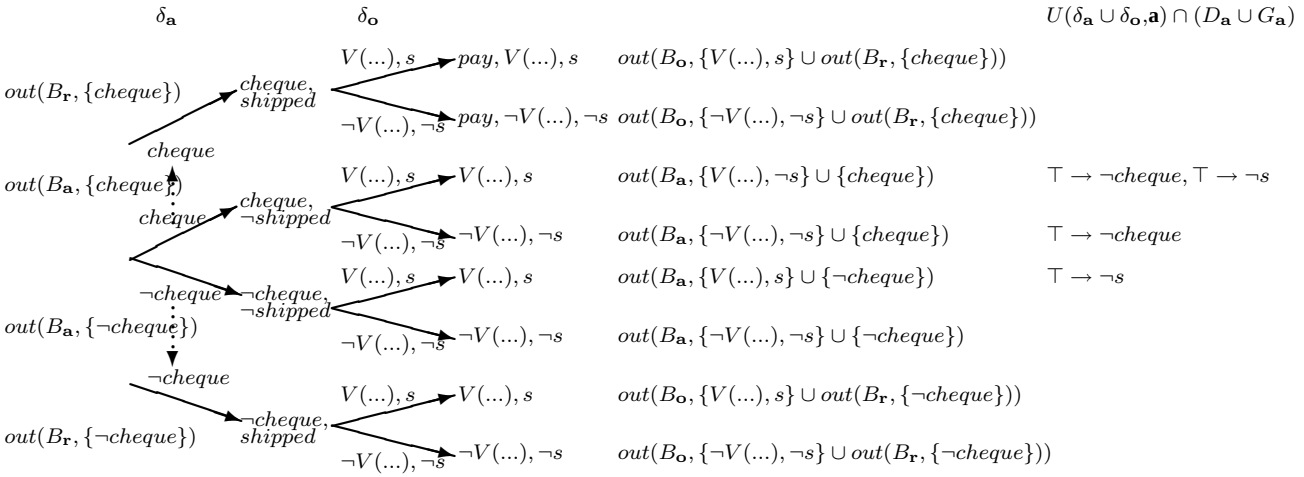

Figure 1: The game between the actor a and the organization o.

- from the point of view of the role **r**:

  $out(B_{\mathbf{o}}, \{\neg V(\neg pay, \mathbf{a}), \neg s\} \cup out(B_{\mathbf{r}}, \{\neg cheque\})) = \{bill, shipped, \neg cheque, \neg pay, \neg V(\neg pay, \mathbf{a}), \neg s\}$

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

Only from the point of view of the role **r** the organization can understand that some desires and goals remain unfulfilled: the agent **a** is not respecting the obligation related to the role it is playing; the organization judges agent **a**'s behavior not from the point of view of what it believes, but from the point of view of what it should believe while playing its role: agent **a** has not acted as if the good has been shipped, as it would follow from the role **r**'s beliefs: $out(B_{\mathbf{r}}, \{\neg cheque\}) = \{bill, shipped, \neg cheque\}$

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

So the optimal decision for the organization is to consider **a**'s behavior as a violation and to sanction it $\delta_{\mathbf{o}} = \{V(\neg pay, \mathbf{a}), s\}$, as the unfulfilled motivations are:
$U(\{V(\neg pay, \mathbf{a}), s\} \cup \{\neg cheque\}, \mathbf{o}) \cap (D_{\mathbf{o}} \cup G_{\mathbf{o}}) = \{shipped \rightarrow pay, \top \rightarrow V(\neg pay, \mathbf{a}), \top \rightarrow \neg s\}$

Instead, given the decision to give the cheque $\delta_{\mathbf{a}} = \{cheque\}$, the optimal decision of the organization is not to consider as a violation the behavior of agent **a** and not to sanction it. The consequences for the organization of its decision $\delta_{\mathbf{o}} = \{\neg V(\neg pay, \mathbf{a}), \neg s\}$ and the unfulfilled mental attitudes are (from role **r**'s point of view):
$out(B_{\mathbf{o}}, \{\neg V(\neg pay, \mathbf{a}), \neg s\} \cup out(B_{\mathbf{r}}, \{cheque\})) = \{bill, shipped, cheque, pay, \neg V(\neg pay, \mathbf{a}), \neg s\}$

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

Note that pay is for the organization the consequence of giving the cheque $cheque$: $cheque \rightarrow pay \in B_{\mathbf{o}}$; we say that $cheque$ for the organization counts as pay. So there is not a violation of the obligation $O_{\mathbf{ao}}(pay, s \mid shipped)$.

How does agent **a** take a decision between giving and not giving the cheque? From its own point of view, the consequences of beliefs and its unfulfilled motivations are:

- if $\delta_{\mathbf{a}} = \{\neg cheque\}$, then $\delta_{\mathbf{o}} = \{V(\neg pay, \mathbf{a}), s\}$:

  $out(B_{\mathbf{a}}, \{\neg cheque\} \cup \{(\neg pay, \mathbf{a}), s\}) = \{bill, \neg cheque, \neg pay, \neg shipped, V(\neg pay, \mathbf{a}), s\}$

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

- if $\delta_{\mathbf{a}} = \{cheque\}$, then $\delta_{\mathbf{o}} = \{\neg V(\neg pay, \mathbf{a}), \neg s\}$:

  $out(B_{\mathbf{a}}, \{cheque\} \cup \{\neg V(\neg pay, \mathbf{a}), \neg s\}) = \{bill, \neg shipped, cheque, pay, \neg V(\neg pay, \mathbf{a}), \neg s\}$

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

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

6. RELATED WORK AND DISCUSSION

According to Zambonelli *et al.* [33] "a multiagent system can be conceived in terms of an organized society of individuals in which each agent plays specific roles and interacts with other agents". For this reason in the last years many proposals on roles and organizations appeared in the multiagent field.

With respect to Sichman and Conte [30], we do not only attribute mental attitudes to roles, but we also consider them as autonomous agents. With respect to Dastani *et al.* [14, 15] we attribute to agents beliefs and we consider the problem of decision making when playing a role also in the situation in which the private goals of the actor are incompatible with those of the role. With respect to Pacheco and Carmo [28], we attribute beliefs and motivations to institutional agents. The representation and mandate problem is addressed in [11]. Moreover, the problem of the transmission of obligations is solved by considering obligations which are created at the moment an agent becomes an actor according to the goals attributed to the role. Finally, since we model roles as agents, we distinguish between obligations addressed to roles and obligations addressed to actors: the difference rests in the agent who is sanctioned.

In our model obligations are associated with roles since the specification of an organization cannot assume that all agents will stick to their role, as, in contrast, it happens in groups. This choice means that we identify organizations with burocracies in the classification proposed by Ouchi [27], while the notion of group in our model is

based on the notion of shared values. The issue of the relation between the goals attributed to roles and the agents' goals is related to the notion of agent type proposed in [12]: agents can be classified according to the priority they give to different components of mental attitudes. So it is possible to distinguish selfish agents who give priority to their own goals, cooperative agents who give priority to the goals of the group [9] and trusted agents who give priority to the goals of the roles they play. The association of obligations and roles is necessary only in case of selfish agents. A similar distinction is made also by [14].

The fact that the actors may not adopt the goals of the roles they play raises the problem of the discrepancy between the specification of the organization in terms of goals and the actual goals of the agents composing the multiagent system. This possible discrepancy fits the idea of some economists who argue that the functioning of an organization is based also on habits which are not accounted for by the specification of the organization [31]. Taking into account the possibility that the behavior stemming from the agents' goals is different from the desired one is necessary to diagnose the malfunctioning of an organization [22, 26].

It must be noted that "an organization is more than simply a collection of roles (as most methodologies assume) [...] further organization-oriented abstractions need to be devised and placed in the context of a methodology [...] As soon as the complexity increases, modularity and encapsulation principles suggest dividing the system into different suborganizations" [33]. According to Jennings [21], however, most current approaches "possess insufficient mechanisms for dealing with organisational structure". Moreover, what is the semantic principle which allows decomposing organizations into suborganizations must be still made precise. The current work is elaborated in [10] by introducing besides roles further concepts to structure an organization using the agent metaphor.

7. CONCLUSION

In this paper we address the problem of defining the notion of role in organizations. Using the methodology of attributing mental attitudes to social entities we show that the beliefs of roles correspond to their expertise and their goals to their responsibilities. The actors playing roles take decisions playing a game with the organization they are employed in: they know that if they do not fulfill the obligations related to their role they are sanctioned or not rewarded.

This paper builds on our work on normative multiagent systems. Starting from Boella and Lesmo [3]'s observation that a normative system behaves like an agent who monitors and sanctions violations, we developed 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 [4] and constitutive rules as beliefs of the normative systems [11]. This reduction distinguishes our approach from other models of constitutive rules in 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 system as any other belief.

Future work concerns the assignment of a role to an agent and, in particular, how the obligations addressed to actors are created. In [7] we propose contracts as a solution to this problem and to the related problem of how an agent exits from a role. Finally, the distribution of the obligations addressed to organizations or to sets of roles can be analyzed in terms of a negotiation process, as we discuss in [8].

8. REFERENCES

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

[2] A. Artikis, J. Pitt, and M. Sergot. Animated specifications of computational societies. In *Procs. of AAMAS'02*, pages 1053–1061. ACM press, 2002.

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

[4] G. Boella and L. van der Torre. Attributing mental attitudes to normative systems. In *Procs. of AAMAS'03*, pages 942–943. ACM Press, 2003.

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

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

[7] G. Boella and L. van der Torre. Contracts as legal institutions in organizations of autonomous agents. In *Procs. of AAMAS'04*, pages 948–955, New York, 2004.

[8] G. Boella and L. van der Torre. The distribution of obligations by negotiation among autonomous agents. In *Procs. of ECAI'04*, pages 13–17. IOS Press, 2004.

[9] G. Boella and L. van der Torre. Groups as agents with mental attitudes. In *Procs. of AAMAS'04*, pages 964–971, New York, 2004.

[10] G. Boella and L. van der Torre. Organizations as socially constructed agents in the agent oriented paradigm. In *Procs. of ESAW'04*, Berlin, 2004. Springer Verlag.

[11] G. Boella and L. van der Torre. Regulative and constitutive norms in normative multiagent systems. In *Procs. of KR'04*, pages 255–265, 2004.

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

[13] L. Cavedon and L. Sonenberg. On social commitments, roles and preferred goals. In *Procs. of ICMAS'98*, pages 80–87, Berlin, 1998. Springer Verlag.

[14] M. Dastani, V. Dignum, and F. Dignum. Role-assignment in open agent societies. In *Procs. of AAMAS'03*, pages 489–496, Melbourne, 2003. ACM Press.

[15] M. Dastani, B. van Riemsdijk, J. Hulstijn, F. Dignum, and J.-J. Meyer. Enacting and deacting roles in agent programming. In *Procs. of AOSE'04*, New York, 2004.

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

[17] M. Esteva, J. Padget, and C. Sierra. Formalizing a language for institutions and norms. In *Intelligent Agents VIII: Procs. of ATAL'01*, pages 348–366, Berlin, 2001. Springer Verlag.

[18] M. Fasli. On commitments, roles and obligations. In *From Theory to Practice in Multi-Agent Systems, LNAI 2296*, pages 93–102, Berlin, 2001. Springer-Verlag.

[19] E. Goffman. *Strategic interaction*. Basil Blackwell, Oxford, 1970.

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

[21] N. R. Jennings. On agent-based software engineering. *Artificial Intelligence*, 117(2):277–296, 2000.

[22] M. Kalech and G. Kaminka. On the design of social

diagnosis algorithms for multi-agent teams. In *Procs. of IJCAI'03*, pages 370–375, 2003.

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

[24] G. Lindemann and I. Munch. The role concept for agents in multi-agent systems. In *Procs. of MASHO Workshop*, 2001.

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

[26] R. Micalizio, P. Torasso, and G. Torta. On line monitoring and diagnosis of multi-agent systems: a model based approach. In *Procs. of ECAI'04*, pages 848–852. IOS Press, 2004.

[27] W. Ouchi. A conceptual framework for the design of organizational control mechanisms. *Management Science*, 25(9):833–848, 1979.

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

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

[30] J. Sichman and R. Conte. On personal and role mental attitudes: A preliminary dependence-based analysis. In *Procs. of SBIA'98 Conference*, pages 1–10, 1998.

[31] P. Terna. Economic simulations in swarm: Agent-based modelling and object oriented programming - a conceptual framework for the design of organizational control mechanisms. *The Electronic Journal of Evolutionary Modeling and Economic Dynamics*, 1013, 2002.

[32] R. Tuomela and K. Miller. Groups beliefs. *Synthese*, 91:285–318, 1992.

[33] F. Zambonelli, N. Jennings, and M. Wooldridge. Developing multiagent systems: The Gaia methodology. *IEEE Transactions of Software Engineering and Methodology*, 12(3):317–370, 2003.