

# What's in it for me? Agent deliberation on taking up social roles\*

Virginia Dignum, Frank Dignum  
Institute of Information and Computing Sciences  
Utrecht University, The Netherlands  
{virginia, dignum}@cs.uu.nl

**Abstract** – *Open systems are characterized by heterogeneous participants which can enter or leave the system at will. Agents (e.g. personal assistants for buying things on the Internet) will only temporarily take up roles (e.g. a buyer in on-line auctions). Taking up a role in a society brings with it norms, obligations and objectives that the agent must decide on how to realize. On the other hand, the agent's own norms will dictate how the agent will fulfill its obligations with respect to the society, and how the agent acts in case of violation. This paper focuses in exploring how the goals of an agent arise from the various sources of motivation to which the agent may be subject. Furthermore, we'll consider how the agent can possibly negotiate social changes that increase its utility and, in the agent's eyes, the overall utility of the society.*

## 1 Introduction

Currently, in most MAS, agents are simply designed from scratch so that their behavior complies with the behavior described by the role(s) it will take up in the society. Such solution is not applicable to open systems, which assume the heterogeneous agents designed independently from the society framework. Typical examples are e-commerce applications or information agent systems. The aims and requirements of the shop owners determine the design and functionality of a web shop. However, participating agents (e.g. personal assistants for buying things on the Internet) will be designed from the perspective of the individuals that own them, and take temporarily the role of e.g. a buyer in that shop. Through a deliberation process, an agent will determine the utility for itself of taking up a role in an agent society. For instance, its utility will increase if the objectives associated with a role contribute to the agent's own goals, and the norms of the role do not conflict with its own norms. Comprehensive solutions for this point require complex agents that are able to reason about their own objectives and desires and thus decide and negotiate their participation in a society. A first step on the road to this solution (cf. [4]) is to have a formalism to compare the specification of agents and roles and determine whether an agent is suitable to enact a role.

Another important aspect concerning role enacting agents is that of modifying the agent to include the characteristics of the assumed role(s). A possible solution for this point has been proposed in [9], which extends agents with an interface to the society. This interface prevents any action not allowed by the role definition. However, it does not ensure the proactive behavior expected from the role and is not flexible enough to incorporate different enacting styles, capabilities and requirements of the agents. It actually makes the actual agent 'invisible' to the society and only its enactment of the role behavior apparent in the society. We think that the consequence of an agent adopting a role is more drastic than this. The actual agent behavior must often be modified according to the objectives, norms and rights specified by the role.

In this paper, we assume that agents have goals of their own (e.g. gotten by design, or by participation in other societies), and are able to form (either by design or by deliberation) plans to achieve those goals. We further assume that society models, describe agent societies from a global perspective, rather than from the perspective of the individual agents. The OperA model [5] will be used as illustration for the specification of organizations.

The paper is organized as follows. In section 2 we give the descriptions and definitions of agents and roles. In section 3 we describe the different ways in which an agent can fulfill a role. In section we discuss possible points of negotiation between the agent and the society when taking up a role. Finally we give some conclusions in section 5.

## 2 Agents and roles

In [5] we presented the OperA model for agent societies that separates organizational design, reflecting the aims and requirements of the organization, from the individuals that animate it. Organizational design can be more or less elaborate, depending on the type of society. It describes the different roles participating in the society and the way interaction between roles is expected to happen in order for the society goals to be realized. Agents are designed independently of the society as avatars for different systems, individuals or groups. An agent seeks admittance to a society at a given moment in its life-cycle if it perceives that participation will increase in some way its utility, that is, contributes to the

realization of its goals. In our model a population of the society is described through contracts that represent the commitments of agents concerning role enactment, which enable the verification of the society behavior. We will not give a complete description of the organizational model or even the contracts, but will concentrate on those elements of a role that form essential elements for forming such a contract.

A **role** is the abstract representation of a policy, service or function. Role descriptions identify necessary activities and services necessary to achieve society goals and enable us to abstract from the individuals that eventually will perform the role. Furthermore, roles must describe the necessary capabilities that must be enacted by any agent pretending to play that role. Roles interact with each other in different interaction scenes. Typically, interaction is described using "landmarks" that specify the main features (conditions, obligations, and possibly partial plans) of interaction between roles. For the sake of simplicity, in this paper we assume that a society structure is completely defined by a set of roles. For a complete specification of the organizational model we refer the reader to [5].

For the objectives of this paper, we define a role through its set of objectives, plans (to attain some of its objectives), its norms, and the ordering of its objectives. The objectives of a role describe the results that agents playing the roles must seek to obtain. The plans can be seen as actions that agents playing the role are allowed (and capable) to perform. Norms of a role specify the duties and privileges of agents enacting the role.

**Definition 1 (Role)**

A **role**  $r$  is a tuple  $\langle G_r, N_r, P_r, \succ^{G_r} \rangle$ , where  $G_r$  is the set of objectives that should be achieved when enacting the role,  $N_r$  is the set of norms that need to be satisfied when enacting the role, and  $\succ^{G_r}$  is the ordering on the objectives of  $r$ .  $P_r$  is the set of plans  $p^{g_r}$  for the objectives of  $r$ , defined as  $p^{g_r} = \{p_i \subseteq P_r : \exists g_r \in G_r, [p_i]g_r\}$ .

The informal meaning of  $[p]g$  is that  $g$  is true after  $p$  has been realized.

**Agents** are active entities that are able to enact roles described in a society model. Agents join a society by adopting some of its roles. In order to be able to take up a role in the society, it is necessary to alter (extend, modify or limit) the agent's own behavior such that it will react within the society in ways that are in accordance to the society's expectations of the role. We assume that agents are designed outside the scope of the society, and have their own goals, beliefs, reasoning capabilities and behavior rules. The role(s) an agent plays determine the current behavior and actions of the agent, and will influence the possibilities of further action of the agent in the society.

Similarly to the agent definition, an agent is defined through its set of goals, its belief base, and its ordering on the goals.

**Definition 2 (Agent)**

An **agent**  $a$  is a tuple  $\langle G_a, B_a, N_a, P_a, \succ^{G_a} \rangle$ , where  $G_a$  is a set of goals,  $B_a$  is a set of beliefs,  $N_a$  the set of agent norms

and  $\succ^{G_a}$  is the ordering on the goals of  $a$ . As in definition 1,  $P_a$  is the set of plans  $p^g$  for the goals of  $a$ , defined as the plans  $p^g = \{p_i \subseteq P_a : \exists g_a \in G_a, [p_i]g_a\}$ .

For the purpose of this paper, we consider that agents have no norms of themselves, that is  $N_a = \{\}$ . The case of normative agents is more complex, because agent and role norms will influence the reasoning about and behavior of role enactment. This will be object of future research. The main question in this paper is: can any agent enact any role in any way it wants? And if not, which are the conditions under which role enactment is possible for an agent? An appropriate enacting relation presupposes that both the agent and the role are internally coherent, that is, that there are no internal conflicts between its components and that the agent goals are achievable. Informally, internal coherence of a role or agent is as in definition 3. The formal definitions are presented in [4].

**Definition 3 (Internally coherent role, resp. agent)**

Given a role/agent, described by its objectives/goals, plans and norms, we say that the role/agent, is internally coherent iff:

1. The objectives/goals of the role/agent are not conflicting
2. The objectives/goals and norms of the role/agent do not conflict
3. Sub-plans in the same plan do not conflict
4. For each objective/goal for which a plan is specified, the objective/goal does not conflict with the plan.

Given an internally coherent agent and an internally consistent role, we must now describe the conditions under which it is possible for the agent to fulfil the role. In the following, we consider two relationships between agents and roles. The first is called compatibility, and is based on a subset relation between the agent and the role. Intuitively an agent is compatible with a role when the goals of the agent are a subset of the objectives of the role, that is, the agent naturally fulfils (part of) the objectives or sub-objectives of the role. For example, an agent that has as only goal the goal of reading a certain paper is compatible with the PC-member role that has the objective of reviewing that paper. In the same way, a role is compatible with an agent when the role objectives are a subset of the goals of the agent. Compatibility indicates that the agent is suitable to fulfil the role. However, such ideal match is often not possible. We therefore introduce a weaker relation between agents and roles. This relation, consistency, indicates that the characteristics of the agent and the role do not oppose each other.

Informally, an agent is consistent with a role if the goals and plans of the agent do not conflict with the objectives, sub-objectives and norms of the role. Similarly, a role is consistent with an agent if the objectives, sub-objectives and norms of the role do not conflict with the goals and plans of the agent. Consistency indicates that it is possible for the

agent to fulfil the role. A formal treatment of conflict and consistency between plans and goals can be found in [3, 6].

**Definition 4** (*Compatibility and consistency of agents and roles*). Given an internally coherent agent  $a$ , and an internally coherent role  $r$ :

1. Agent  $a$  is compatible with role  $r$ , if the goals of  $a$  are a subset of the objectives of  $r$ , and all plans of  $a$  can be formed using plans or sub-plans of  $r$ .
2. Role  $r$  is compatible with agent  $a$ , if the objectives of  $r$  are a subset of the goals of  $a$ , and all plans of  $r$  can be achieved using the plans or sub-plans of  $a$ .
3. Agent  $a$  is consistent with role  $r$  if the goals and plans of the agent do not conflict with the objectives and plans of the role.

Using point 1 in definition 4 above, it can be guaranteed that the agent will only achieve results that are in accordance with the role objectives (all what the agent aims for, is indeed also an objective of the role). In the converse case, that is, if the objectives of the role would be a subset of the goals of the agent, there would be valid and rational plans of the agent that would not guarantee the achievement of the objectives of the role. An agent that is neither coherent nor consistent with a role has apparently goals and planning rules that, when enacting the role, may violate some norms that are associated with the role. A critical case is when an agent is consistent with a role but not compatible. In such a case, the agent has apparently some additional goals or planning rules that are not associated with the role. Although this agent will not violate any norms associated with the role, the agent may use its own preference ordering to achieve its own goal preferably and thus ignore the goals that are prescribed by the role. It is clear that in such a case, the agent enacts the role inappropriately. A similar situation can occur even when the agent and the role are compatible. In such cases, the appropriate enacting relation is not guaranteed since the enacting relation depends on the ordering on goals and roles as well. An agent can in fact use its own preference relation and thereby never achieve the goals associated to a role.

### 3 Role enactment

From the perspective of an (OperA) society, it is up to the agent how to manage and prioritize its own goals. That is, by assuming a role, the agent will receive the objectives from that role. How the agent will handle those objectives, whether it interprets them as goals or as norms, what priority it gives them, is up to the agent self. However, the society model is based on the assumption that agents that take up roles are expected to eventually realize the assumed objectives.

The above classification is based on the assumption that the ordering on goals and rules of the enacting agent is in accordance with the ordering that is prescribed by the role. This assumption can be relaxed in which case the agent can

either use its own ordering, the ordering of the role, or a combined ordering. Moreover, we may assume an ordering on the obligations and prohibitions associated to the role. In general, the possible choices to use these orderings result in a variety of agent types. In the case of conflicting orders, a social agent will adopt the order associated to the role and not its own. In contrast, a selfish agent will use its own order rather than the order associated with the role. We conclude this section with a proposition that indicates some relations between the different relations between agents and roles.

**Proposition 1**. Let  $a$  be an internally coherent agent and  $r$  be an internally coherent role such that  $rea(a, r, s)$ , for a scene  $s$ . Then:

1.  $a$  is compatible with  $r$ , implies  $a$  is consistent with  $r$ .
2. If  $a$  is consistent with  $r$ , then no violation of  $r$  can occur when  $a$  enacts  $r$ .
3. Conversely, if  $a$  is not consistent with  $r$ , then violation of  $r$  can occur when  $a$  enacts  $r$ .
4. If  $a$  is compatible with  $r$ , then  $a$  can only maximally socially enact  $r$ .
5. If  $a$  is compatible with  $r$  and  $a$  is a social agent, then the enacting agent does not violate any norm that is associated with the role  $r$ .
6. If  $a$  is consistent with  $r$  and  $a$  is a social agent, then the enacting agent does not violate any norm that is associated with the role  $r$ .
7. If  $a$  is a maximally social agent, then  $a$  will never violate  $r$ , whether or not  $a$  and  $r$  are consistent and/or compatible.

#### 3.1 Social attitudes

Agent literature discusses extensively different types of social attitudes of agents: selfish, altruistic, honest, dishonest, etc [1, 8, 7]. Different types of agents result in different role performances, because the way an agent will plan its goals, is dependent on his social attitude, which influence the realization of its role objectives and the fulfillment of the role norms. For instance, some types of agents will only attempt to achieve the goals of their adopted roles and forget their own private goals, while others will only attempt to achieve the goals from the role after all their own goals have been satisfied. Furthermore, the relations between agent plans and role objectives, and of agent goals and role sub-objectives must be considered, as well as the influence of the role norms on the behavior of agents.

Given an internally coherent agent  $a$  and an internally coherent role  $r$ , we should consider what it means for agent  $a$  to enact role  $r$  appropriately, that is, in a way that meets the expectations of the society. The most simple case is that of total adoption, that is, when agent  $a$  enacts role  $r$ ,  $a$  will adopt all the goals and the norms associated with  $r$ . Agent  $a$  will also

include the norms of the role in its own model. These will trigger the agent to fulfill the obligations of the role. In addition, the agent can keep some or all of its own goals and rules, as long as it keeps its internal coherence. Furthermore, in order to achieve its goals, the agent must select and plan them. This selection mechanism is usually based on the ordering on goals and planning rules. When  $a$  adopts the goals and rules of a role  $r$ ,  $a$  must also extend its orderings to include the goals and rules of the role. Of course, this can be done in many different ways which indicate how the agent assumes the role. For instance, this ordering can give preference to the agent's own goals (a selfish agent), or to the goals of the role (a social agent). In the most simple case, such ordering will be imposed on the agent by the society. Nevertheless, we must consider the case of agents that are able to reason and negotiate about combining these orderings.

In summary, the following basic types of role enactment by the agents can be distinguished [4]:

1. **Social enactment:** The agent includes as many of its own goals as possible, but gives priority to the objectives of the role over its own. The agent gives priority to the plans of the role over its own plans, which are only used if role sub-objectives cannot be used.
2. **Maximally social enactment:** The agent only uses the objectives from the role and ignores its own goals, for the duration of the role enactment. The agent only uses the plans from the role and ignores its own plans.
3. **Selfish enactment:** The agent includes as many of the goals of the role as possible and gives priority to its own goals over the objectives of the role. The agent includes as many of its own plans as possible and gives priority to its own plans over those of the role.
4. **Maximally selfish enactment:** The agent only uses its own goals, and ignores any objectives of the role. The agent only uses its own plans, and ignores those of the role.

Formally, we define a role enacting agent as:

**Definition 5** (Role Enacting Agent)

Given an agent  $a = \langle G_a, B_a, P_a, \succ^{G_a} \rangle$  and a role  $r = \langle G_r, N_r, P_r, \succ^{G_r} \rangle$ , the role enacting agent  $r_a$  is defined as a tuple  $\langle G, B_a, N_r, P, \succ^G \rangle$  where the sets of goals and plans, and the preference relation depend on the role enactment type, as follows:

1. **Social enactment:**  $G = G_r \cup \max(G_a)$ , such that  $G$  is consistent,  $P = P_r \cup \max(P_a)$ , such that  $P$  is consistent,  $\succ^G = \succ^{G_r}$  and  $g_r \succ^g g_a, \forall g_r, g_a \in G$ .
2. **Maximally social enactment:**  $G = G_r$ ,  $P = P_r$  and  $\succ^G = \succ^{G_r}$ .
3. **Selfish enactment:**  $G = G_a \cup \max(G_r)$ , such that  $G$  is consistent,  $P = P_a \cup \max(P_r)$ , such that  $P$  is consistent,  $\succ^G = \succ^{G_a}$  and  $g_a \succ^g g_r, \forall g_a, g_r \in G$ .

4. **Maximally selfish enactment:**  $G = G_a$ ,  $P = P_a$  and  $\succ^G = \succ^{G_a}$ .

### 3.2 Individual motivation

Agents will enact roles if those contribute for an increase of their individual utility. On the other hand, often role enactment brings along extra commitments and/or constraints to the activity of the agent. A rational agent will determine the utility gain based on its own 'personality'. Based on [8], possible motivations for enacting a role are:

- **Personal enrichment by role playing:** The agent did not have a plan on how to achieve one of its goals, which is provided to it by playing the role. In this case, it is the agent that profits from the society. Formally:  $\exists g \in G_a : \neg(p_g \in P_a) \wedge \exists p \in P_r : [p]g$ .
- **Increasing power by role playing:** The role provides the actor with a 'better' way to realize its goals, for instance, the rights associated with the role enhance the activity of the agent. Formally,  $\exists g \in G_a \wedge \exists p_g \in P_r : p_g \succ^G p_{g_a}, \forall p_{g_a} \in P_a$ .

The agent might also need to adjust its priorities to fit the requirements of the role. That is, the obligations arising from the role to be enacted influence an agent's prioritizing of its goals [2]. Norm-related considerations for an agent to decide on role enactment can be grouped along the following areas:

- **Diminishing power by role playing:** The society limits the possibilities of activity of the agent. Some plans the agent may have had are not applicable within the role description. Formally,  $\exists p_a \in P, g \in G : [p_a]g \wedge N \rightarrow Fp_a$ , where  $F$  is the prohibition operator.
- **Limitation of personal goals:** The role norms forbid that some of the goals of the agent cannot be achieved within the society. Formally,  $\forall p \in P, g_a \in G : [p]g_a \wedge N \rightarrow Fp$ .
- **Extension to personal goals:** The role norms create some extra goals for the agent, which it did not have previously. Formally,  $\exists g, N \rightarrow g : g \notin G_a \wedge g \in G_{a_r}$ .
- **Alteration of personal plans:** The norms of the role cause an alteration on the plans of the agent; these must be extended with extra activity, some of the actions cannot be taken or the order in which actions were planned is not allowed. Formally,  $\exists p_1, p_2 : p_1 \succ^{G_a} p_2 \wedge N \rightarrow Fp_1$ .

Possible objections for role enactment, concern limitations of agent behavior, or, conversely, excessive added obligations. On this aspect, role norms play an important role, as they may affect the behavior of the agent playing the role. Of course, depending on the beliefs of the agent, these possible objections, might also be perceived as motivations be the agent.

### 3.3 Society motivation

In the same way as individual agents, societies will also attempt to choose the most adequate agent to enact a society role. Agent adequacy in this sense means the agent whose role enactment will most contribute to the utility of the society. In this paper, we consider enactment from the perspective of an individual agent. In this sense, increase of society utility is not relevant, as agents are not necessarily interested in the utility of the society but just in their own utility. More research on the motivations of the society itself is needed. A possible motivation for the society side is the following:

- **Role enrichment by personal plans:** When the role has an objective for which no plan is specified, and the agent playing the role has a plan on how to achieve that role objective. This shows a good adequacy of the player to the role, from the society viewpoint (the actor really adds something to the society activity). Formally,  $\exists g \in G_r : \neg(p_g \in P_r) \wedge \exists p \in P_a : [p]g$ .

Given that, in this paper, we assume that agents have no norms of themselves, we do not consider the case of agents that have norms of themselves, we do not evaluate how agent norms affect expectations on the role. The case of normative agents will be the subject of further research.

## 4 Negotiating change

In the previous section we described different ways in which an agent can enact a role and also indicated some reasons why an agent would decide to enact a role and a society would accept that agent enacting that role. The next step is to check the possibilities for negotiating the conditions under which a role is enacted. First of all we have to check the function of a role in the society. If the role has a crucial function on which the goal of a society depends the agent has more leverage in negotiating the conditions. E.g. the role of program chair is crucial for a conference organization. The role of reviewer is important but because there will be many agents enacting that role it is less crucial. Finally the role of an author is even less important (assuming it is a big conference with many submissions). In the last case an agent that wants to enact the role of author usually cannot negotiate the deadline for submitting a paper. The basic idea behind this is that the conference goal of getting a high quality program does not depend (a lot) on the submission of one paper. Thus the utility of the society does not rise considerably with that submission. Therefore the society is not willing to give anything in return (like a deadline extension) for a submission. The role of a reviewer is more crucial for the goals of the society. If a famous agent is willing to be in the PC of the conference it is good publicity and therefore will attract high quality papers. This in turn contributes to the goal of a high quality program. Therefore the society will be willing to negotiate with this agent. E.g. the agent can negotiate a maximum number of papers to review. However, the boundaries are still such that the remaining objectives of the role contribute to the society and the agent and role are consistent.

E.g. the society will not accept that the agent does no reviewing at all or uses the material of the reviewed papers to write his own papers.

It should be noted that the society does not know the agent's goals and plans. Therefore the society has no direct way of checking on forehand whether an agent will perform all functions of the role in a proper way. For non-crucial roles the norms of the role will be specified in a way that the enacting agent cannot perform any unwanted action. Whether it actually performs the wanted actions of the role is of no concern to the society. E.g. whether an agent will actually submit a paper or not after enacting the author role is not important. But this agent should not be allowed to ever review its own paper.

In order for the society to check whether an agent will also perform the wanted actions of a role it can demand a proof of the capabilities of the agent. E.g. it should provide a possible plan or schedule for the execution of the tasks of the role. In principle the society can also demand the agent to commit to giving the objectives of the role the highest priority or fix deadlines for the performance of the tasks of the role in the contract of the agent.

Finally it should be remarked that the amount of agents that apply for enacting a role has of course a large influence on the negotiation as well as the number of options for an agent to fulfill its own goals. As in any other negotiation these elements determine the negotiation power of the parties. E.g. if a society has a choice between agents for the enactment of a role in which one agent is compatible while the other is only consistent with the role the society will always choose the compatible agent and does not negotiate with the consistent agent about enactment conditions.

## 5 Conclusions

In this paper we argued that in open agent systems the agents can enact roles in different ways. Both the agent as well as the society should benefit from the enactment of the role by the agent. The agent can benefit from it because enacting the role gives him means to accomplish a goal that could not be attained otherwise. The society can benefit because the agent will achieve some of the objectives of the role that are contributing to the goals of the society. How much each party benefits depends largely on how the agent enacts the role. We have shown that even when the agent's goals and roles objectives are consistent there exist many ways to enact the role.

Finally we have discussed a few points on the negotiation process between the agent and the society. In principle one could use the theories on multi-attribute negotiation to devise strategies for agents and society to achieve the best results. We leave further exploration of this issue for future research.

## References

- [1] C. Castelfranchi. Commitments: from individual intentions to groups and organizations. In V. Lesser, editor,

*Proc. 1st. International Conference on Multi-Agent Systems (ICMAS'95)*, pages 41 – 48. MIT Press, 1995.

- [2] L. Cavedon and L. Sonenberg. On social commitments, roles and preferred goals. In *Proc. of the Third Int. Conference on Multi-Agent Systems (ICMAS-98)*, pages 80–87. IEEE Computer Society, 1998.
- [3] M. Dastani, V. Dignum, and F. Dignum. Organizations and normative agents. In *Proc. Eurasia-ICT, First Eurasian Conference on ICT*, Shiraz, Iran, October, 2002.
- [4] M. Dastani, V. Dignum, and F. Dignum. Role assignment in open agent societies. In *Proc. AAMAS'03*, 2003.
- [5] V. Dignum. *A Model for Organizational Interaction: based on Agents, founded in Logic*. SIKS Dissertation Series 2004-1. SIKS, 2004. PhD Thesis.
- [6] L. Padgham J. Thangarajah, M. Winikoff and K. Fischer. Avoiding resource conflicts in intelligent agents. In F. van Harmelen, editor, *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI-2002)*. IOS Press, 2002.
- [7] M. Miceli, A. Cesta, and P. Rizzo. Distributed artificial intelligence from a socio-cognitive standpoint: Looking at reasons for interaction. *Artificial Intelligence and Society*, 9:287 – 320, 1996.
- [8] J. Sichman and R. Conte. On personal and role mental attitudes: A preliminary dependence-based analysis. In F. Oliveira, editor, *Advances in AI*, volume LNAI-1515, pages 1–10. Springer-Verlag, 1998.
- [9] W. Vasconcelos, J. Sabater, C. Sierra, and J. Querol. Skeleton-based agent development for electronic institutions. In *Proc. AAMAS'02*, 2003.