Reasoning about the dynamics of social behaviour

Maria Fasli
University of Essex
Department of Computer Science
Wivenhoe Park
Colchester CO4 3SQ,UK

ABSTRACT

Formal theories of multi-agent systems require a rich ontology for modelling the dynamics of social behaviour. In this paper a formal analysis of the social behaviour of individual and social agents within the BDI paradigm is provided. The central idea behind this approach is that stability and regulation of activity within a group of agents can be accounted for by means of a complex web of roles, commitments obligations and rights. In particular, collective commitments are considered to be the attitudes that hold a group of agents together. In pursuit of their own objectives as well as in order to support their collective commitments, agents adopt roles and undertake social commitments. Being semi-autonomous they may decide to drop their commitments and roles, but they may have to bear the consequences of the other agents' prerogative to exercise their rights.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*; F.4.1 [**Mathematical Logic and Formal Languages**]: Mathematical Logic—*Modal logic*

General Terms

Theory

1. INTRODUCTION

Multi-agent systems have become increasingly popular as a means of providing solutions to inherently complex and distributed problems that require the cooperation and coordination of a number of loosely-coupled individual agents [19, 31]. Consequently, theoretical and practical research in multi-agent systems needs to address issues such as stable group activity and regulation of behaviour that arise in cooperative problem solving and teamwork. From the theoretical perspective these issues need to be investigated in the context of formal theories of agents. Most theoretical models view agents as *intentional systems* that are characterised by a mental state. The most prominent such framework is the BDI paradigm [3, 25, 26, 35] which models agents as having beliefs, desires and intentions. Reasoning about cooperative activity and teamwork in the context of such theories requires a very rich ontology of social and collective attitudes such as mutual beliefs and intentions, commitments as well as normative concepts such as obligations and rights.

Most traditional work in this area has concentrated on the concept of commitment [19, 29]. As noted by Castelfranchi [4], commitments are very important since they link an individual agent's activity with the group's objectives. Works addressing collective attitudes such as joint intentions that lead to teamwork and cooperative problem solving include [8, 33, 34]. Joint intentions and social plans have been considered in [27] in an extension of the original BDI paradigm to the multi-agent case. Another extension of the BDI model considers social and collective commitments [12] in which a collective commitment is the strongest notion of teamwork. Recently, normative concepts have also been explored in the context of collective activity. Krogh [20] argued in favour of obligations in multi-agent systems, traditionally the subject of deontic logic [1, 17], in order to formalise normative aspects of the agents' behaviour. Works that consider issues regarding obligations and commitments include [11, 10]. Other approaches to obligations and other normative concepts such as rights and duties include [16, 22, 32].

These studies are informative and they have offered important insights into cooperative problem solving and teamwork. However, the analysis of commitment provided, which is considered to be the cornerstone of collective activity, is insufficient and unsatisfactory. In addition, there seems to be confusion between the concepts of social and collective commitment; the terms are used interchangeably to describe two different things while some researchers use one term when they actually mean the other [4]. These theoretical models tackle the problem from different perspectives and they offer complementary, but not comprehensive views of collective activity; none of them covers the full spectrum of social and collective attitudes that are typically involved in teamwork. For instance, works that deal with commitments very often ignore the relevance of normative concepts such as obligations, while other works that examine joint intentions or goals do not explicitly consider commitments or obligations. In addition, very few studies have attempted to develop models of collective activity using organisational approaches. Concepts such as roles are used in a very informal way and practically there has been little effort to incorporate organisational terms and concepts into multi-agent systems with the exception of works such as [28, 6, 5] and [24, 9].

In this paper a formal analysis of the dynamics of social behaviour of individual and social agents within the BDI paradigm is presented. The paper is organised as follows. Next the basic ideas behind this approach are discussed. The following section describes the basics of the multi-modal logical framework that is based on the classical BDI paradigm. An analysis of the primary individual and group attitudes follows. Next a formal analysis of the basic ingredients of the theory starting from obligations, rights and preferences and moving subsequently to commitments and roles is presented. Finally the paper ends with a summary and pointers to future work.

2. SOCIAL DYNAMICS

The central idea behind the approach adopted here is that stability and regulation of activity within a group of agents (team or organisation) can be accounted for by means of a complex web of roles, commitments, obligations and rights. There are two types of agents in a multi-agent environment: individual and social. An individual agent is simply an agent, whereas a social agent consists of other individual or social agents; hence a team and an organisation are both social agents. In the discussion that follows the terms "group" and "social agent" are used interchangeably.

Stability and fairness in a multi-agent environment is crucial and as a consequence some form of general rules and norms should restrict the behaviour of individual agents. These general obligations express normative sentences and can be seen as rules that provide the minimal means of social interaction. They express what ought to be the case for all agents and they are impersonal, that is no explicit reference is being made to a particular agent.

While in pursuit of its own objectives an individual agent may decide to join groups and thus engage in teamwork and cooperative problem solving. As has been extensively argued in the literature, commitments play a very important role in such activities [4, 34]. In particular, the position that collective commitments are the attitudes that hold a group of agents together is endorsed here. These are the internal commitments of a group and they can be viewed as expressing the purpose and objectives of the whole group. Collective commitments depend on other group attitudes such as mutual beliefs and intentions.

In order to achieve their individual objectives as well as to support their collective commitments, agents adopt roles and undertake social commitments. When an agent joins a social agent, it assumes or is given a specific role. This role entails a set of commitments, obligations and rights, and specifies the position of the individual in the social agent as well as its commitments towards the group and the rest of the individual and social agents. Each member of the social agent knows its place and acts accordingly and furthermore each knows the implications of exercising rights and breaking commitments. Individual and social agents may adopt roles in relation to other individual or social agents. However, when an agent adopts a role, it does not necessarily mean that it has to adhere to it forever. Circumstances may arise when an agent may decide to abandon a role, although this may not be without consequences. Moreover, agents may hold different roles in different groups and as a result conflicts of interest may arise. Futhermore, roles may also be associated with authority relations thus giving rise to a notion of power.

In addition to social commitments resulting from the adoption of roles, agents take up social commitments of their own accord as a result of promises towards other agents. Often there is confusion in the use of the terms "social" and "collective" commitment, and they are regarded as meaning one and the same thing or they are used interchangeably. However, we consider the notion of a collective commitment to be very different from that of a social one, albeit related. A social commitment characterises the relation of an agent (bearer) towards another (counterparty) with respect to an action or state of affairs. Often a third agent is involved, the witness or authority in whose presence the commitment is taken and who may have responsibility for punishing the bearer agent in case of failing to fulfil its commitment [4]. Social commitments may arise between individual (or social) agents and other individual or social agents.

Social commitments involve the creation of relativised obligations and rights between the bearer and counterparty agents. Relativised obligations and social commitments are different in the following sense: if an agent commits to another agent to bring about a certain state, then this involves not only a relativised obligation on behalf of the bearer towards the counterparty, but also an intention (personal commitment) of the bearer. On the other hand, a relativised obligation may not necessarily mean that the bearer is personally committed to bring about the state of affairs. In contrast to general obligations that apply to all agents, relativised obligations involve specific individual and counterparty agents and can arise between a combination of individual and social agents. Moreover, obligations go hand-in-hand with other normative concepts such as permissions and rights. When a bearer agent violates its obligations it inevitably frustrates the expectations of the counterparty agent who now has the right to impose sanctions.

A multi-agent system can be viewed as a collection of agents who can join their forces under the umbrella of a social agent as long as they fulfil the commitments that they undertake as part of the group and through their roles. The system can be described in terms of its structure; that is by the way social agents, roles and authority relations are arranged to form a whole. Agents are free to join social agents while in pursuit of their own objectives, but at the same time they have to balance the commitments that they undertake while their behaviour is regulated via the relativised and general obligations that they hold.

In the sections that follow some aspects of the dynamics of social behaviour are formalised. In particular, obligations, rights, commitments and roles are formalised, while authority relations will be addressed as part of future work.

3. FORMAL FRAMEWORK

The logical framework is based on the BDI paradigm which we extend into a many-sorted first order modal logic. The basic ideas and the extensions made to the original framework are briefly described here; the reader is referred to [25, 26] for the full details of the BDI paradigm.

The logical language $\mathcal{L}$ includes, apart from the usual connectives and quantifiers, three modal operators B, D, and I for expressing beliefs, desires and intentions respectively. There are three sorts: $Agents$, $SAgents$, and $Other$. $Agents$ is the set of individual agents while $SAgents$ is the set of *social agents*, which may be groups of agents or individual agents. In fact, each individual agent is considered to be a social agent and is included as such in *SAgents*. *Other* indicates all the other objects/individuals in the universe of discourse. The framework includes a branching temporal component based on CTL* logic [13], in which the belief-, intention-, and desire-accessible worlds are themselves branching time structures. The operator *inevitable* is said to be true of a path formula γ at a particular point in a time-tree if γ is true of all paths emanating from that point. O-formulas are wffs that contain no positive occurrences of *inevitable* outside the scope of the modal operators B, D and I. The temporal operators *optional*, $\bigcirc$ (next), $\Diamond$ (eventually), $\Box$ (always), U (until) are also included. Furthermore the operators: *succeeds*(e), $fails(e)$, *does*(e), *succeeded*(e), $failed(e)$ and *done*(e), express the present and past success or failure of an event e. The additional operator $\in$ expresses membership in a social agent.

Semantics is given in terms of possible worlds relativised to time points. A model for $\mathcal{L}$ is a tuple $M =< W, E, T, \prec, S, U, \mathcal{B}, \mathcal{D}, \mathcal{I}, \pi >$ where W is a set of worlds, E is a set of primitive event types, T is a set of time points, $\prec$ is a binary relation on time points, S is the set of all situations $S \subseteq W \times T$, i.e. a situation is a world at a particular time point, U is the universe of discourse which is a tuple itself $U =< U_{Agents}, U_{SAgents}, U_{Other} >$, $\mathcal{B}$ is the belief accessibility relation, $\mathcal{B} : U_{Agents} \rightarrow \wp(W \times T \times W)$, and $\mathcal{D}$ and $\mathcal{I}$ similarly for desires and intentions and finally π interprets the atomic formulas of the language. Satisfaction of formulas is given in terms of a model M a mapping v of variables into elements of U, a world w and a time point t (i.e. a situation w_t):

$M_{v,w_t} \models P(\tau_1, ...\tau_k)$ iff $< v(\tau_1), ..., v(\tau_k) >\in \pi(P^k, w_t)$

$M_{v,w_t} \models \neg\phi$ iff $M_{v,w_t} \not\models \phi$

$M_{v,w_t} \models \phi \wedge \psi$ iff $M_{v,w_t} \models \phi$ and $M_{v,w_t} \models \psi$

$M_{v,w_t} \models \forall x\phi$ iff $\forall d \in U$, $M_{v[d/x],w_t} \models \phi$

$M_{v,w_t} \models B(i, \phi)$ iff $\forall$ w'_t such that $\mathcal{B}_i(w_t, w'_t)$, $M_{v,w'_t} \models \phi$

$M_{v,w_t} \models (\tau_1 = \tau_2)$ iff $\| \tau_1 \|=\| \tau_2 \|$

$M_{v,w_t} \models (i \in g)$ iff $\| i \|\in\| g \|$

$M_{v,w_t} \models$ *optional*(ϕ) iff $\exists$ a fullpath $w_{t_0}, w_{t_1}, ..$ such that $M_{v,w_{t_0},w_{t_1},..} \models \phi$

$M_{v,w_t} \models$ *succeeded*(e) iff $\exists$ t_0 such that $S_w(t_0, t_1) = e$

Similarly for the other connectives and operators.

4. INDIVIDUAL MENTAL ATTITUDES

An agent's information about the state of the world and the other agents is represented in terms of beliefs. This reflects the fact that what an agent believes may not necessarily hold and thus its picture of the world may not be correct. For the belief operator the standard KD45$_n$ system is adopted by requiring the accessibility relation $\mathcal{B}$ to be serial, transitive and euclidean:

B-K. $B(i, \phi) \wedge B(i, \phi \Rightarrow \psi) \Rightarrow B(i, \psi)$

B-D. $B(i, \phi) \Rightarrow \neg B(i, \neg\phi)$

B-S4. $B(i, \phi) \Rightarrow B(i, B(i, \phi))$

B-S5. $\neg B(i, \phi) \Rightarrow B(i, \neg B(i, \phi))$

B-N. if $\vdash \phi$ then $\vdash B(i, \phi)$

Desires express an agent's motivation. They are the states of affairs that the agent would ideally like to bring about. As such they may not be consistent with each other. However, are among the driving forces in the decision making process of an agent. Desires can be either present or future-directed. The axiom system adopted for desires is the K$_n$:

D-K. $D(i, \phi) \wedge D(i, \phi \Rightarrow \psi) \Rightarrow D(i, \psi)$

D-N. if $\vdash \phi$ then $\vdash D(i, \phi)$

There are no restrictions imposed on the accessibility relation for desire $\mathcal{D}$. In particular seriality is not imposed since desires may not be consistent with each other.

Intentions express an agent's commitment to itself to bring about a particular state of affairs. An agent may decide to adopt one of its desires as intention, but not all desires may be upgraded to the status of intentions. Intentions need to be consistent with each other and they may be present or future-directed. The axiom system adopted is D$_n$:

I-K. $I(i, \phi) \wedge I(i, \phi \Rightarrow \psi) \Rightarrow I(i, \psi)$

I-D. $I(i, \phi) \Rightarrow \neg I(i, \neg\phi)$

I-N. if $\vdash \phi$ then $\vdash I(i, \phi)$

The I-D axiom expresses the consistency of intentions. The K axiom and the necessitation rule which are included in the axiomatisation of all three attitudes are inherent of the possible worlds approach and they hold in normal modal logics regardless of any restrictions imposed on the accessibility relation. Thus agents are logically omniscient with respect to their attitudes. For a detailed discussion on the logical omniscience problem see [14] .

The interrelations between the three attitudes are described by a variation of the notion of *strong realism* which comes closer to the desiderata for rational reasoning agents [3, 26] than the original notion (for alternative notions of realism see [15]):

$I(i, \gamma) \Rightarrow B(i, \gamma)$

$D(i, \gamma) \Rightarrow \neg B(i, \neg\gamma)$

These correspond to the following semantic conditions:

C1. $\forall i \in U_{Agents}$, $\forall w_t, w'_t$ if $\mathcal{B}_i(w_t, w'_t)$ then $\exists w''_t$ s.t. $\mathcal{I}_i(w_t, w''_t)$ and $w''_t \sqsubseteq w'_t$

C2. $\forall i \in U_{Agents}$, $\forall w_t, \exists w'_t \mathcal{B}_i(w_t, w'_t)$ s.t. $\exists w''_t \mathcal{D}_i(w_t, w''_t)$ and $w''_t \sqsubseteq w'_t$

C1 requires that for all belief-accessible worlds w'_t from w_t, there is an intention-accessible world w''_t from w_t which is also a sub-world of w'_t, while C2 that there is at least one belief-accessible worlds w'_t from w_t, such that there is a desire-accessible world w''_t from w_t which is also a sub-world of w'_t. A world w'_t is a sub-world of w_t ($w'_t \sqsubseteq w_t$) if the tree structure of w'_t is a subtree of w_t, and w'_t has the same truth assignment and accessibility relations as w_t. By imposing the sub-world restriction between worlds the application of these axioms is restricted to O-formulas γ [25, 26].

Apart from the realism constraints directly relating the three attitudes the following properties are also adopted:

Belief of Intentions

If an agent intends ϕ then it believes that it intends it
$I(i, \phi) \Rightarrow B(i, I(i, \phi))$
BI. $\forall w_t, w'_t, w''_t$ $\mathcal{B}_i(w_t, w'_t) \wedge \mathcal{I}_i(w'_t, w''_t) \Rightarrow \mathcal{I}_i(w_t, w''_t)$

Belief of Desires

If an agent desires ϕ then it believes that it desires it
$D(i, \phi) \Rightarrow B(i, D(i, \phi))$
BD. $\forall w_t, w'_t, w''_t$ $\mathcal{B}_i(w_t, w'_t) \wedge \mathcal{D}_i(w'_t, w''_t) \Rightarrow \mathcal{D}_i(w_t, w''_t)$

The BDI system with the aforementioned connection axioms will be called SV-BDI. Strategies for the maintenance of intentions as in [25] can be adopted here as well.

5. BASIC GROUP ATTITUDES

Social agents are usually aggregations of agents and they may consist of individual agents, or individual agents and other social agents. The fact that an individual agent i is a member of a social agent si is expressed simply as ($i \in si$). In order to be able to reason about the members of a social agent the following set-theoretic relations are introduced:

$(sj \subseteq si) \equiv_{def} \forall i(i \in sj) \Rightarrow (i \in si)$

$(sj \subset si) \equiv_{def} (sj \subseteq si) \wedge \neg(sj = si)$

$singleton(si, i) \equiv_{def} \forall j(j \in si) \Rightarrow (j = i)$

$singleton(si) \equiv_{def} \exists i\ singleton(si, i)$

In order to be able to reason about a social agent's information state two additional modal operators $EB(si, \phi)$ and $MB(si, \phi)$ are introduced for "Every member of social agent si believes ϕ" and "ϕ is a mutual belief among the members of social agent si" respectively. Following [14]:

$EB(si, \phi) \equiv_{def} \forall i(i \in si) \Rightarrow B(i, \phi)$

Intuitively every member of a social agent believes ϕ if and only if every individual agent i in this social agent believes ϕ. Then a proposition ϕ is mutually believed by a social agent if every member believes it, and every member believes that every member believes it,... and so on. However, infinite formulas cannot be expressed in the language. Let $EB^1(si, \phi)$ be an abbreviation for $EB(si, \phi)$ and $EB^k(si, \phi)$ for $k \geq 1$ be an abbreviation for $EB(si, EB^{k-1}(si, \phi))$. Thus, if EB^k expresses the k-th level of nesting of belief of the agents in social agent si, then the social agent has mutual belief of ϕ:

$M_{v,w_t} \models MB(si, \phi)$ iff $M_{v,w_t} \models EB^k(si, \phi), k = 1, 2, ..$

Now define w'_t to be belief-si-reachable from w_t if there is a path in the Kripke model from w_t to w'_t along accessibility arrows $\mathcal{B}_i$ that are associated with members $i \in si$ [14]. Then the following property holds:

$M_{v,w_t} \models MB(si, \phi)$ iff $M_{v,w'_t} \models \phi$ for all w'_t that are belief-si-reachable from w_t

Using this property and the notion of reachability the following axiom and rule can be soundly added to the KD45$_n$ system of belief:

$MB(si, \phi) \Leftrightarrow EB(si, \phi \wedge MB(si, \phi))$

From $\phi \Rightarrow EB(si, \psi \wedge \phi)$ infer $\phi \Rightarrow MB(si, \psi)$

Next two modal operators for expressing what every member of a social agent intends, and what is mutually intended by a social agent $EI(si, \phi)$ and $MI(si, \phi)$ respectively are introduced. Similarly to EB, every member of si intends ϕ, if and only if every individual agent intends ϕ:

$EI(si, \phi) \equiv_{def} \forall i(i \in si) \Rightarrow I(i, \phi)$

Based on the definition of what everyone intends, ϕ is mutually intended by a social agent if every member intends it, and every member intends that every member intends it,... and so on. If EI^k expresses the k-th level of nesting of intentions of the agents in the social agent si, then:

$M_{v,w_t} \models MI(si, \phi)$ iff $M_{v,w_t} \models EI^k(si, \phi), k = 1, 2, ..$

A world w'_t is defined to be intention-si-reachable from w_t similarly to being belief-si-accessible and the following property holds:

$M_{v,w_t} \models MI(si, \phi)$ iff $M_{v,w'_t} \models \phi$ for all w'_t that are intention-si-reachable from w_t

Finally, the following axiom and rule can be soundly added to the D$_n$ system of intentions:

$MI(si, \phi) \Leftrightarrow EI(si, \phi \wedge MI(si, \phi))$

From $\phi \Rightarrow EI(si, \psi \wedge \phi)$ infer $\phi \Rightarrow MI(si, \psi)$

If the social agent is a singleton (individual agent) then the MB and MI operators reduce to their individual constituents respectively. In other words, the mutual belief of a single agent is simply its belief, and its mutual intention is simply an intention.

6. OBLIGATIONS AND RIGHTS

General obligations are expressed via an obligation operator O that prefixes propositions $\phi, \psi, ...$ as in standard deontic logic (SDL). A formula $O(\phi)$ is read "It ought to be the case that ϕ". Relativised obligations are expressed via an operator $O(si, sj, \phi)$ read as "Social agent si is obligated to sj to bring about ϕ". The model for the language is extended as follows: $M =< W, E, T, \prec, U, \mathcal{B}, \mathcal{D}, \mathcal{I}, \pi, \mathcal{O}, \mathcal{O}^* >$ where $\mathcal{O}$ is the accessibility relation for general obligations and $\mathcal{O}^* = \{\mathcal{O}_{si,sj} | \forall si, sj \in U_{SAgents} \wedge si \neq sj\}$ is the accessibility relation for relativised obligations between pairs of social agents. $\mathcal{O}$ is considered to yield the deontically ideal worlds relative to a world w at time point t:

$M_{v,w_t} \models O(\phi)$ iff $\forall$ w'_t s.t. $\mathcal{O}(w_t, w'_t)$, $M_{v,w'_t} \models \phi$

$M_{v,w_t} \models O(si, sj, \phi)$ iff $\forall w'_t$ s.t.$\mathcal{O}_{si,sj}(w_t, w'_t)$,$M_{v,w'_t} \models \phi$

For the general obligations operator we adopt the D system. This ensures that deontic conflicts are not allowed, that is not both ϕ and $\neg\phi$ ought to be the case:

$O(\phi) \Rightarrow \neg O(\neg\phi)$

The principle of veracity $O(\phi) \Rightarrow \phi$ is rejected since what ought to be the case may not be the case after all. For the relativised obligations operator the K$_n$ system is adopted. As a consequence deontic conflicts are allowed for relativised obligations. Finally, a permission operator is defined as the dual of the general obligation operator as follows:

$P(\phi) \equiv_{def} \neg O(\neg\phi)$

It seems reasonable to suggest that if ϕ is a general obligation then each agent believes that this is the case (special constant s_0 denotes the set of all agents, i.e. the social agent that constitutes the society of the domain):

$\forall i(i \in s_0) \Rightarrow (O(\phi) \Rightarrow B(i, O(\phi)))$ (*)

In other words, if ϕ ought to be the case, then each agent i believes that it ought to be the case. This axiom requires the following semantic condition:

$\forall i \in U_{Agents}$, $\forall w_t, w'_t, w''_t$ if $\mathcal{B}_i(w_t, w'_t)$ and $\mathcal{O}(w'_t, w''_t)$ then $\mathcal{O}(w_t, w''_t)$

Since general obligations ought to be believed by all agents we also derive the following from (*) by the axiom defining EB and the induction rule for MB:

$O(\phi) \Rightarrow MB(s_0, O(\phi))$

This means that normative statements are mutually believed (ideally) by all agents. It also seems reasonable to suggest that if such an ought-to relation between an agent (counterparty) and another agent (bearer) is in place, both of them should be believe that this is the case:

$O(si, sj, \phi) \Rightarrow \forall i(i \in si) \Rightarrow B(i, O(si, sj, \phi))$

$O(si, sj, \phi) \Rightarrow \forall j(j \in sj) \Rightarrow B(j, O(si, sj, \phi))$

Moreover we can accept the stronger axiom that such a relativised obligation is a mutual belief between the bearer and counterparty agents:

$O(si, sj, \phi) \Rightarrow MB(\{si, sj\}, O(si, sj, \phi))$

Another plausible principle is that if i is obligated to j to bring about ϕ, at least j should not desire that $\neg\phi$:

$O(i, j, \gamma) \Rightarrow \neg D(j, \neg\gamma)$

This in turn requires the following semantic restriction:

$\forall i, j \in U_{Agents}$, $\forall w_t, \exists w'_t \mathcal{D}_j(w_t, w'_t)$ s.t. $\exists w''_t$ $\mathcal{O}_{ij}(w_t, w''_t)$ and $w''_t \sqsubseteq w'_t$

The application of the axiom is restricted to O-formulas and accordingly it states that if agent i is obligated to j to bring about $optional(\psi)$, then j does not desire $\neg optional(\psi)$. Although there seem to be counter-arguments (parents may have relativised obligations regarding their children's education, which the children may not desire), we assume that the counterparty agent will take the necessary steps to free the bearer from the obligation (although this is not present in the current formalism), if it doesn't desire ϕ to be brought about. The following are also theorems:

$O(\phi) \Rightarrow \neg D(i, \neg O(\phi))$
$O(\phi) \Rightarrow \neg I(i, \neg O(\phi))$
Counterparty Agent
$O(i, j, \phi) \Rightarrow \neg I(j, \neg O(i, j, \phi))$
$O(i, j, \phi) \Rightarrow \neg D(j, \neg O(i, j, \phi))$
Bearer Agent
$O(i, j, \phi) \Rightarrow \neg I(i, \neg O(i, j, \phi))$
$O(i, j, \phi) \Rightarrow \neg D(i, \neg O(i, j, \phi))$

Once a social agent si has managed to bring about the desired state of affairs for agent sj, or it has come to its attention that the state of affairs is not an option any more, it needs to take some further action in order to ensure that the counterparty agent is aware of the situation. The social agent successfully de-commits itself from a relativised obligation in the following way:

$\mathit{succeeded}(\mathit{decommit}(si, sj, \mathit{inevitable}\Diamond\phi)) \Rightarrow$
$(\neg O(si, sj, \mathit{inevitable}\Diamond\phi)$
$\wedge\ \mathit{done}(\mathit{communicate}(si, sj, MB(si, \phi))))$
$\vee\ (\neg O(si, sj, \mathit{inevitable}\Diamond\phi)$
$\wedge\ \mathit{done}(\mathit{communicate}(si, sj, \neg MB(si, \mathit{optional}\Diamond\phi)))$
$\wedge\ \mathit{done}(\mathit{communicate}(si, sj, \neg O(si, sj, \mathit{inevitable}\Diamond\phi))))$

Accordingly a social agent si can successfully de-commit itself from a previously adopted relativised obligation towards another agent sj if: i) the social agent has come to believe that it has achieved its ϕ and in this case it drops its obligation towards sj and lets it know that the state of affairs has been achieved, or ii) the social agent has come to believe that the state of affairs that is committed to is not an option anymore, and in this case it successfully decommits itself by dropping the obligation and by letting the other agent know that the state of affairs is not an option and finally that it no longer holds the relativised obligation to bring about that state of affairs. *communicate* is used in a generic way to indicate that the agent needs to communicate with the other agent involved in order to de-commit successfully. This may be done by a member of the social agent, perhaps the individual agent that has the role of the representative or spokesperson in si. On the other hand the counterparty agent may reserve the right to impose sanctions on the bearer agent since it has the "right" to do so.

To this end another relativised operator *Right* is introduced in order to describe that a social agent sj has the right ψ over another social agent si expressed as $\mathit{Right}(sj, si, \psi)$. No particular restrictions are imposed on the accessibility relation for this modality. The formula ψ may express the form of the sanction that sj has the right to impose on si. Obligations and rights as we will see in the subsequent sections are created pairwise. If an agent si drops a previously adopted relativised obligation towards sj, and sj has a right over si, then sj may decide to exercise this right. Agents may have a lenient or a strict policy of exercising their rights. An agent has a lenient policy if it keeps its options open as to whether or not it will exercise its right over another agent:

$\mathit{Right}(sj, si, \psi) \wedge MB(sj, \neg O(si, sj, \mathit{inevitable}\Diamond\phi)) \wedge$
$\neg MB(sj, \phi) \Rightarrow \mathit{optional}(MI(sj, \mathit{optional}\Diamond\psi))$

That is the agent may optionally adopt the intention to optionally eventually bring about ψ. On the other hand a strict policy means that an agent will always exercise its rights on the deviating agent:

$\mathit{Right}(sj, si, \psi) \wedge MB(sj, \neg O(si, sj, \mathit{inevitable}\Diamond\phi)) \wedge$
$\neg MB(sj, \phi) \Rightarrow \mathit{inevitable}(MI(sj, \mathit{inevitable}\Diamond\psi))U\ MB(sj, \psi)$

The agent will keep trying to bring about ψ until it actually comes to believe that it has managed to do so. Agents may or may not reveal their policy on exercising rights to the other agents.

7. PREFERENCES

Agents express preferences when they are presented with a dilemma; when they are in a situation in which not every state of affairs that they would like to bring about is feasible. For instance, when somebody asks you what you would like to drink coffee or tea, this means that you can drink either coffee or tea, not both (of course being an autonomous agent you may decide to have both, but then there is no reason to express a preference). In this sense, preferences express an agent's choice between two states of the world that cannot both be realisable at the same time. This is how preferences should be understood in the context of this paper.

In order to be able to express that an agent prefers ϕ to ψ the language is extended by adding a modal operator $Pref$. Thus a formula of the form $Pref(i, \phi, \psi)$ means that agent i prefers ϕ to ψ. Semantics to this modality is given in terms of a world preference based on von Wright's [36] *conjunction expansion principle*. According to this, to say that an agent i prefers coffee to tea is to say that it prefers situations in which it has coffee and no tea to those in which it has tea but no coffee. In terms of possible worlds, that is to say that agent i prefers $\phi \wedge \neg\psi$-worlds to $\psi \wedge \neg\phi$-worlds. Unfortunately the semantics of preferences which is based on normal Kripke semantics gives rise to paradoxes of disjunction and conjunction: if ϕ is preferred to ψ, then $\phi \vee \chi$ is preferred to ψ, and ϕ is preferred to $\psi \wedge \chi$. χ being an irrelevant state of the world these properties can result in counterintuitive situations. In order to avoid this, the *ceteris paribus* nature of preferences needs to be captured: only $\phi \wedge \neg\psi$-worlds and $\neg\phi \wedge \psi$-worlds that otherwise differ as little as possible to the real world should be compared [2].

Following Bell and Huang [2, 18] and in line with the Stalnaker-Lewis' [30, 21] analysis of conditionals a selection function cw_t is defined as $cw_t(w_t, ||\phi||_v^M)$ where $||\phi||_v^M$ is an abbreviation for $M_{v,w_t} \models \phi$. cw_t gives the set of closest situations to w_t (worlds at time point t) in which ϕ is true. The function cw_t if of type $S \times P(S) \rightarrow P(S)$. Then $\succ$ is a comparison relation for preferences for each agent $\succ$: $U_{Agents} \rightarrow P(S) \times P(S)$. The truth condition for preferences is as follows:

$M_{v,w_t} \models Pref(i, \phi, \psi)$ iff
$cw_t(w_{t,}||\phi \wedge \neg\psi||_v^M) \succ_i cw_t(w_{t,}||\neg\phi \wedge \psi||_v^M)$

Accordingly we have the following properties [2]:

(IR) $\neg Pref(i, \phi, \phi)$
(TR) $Pref(i, \phi, \psi) \wedge Pref(i, \psi, \chi) \Rightarrow Pref(i, \phi, \chi)$
(DL) $Pref(i, \phi, \chi) \wedge Pref(i, \psi, \chi) \Rightarrow Pref(i, \phi \vee \psi, \chi)$
(DR) $Pref(i, \phi, \psi) \wedge Pref(i, \phi, \chi) \Rightarrow Pref(i, \phi, \psi \vee \chi)$
(CEP) $Pref(i, \phi, \psi) \Leftrightarrow Pref(i, (\phi \wedge \neg\psi), (\psi \wedge \neg\phi))$
(AS) $Pref(i, \phi, \psi) \Rightarrow \neg Pref(i, \psi, \phi)$
(CP) $Pref(i, \phi, \psi) \Rightarrow Pref(i, \neg\psi, \neg\phi)$

IR and TR state the irreflexivity and transitivity of preferences respectively. DL and DR describe the left and right disjunction of preferences while CEP states the conjunction expansion principle. Preferences are asymmetric (AS) and contraposable (CP). Moreover, since $Pref(i, \phi, \psi)$ implies neither $Pref(i, \phi \vee \psi, \psi)$ nor $Pref(i, \phi, \psi \wedge \chi)$, the paradoxes involving conjunction and disjunction do not arise [2].

As was mentioned earlier, agents express preferences when they are forced to chose between two (or more) states of affairs that cannot be realisable at the same time. It seems that when an agent prefers ϕ to ψ, it believes that ϕ is an option or can be realised, but it does not believe that ψ can be realised or is an option at the same time:

$Pref(i, \phi, \psi) \Rightarrow B(i, \mathit{optional}\Diamond\phi) \wedge \neg B(i, \mathit{optional}\Diamond\psi)$

Individual agents can express their preferences between states of affairs, but it also seems possible that groups of agents or a social agent can express a preference. To this end a modal operator $EPref(si, \phi, \psi)$ is introduced which is read as "everyone in social agent si prefers ϕ to ψ". Thus:

$EPref(si, \phi, \psi) \equiv_{def} \forall i(i \in si) \Rightarrow Pref(i, \phi, \psi)$

Then a mutual preference among the members of a social agent si is defined as follows:

$MPref(si, \phi, \psi) \equiv_{def}$
$EPref(si, \phi, \psi) \wedge MB(si, EPref(si, \phi, \psi))$

This is a weaker collective attitude compared to mutual belief and mutual intentions.

8. COMMITMENTS

Social commitments will be expressed via an operator $SCom(si, sj, \phi)$ which is read "social agent si is committed to social agent sj to bring about ϕ". Since social commitments can arise between both individual and social agents we require a definition that covers all four cases. Moreover, adopting a commitment is a rights and obligations producing act [4] between the bearer and counterparty agents and this needs to be reflected in the definition:

$SCom(si, sj, \phi) \Leftrightarrow$
$O(si, sj, \phi) \wedge MI(si, \phi) \wedge Right(sj, si, \psi)$
$\wedge\, MB(\{si, sj\}, (O(si, sj, \phi) \wedge MI(si, \phi) \wedge Right(sj, si, \psi)))$

In the simple case of two individual agents the MB and MI operators are reduced to an individual belief and intention respectively while if si and sj are groups then a mutual intention and a mutual belief arise. Intuitively there should be conditions under which an agent is allowed to drop its social commitments as discussed in [19, 12]. In what follows two different commitment strategies for social commitments are described, namely *blind* and *reliable*.

A social agent has a *blind* social commitment strategy if it maintains its commitment until actually it is a mutual belief among the members of the social agent that it has been achieved. The social agent will keep trying to bring about the state of affairs, until there is a mutual belief that this has been achieved:

$SCom(si, sj, \mathit{inevitable}\Diamond\phi) \Rightarrow$
$\mathit{inevitable}(SCom(si, sj, \mathit{inevitable}\Diamond\phi)\ U\ MB(si, \phi))$

Clearly such a strategy towards social commitments is very strong. If this requirement is relaxed then we can define a *reliable* strategy. A social agent follows a *reliable* strategy if it keeps its commitment towards another agent as long as the members of the social agent have mutual belief that it is still an option:

$SCom(si, sj, \mathit{inevitable}\Diamond\phi) \Rightarrow$
$\mathit{inevitable}(SCom(si, sj, \mathit{inevitable}\Diamond\phi)\ U$
$(MB(si, \phi) \vee \neg MB(si, \mathit{optional}\Diamond\phi)))$

We turn now our attention to collective commitments. A collective commitment is the internal commitment of a social agent to itself. Such a commitment seems to involve first of all social commitments on behalf of the individual members of the group towards the group, a mutual intention of the group to achieve ϕ, and finally a mutual belief that the social agent has the mutual intention ϕ. Thus, the definition of collective commitment is as follows:

$CCom(si, \phi) \Leftrightarrow$
$\forall i(i \in si) \Rightarrow SCom(i, si, \phi) \wedge MI(si, \phi) \wedge MB(si, MI(si, \phi))$

9. ROLES

Following Cavedon and Sonenberg [6] roles are first class entities in this framework. Three additional sorts *Relns*, *RelTypes* and *Roles* are introduced. *Relns* constants represent relationship instances. E.g. if *Ray* is in a supervisor-student relationship with two different students then these relationships will be represented by different *Relns* constant symbols. *RelTypes* constants represent a collection or type of relationship. E.g. all student-supervisor relationships will be of the same type. *RelTypes* objects allow us to abstract and associate properties with a collection of such relationships. *Roles* constants represent "role types", e.g. the same constant symbols are used to represent the supervisor and the student roles in each supervisor-student relationship.

Roles are related to relationship types via a predicate $RoleOf(a, R)$ which describes that a is one of the roles in relationship of type R. A three place predicate $In(si, a, r)$ which asserts that social agent si is in role a of relationship r is introduced. Moreover only one agent can fill a role in a given relationship at any given time:

$\forall i, j, a, r\ In(si, a, r) \wedge In(sj, a, r) \Rightarrow (si = sj)$

We require that roles of a relationship type are filled when any role of that type is filled (note: given a relationship r, $\widehat{r}$ denotes its corresponding type):

$\forall r, \forall si, a\ \ In(si, a, r) \Rightarrow$
$\forall b(RoleOf(b, \widehat{r}) \Rightarrow \exists sj\ In(sj, b, r))$

In order to express that a role a involves the adoption of a social commitment ϕ a new modality $RoleSCom(a, \phi)$ is introduced. No particular restrictions are imposed on the accessibility relation for this modality. *RoleSCom* is used in order to define the general social commitments associated with a particular role. This then provides a way of associating relativised-obligations to roles. Intuitively if role a involves the social commitment ϕ and social agent si has the role a in relationship r, then there exists another social agent sj (different to si) that has the role b in relationship r towards whom agent si has the social commitment ϕ:

$RoleSCom(a, \phi) \wedge In(si, a, r) \Rightarrow$
$\exists sj, b\ In(sj, b, r) \wedge SCom(si, sj, \phi) \wedge \neg(si = sj)$

An agent can decide to drop a role if it comes to believe that it has fulfilled its social commitments (e.g. a the supervisor of a Ph.D. student may drop its role once the student has been successfully examined), or if it has come to believe that it cannot fulfil the commitments of its role anymore. This may happen for a variety of reasons, for instance the agent may decide that it is not to its benefit to adhere to the role any longer, or another role is in conflict with the first one. However, the agent that decides to drop a role needs to communicate to the other agent that it is doing so as well as whether or not it believes that it has fulfilled its commitments or not.

$\mathit{succeeded}(\mathit{droprole}(si, sj, a)) \Rightarrow$
$(\neg In(si, a, r) \wedge \neg SCom(si, sj, \phi) \wedge$
$\mathit{done}(\mathit{communicate}(si, sj, (\neg In(si, a, r) \wedge$
$\neg SCom(si, sj, \mathit{inevitable}\Diamond\phi)))))$
$\vee(\neg In(si, a, r) \wedge \neg SCom(si, sj, \phi) \wedge$
$\mathit{done}(\mathit{communicate}(si, sj, (\neg In(si, a, r) \wedge MB(si, \phi)))))$
$\vee(\neg In(si, a, r) \wedge \neg SCom(si, sj, \phi) \wedge$
$\mathit{done}(\mathit{communicate}(si, sj, (\neg In(si, a, r) \wedge$

$\neg MB(si, optional\Diamond\phi)))))$

An agent may also decide to drop a commitment which is part of its role without dropping the role itself and perhaps accepting that a form of sanction will have to be imposed.

10. SIMPLE EXAMPLE

Consider the following scenario. John (J) is a Ph.D. student supervised by Sandy (S). John is also a member of the University football team the Aces (A) for which he plays on Sunday mornings. This situation regarding John's roles and relationships is described below:

$$In(J, student, r1), In(S, supervisor, r1)$$

$$RoleScom(student, followadvice) \wedge In(J, student, r1) \Rightarrow In(S, supervisor, r1) \wedge SCom(J, S, followadvice)$$

$$In(J, player, r2), In(A, team, r2)$$

$$RoleScom(player, playgame) \wedge In(J, player, r2) \Rightarrow In(A, team, r2) \wedge SCom(J, A, playgame)$$

On Friday morning Sandy asks John to finish writing a paper which needs to be sent to a very prestigious conference on Monday morning. If he doesn't, then this will have consequences on his progress. John realises that this needs to be done over the weekend. His commitments are:

$$SCom(J, S, writepaper) \Leftrightarrow O(J, S, writepaper) \wedge I(J, writepaper) \wedge Right(S, J, inhibitprogress(S, J)) \wedge MB(\{J, S\}, O(J, S, writepaper) \wedge I(J, writepaper) \wedge Right(S, J, inhibitprogress(S, J)))$$

$$SCom(J, A, playgame) \Leftrightarrow O(J, A, playgame) \wedge I(J, playgame) \wedge Right(A, J, exclude(A, J)) \wedge MB(\{J, A\}, O(J, A, playgame) \wedge I(J, playgame) \wedge Right(A, J, exclude(A, J)))$$

Given the fact that John plays for the Aces every Sunday it is clear to him that not both of his commitments can be honoured. Thinking of the consequences of dropping each of its commitments and the possible repercussions, John's preferences are as follows:

$$Pref(J, writepaper, playgame)$$

$$Pref(J, exclude(A, J), inhibitprogress(S, J))$$

Although he does not want to disappoint his team, he decides that it is impossible to play in the game while finishing the paper at the same time:

$$B(J, optional\Diamond writepaper) \wedge \neg B(J, optional\Diamond playgame)$$

Since John follows a reliable strategy regarding his commitments the belief that it is not an option any longer to play in the game leads him to drop its commitment. He decommits and also lets the team know about this:

$$succeeded(decommit(J, A, inevitable\Diamond playgame)) \Rightarrow (\neg O(si, sj, inevitable\Diamond playgame) \wedge done(communicate(si, sj, \neg MB(si, optional\Diamond playgame))) \wedge done(communicate(si, sj, \neg O(si, sj, inevitable\Diamond playgame))))$$

Now luckily for John the team has a lenient policy and this time he does not get excluded. Notice, that although John did not fulfil his commitment towards the team which was part of his role in the team, he did not drop this role.

Let the football team Aces (A) consist of the team of players and a coach for simplicity. The collective commitment that characterises the team is that they win the X cup. This is a collective commitment of the whole football team. According to the definition then we have:

$$CCom(A, wincup) \Leftrightarrow \forall i(i \in A)\ SCom(i, A, wincup) \wedge MI(A, wincup) \wedge MB(A, MI(A, wincup))$$

Accordingly, the football team has a collective intention to win the cup iff every member of the social agent has a social commitment towards the social agent to win the cup, and it is a mutual intention among the members to do so, and it is also a mutual belief among the football team that the team has the mutual intention to win the cup. The social commitments involved in this definition give rise to relativised obligations and personal intentions towards the state of affairs which is to win the cup. The structure of the social agent "football_team" is described by the relationship between a team of players and a coach with the corresponding roles. Each of the roles prescribes a set of social commitments which come in support of the collective commitment of the football team.

11. CONCLUDING REMARKS

This paper presented an analysis of some of the aspects of the dynamics of social behaviour among agents within the BDI paradigm. The cornerstones of the approach followed here are commitments, roles, obligations and rights. Although, the formalism presented so far is by no means a complete characterisation of these dynamics, it adds to the literature in an essential way. Firstly, normative attitudes such as obligations and rights are considered in relation to social attitudes such as commitments and roles. Secondly, the definition of a social commitment provided is a generic one, that is, it covers all four cases of social commitments arising between a combination of social and individual agents. Thirdly, roles are associated with social commitments which in turn give rise to rights and obligations and as a result this approach provides a way of explaining how these attitudes arise in a unified way. Moreover, the formalism provides the means for de-commiting from an obligation, a social commitment or even a role in a variety of situations. So for instance, an agent may drop a social commitment which is part of an adopted role, without dropping the role itself. However, an agent may have to bear the consequences of the other agents' rights on itself.

Although this work leaves many unanswered questions, we believe it is a first step towards a comprehensive formal model of activity and regulation of behaviour within a group of agents. Towards this direction there are a number of possible avenues for future research. A formalisation of authority relations within roles is pending. Authority relations seem to create power relations within a group and issues such as whether authority relations can be transitive need to be looked into. Moreover, we need to provide a formal definition of a social agent based on roles and their interaction. As a first step, we can define a social agent structure to be a generic description of the roles and the relationships between them in a particular type of social agent. Formally, a social agent structure is a tuple $SA =< R, RI >$ where R is a finite set of roles $R \subseteq Roles$, i.e. R is the set of all possible roles that can be played by agents within a social agent of this type and RI is the relationship interaction graph that specifies all the valid generic relationship types between roles, $RI : R \times R \rightarrow RelTypes$. Each edge of the graph represents a relationship type (a, b) between roles $a, b \in R$. For the special case that a social agent is a singleton, the set of roles is the empty set and the interaction graph has no edges. This concept of a social structure needs to be further extended so as to take into account authority relations as well as collective commiments.

12. REFERENCES

[1] Aqvist, L. Deontic Logic. In *Handbook of Philosophical Logic* Vol.II (D. Gabbay and F. Guenthner eds), Reidel Publishing Company. pp.605-714, 1983.

[2] Bell, J. and Huang, Z. Dynamic goal Hierarchies. In *Intelligent Agent Systems: Theoretical and Practical Issues* (L. Cavedon, A. Rao and W. Wobcke eds), pp.88-103, 1997.

[3] Bratman, M.E. *Intentions, Plans, and Practical Reason.* Harvard University Press, 1987.

[4] Castelfranchi, C. Commitments: From Individual Intentions to Groups and Organisations. In *Proceedings of the First ICMAS Conference.* pp. 41-48, 1995.

[5] Castelfranchi, C. and Falcone, R. From Task Delegation to Role Delegation. In *Proceedings of the AI*IA 97: Advances in Artificial Intelligence Congress.* pp.278-289, 1997.

[6] Cavedon, L. and Sonenberg, L. On Social Commitments, Roles and Preferred Goals. In *Proceedings of the Third ICMAS Conference.* pp. 80-87, 1998.

[7] Cohen, P.R. and Levesque, H.J. Intention is Choice with Commitment. *Artificial Intelligence*, 42:213-261, 1990.

[8] Cohen, P.R. and Levesque, H.J. Teamwork. *Nous*, 25:485-512, 1991.

[9] Dignum, V., Meyers, J.-J. and Weigand H. Towards an Organizational Model for Agent Societies Using Contracts. In *Proceedings of the AAMAS'02 Conference.* pp. 694-,695, 2002.

[10] Dignum,F., Kinny,D. and Sonenberg, L. Motivational Attitudes of Agents: On Desires, Obligations and Norms. In *Proceedings of the Second International Workshop of Central and Eastern Europe on Multi-Agent Systems.* pp. 61-70, 2001.

[11] Dignum, F.; Meyer, J.-J Ch.; Wieringa, R.J. and Kuiper, R. A. A Modal Approach to Intentions, Commitments and Obligations: Intention plus Commitment yields Obligation. In *Deontic Logic, Agency and Normative Systems*, (M.A Brown and J. Carmo eds). pp.80-97, 1996.

[12] Dunin-Keplicz, B. and Verbrugge, R. Collective Motivational Attitudes in Cooperative Problem Solving. In *Proceedings of the First International Workshop of Central and Eastern Europe on Multi-Agent Systems.* pp.22-41, 1999.

[13] Emerson, E.A. and Srinivasan J. Branching Time Temporal Logic. In *Linear Time, Branching Time and Partial Order in Logics and Models for Concurrency* (J.W.de Bakker, W.P. de Roever and G.Rozenberg eds). pp.123-172, 1989.

[14] Fagin, R., Halpern, J.Y., Moses, Y., and Vardi, M.Y. *Reasoning about Knowledge.* MIT Press. Cambridge, MA, 1995.

[15] Fasli, M. Heterogeneous BDI Agents. Cognitive Systems Ressearch Journal. (forthcoming)

[16] Herrestad, H. and Krogh, C. Deontic Logic Relativised to Bearers and Counterparties. In *Anniversary Anthology in Computers and Law.* pp. 453-522, 1995.

[17] Hilpinen, R. (ed.) *Deontic Logic: Introductory and Systematic Readings.* Reidel Publishing Company, 1971.

[18] Huang, Z., Masuch, M. and Polos, L. ALX: An action logic for agents with bounded rationality. *Artificial Intelligence*, 82:101-153, 1996.

[19] Jennings, N.R. Commitments and Conventions: The Foundation of Coordination in Multi-Agent Systems. *Knowledge Engineering Review*, 8(3):223-250, 1993.

[20] Krogh, C. Obligations in Multi-Agent Systems. In *Proceedings of the 5th Scandinavian Conference on AI, 1995.*

[21] Lewis, D. *Counterfactuals.* Basil Blackwell, Oxford, 1973.

[22] Ma, G. and Shi, C. Modelling Social Agents in BDO Logic. In *Proceedings of the 4th ICMAS Conference.* pp. 411-412, 2000.

[23] Norman, T.J. and Reed, C.A. Delegation and Responsibility. *Intelligent Agents VII.* Springer Verlag. pp.136-149, 2000.

[24] Panzarasa, P., Jennings, N.J. and Norman T.J. Formalizing Collaborative Decision-making and Practical Reasoning in Multi-agent Systems. *Journal of Logic and Computation*, 11(6), pp. 1-63, 2001.

[25] Rao, A. and Georgeff, M. Modeling Rational Agents within a BDI-Architecture. In *Proceedings of the Second Int. Conf. on Principles of Knowledge Representation and Reasoning.* pp.473-484, 1991

[26] Rao, A. and Georgeff, M. Decision Procedures for BDI Logics. *Journal of Logic and Computation*, 8(3), pp. 293-343, 1998.

[27] Rao, A., Georgeff, M. and Sonenberg, E.A. Social Plans: A Preliminary Report. In *Decentralised A.I.-3.* pp. 57-76, 1992.

[28] Royakkers, L. and Dignum, F. Organisations and Collective Obligations. In *Proceedings of the Database and Expert Systems Applications Conference.* pp.302-311, 2000.

[29] Singh, M.P. An Ontology for Commitments in Multiagent Systems: Towards a Unification of Normative Concepts. *AI and Law*, 7:97-113, 1999.

[30] Stalnaker, R. A theory of conditionals. *Studies in Logical Theory, American Philosophical Quarterly*, 2:98-122, 1968.

[31] Tambe, M. Implementing agent teams in dynamic multi-agent environments. *Applied AI*, 12:189-210, 1998.

[32] van der Torre, L. and Tan, Y.-H. Rights, Duties and Commitments between Agents. In *Proceedings of the International Joint Conference on Artificial Intelligence* (IJCAI'99), pp. 1239-1244, 1999.

[33] Tuomela, R. and Miller,K. We-intentions. *Philosophical Studies,* 53:367-389, 1988.

[34] Wooldridge, M. and Jennings, N.R. The Cooperative Problem-solving Process. *Journal of Logic and Computation,* 9:563-592, 1999.

[35] Wooldridge, M. *Reasoning about Rational Agents.* The MIT Press, 2000.

[36] von Wright, G. H. *The Logic of Preference.* Edinburgh University Press, Edinburgh, 1963.