

Collective Commitments

Barbara Dunin-Kępicz

Institute of Informatics

Warsaw University

Banacha 2

02-097 Warsaw

Poland

E-mail: kepicz@mimuw.edu.pl

Rineke Verbrugge

Dept. of Math. and Computer Science

Vrije Universiteit

De Boelelaan 1081

1081 HV Amsterdam

The Netherlands

E-mail: rineke@cs.vu.nl

Abstract

In this paper we aim to arrive at an understanding of the notion of collective commitments in groups of agents. Collective intentions and collective commitments are formalized in the logical framework designed by Rao and Georgeff in (Rao & Georgeff 1991). Then, three commitment strategies are defined and investigated in the context of planning and plan execution in different kinds of groups of agents within a dynamically changing environment. It turns out that already at the stage of group formation, it is crucial to choose agents with appropriate commitment strategies as members. Next, during task delegation all members' commitment strategies should be taken into account in the proper way. Finally, during plan execution agents of different commitment characteristics play different roles.

Introduction

In this paper we discuss and formalize social and collective aspects of intentions and commitments. Our goal is to arrive at an understanding of the notion of commitments in different groups of agents.

In order to remain as generic as possible, we impose only a few restrictions on individual agents. With respect to social capabilities they should be able to communicate, cooperate and coordinate. Our preliminary choice is to focus on strictly cooperative groups. All individual motivational attitudes are viewed as primitive notions, but in contrast to (Rao & Georgeff 1991) we investigate logical aspects of relations between individual intentions and social commitments on the one hand, and collective intentions and commitments on the other hand.

Starting from individual intentions, we first define a notion of a *collective intention* for a group. Together with individual and mutual belief, a collective intention constitutes a basis for preparing a plan (or a set of plans). Here we abstract from methods of constructing plans, which form a separate subject. We assume that during planning each agent is delegated to realize particular goals (actions) and that all necessary aspects of the cooperation are well defined.

What we achieve next is to characterize the strongest motivational attitude, namely the *collective commitment* of a group. We assume that bilateral aspects of a plan — mutual obligations between agents — are reflected in *social commitments*. Thus collective commitments are defined on the basis of, previously characterized, collective intentions and social commitments together with mutual beliefs about them.

A collective commitment is meant to reflect the way a plan is to be executed, but one should take into account that agents are autonomous entities taking individual decisions in the light of the current situation and the commitment strategies they follow. We provide a classification of agents' commitment strategies based on an analogical characterization of intentional strategies given by Rao and Georgeff in (Rao & Georgeff 1991). Let us stress, however, that because the change of an agent's social commitment directly influences the way the plan is executed, the conditions imposed on agents dropping social commitments are more restrictive than in the case of changing individual intentions.

The next question addressed in this paper is how to maintain cooperation within different types of groups in the case that some members change their motivational attitudes. It will be shown that the cooperative possibilities depend on the strength of the commitment strategies of the members of the group. These strategies play a crucial role both during the stage of planning and during plan execution.

The paper is structured in the following manner. In the section "Beliefs", notions of a group and common belief are introduced. In the section "Individual and social motivational attitudes", some motivational attitudes like individual goals and intentions and social commitments are discussed as well as different types of agents' commitment strategies. On this basis collective motivational attitudes are defined in the following section, "Collective motivational attitudes". Next, the section "Planning and acting in different groups" provides a preliminary analysis of planning and acting in different types of groups. Finally, the last section focuses on discussion and options for further research.

Beliefs

To represent beliefs, we adopt a standard $KD45_n$ -system for n agents as explained in (Fagin *et al.* 1995). We take $BEL(a, \varphi)$ to have as intended meaning “agent a believes proposition φ ”. In the semantics, there are the usual accessibility relations B_i that lead from worlds w to worlds that are consistent with agent i 's beliefs in w . These accessibility relations are transitive, serial and euclidean.

One can define modal operators for group beliefs. The formula $E-BEL_G(\varphi)$ is meant to stand for “every agent in group G believes φ ”. It is defined semantically as $w \models E-BEL_G(\varphi)$ iff for all $i \in G$, $w \models BEL(i, \varphi)$.

A stronger operator is the one for common belief, where $C-BEL_G(\varphi)$ is meant to be true if everyone in G believes φ , everyone in G believes that everyone in G believes φ , etc. Let $E-BEL_G^0(\varphi)$ be an abbreviation for φ , and let $E-BEL_G^{k+1}(\varphi)$ be an abbreviation for $E-BEL_G(E-BEL_G^k(\varphi))$. Thus we have $w \models C-BEL_G(\varphi)$ iff $w \models E-BEL_G^k(\varphi)$ for all $k \geq 1$.

Define t to be G -reachable from s if there is a path in the Kripke model from s to t along accessibility arrows that are associated with members of G . Then the following property holds:

$$s \models C-BEL_G(\varphi) \text{ iff } t \models \varphi \text{ for all } t \text{ that are } G\text{-reachable from } s.$$

Using this property, it can be shown that the following axioms and rule can be soundly added to the $KD45_n$ -system. The resulting system is called $KD45_n^C$, and it is sound and complete with respect to Kripke models where all n accessibility relations are transitive, serial and euclidean (Fagin *et al.* 1995).

$$E-BEL_G(\varphi) \leftrightarrow \bigwedge_{i \in G} BEL(i, \varphi)$$

$$C-BEL_G(\varphi) \rightarrow E-BEL_G(\varphi \wedge C-BEL_G(\varphi))$$

From $\varphi \rightarrow E-BEL_G(\psi \wedge \varphi)$ infer $\varphi \rightarrow C-BEL_G(\psi)$
(Induction Rule).

Some of the ways in which individual beliefs can be established are updating, revision, and contraction. The establishment of common beliefs among a group is more problematic. In (Halpern & Moses 1990) it is shown that bilateral sending of messages does not suffice to determine common belief. We assume that in our groups, a more general type of communication, e.g. by a kind of global announcement, can be achieved. A good reference to the problems concerning common belief and to their possible solutions is (Fagin *et al.* 1995, Chapter 11)

COMM(a, b, φ)	agent a commits to agent b to make φ true
COMM(a, b, α)	agent a commits to agent b to do α
GOAL(a, φ)	agent a has as a goal that φ be true
GOAL(a, α)	agent a has as a goal to do α
OUGHT(a, φ)	agent a ought to make φ true
OUGHT(a, α)	agent a ought to do α
<i>still</i> (a, φ)	agent a sees to it that φ holds
<i>done</i> (a, α)	agent a has just done α
INT(a, φ)	agent a has the intention to make φ true
INT(a, α)	agent a has the intention to do α
E-INT(G, φ)	every agent in group G has the individual intention to make φ true
E-INT(G, α)	every agent in group G has the individual intention to do α
C-INT(G, φ)	group G has the collective intention to make φ true
C-INT(G, α)	group G has the collective intention to do α
C-COMM(G, P, φ)	group G has a collective commitment to make φ true by plan P
C-COMM(G, P, α)	group G has a collective commitment to do α by plan P

Table 1: Formulas and their intended meaning

Individual and social motivational attitudes

After an introduction of the notation used in this paper and a short discussion of individual goals and intentions, this section focuses on a definition of social commitments and three possible commitment strategies that agents may follow.

Notation

In our framework, most axioms relating motivational attitudes of agents appear in two forms: one with respect to *propositions*, the other with respect to *actions*. These actions are interpreted in a generic way we abstract from any particular form of actions: they may be complex or primitive, viewed traditionally with certain effects or with default effects (Dunin-Kępicz & Radzikowska 1995a; 1995b; 1995c), etc.

A proposition, on the other hand, reflects the particular state of affairs that an agent aims for. In other words, propositions represent the agent's higher level goals. Again, we abstract from particular methods of achieving them; e.g. they may be realized by particular plans.

Table 1 gives the formulas appearing in this paper, together with their intended meanings. The symbol φ denotes a proposition and α an action.

Individual goals and intentions

In Rao and Georgeff's (Rao & Georgeff 1991), individual beliefs, goals, and intentions are formalized as primitive notions and given a formal semantics. We take their semantics as a basis for our formalization of collective motivational attitudes, and refer the reader to (Rao & Georgeff 1991) for details. As a reminder, the temporal structure is a discrete tree branching towards the future. The temporal operators include *inevitable*(φ) (in all paths through the point of reference φ holds), *optional*(φ) $\equiv \neg$ *inevitable*($\neg\varphi$), $\Diamond\varphi$ (somewhere later on the same path, φ holds) and $\varphi \mathbf{U} \psi$ (either φ will hold forever on this path, or, as soon as it stops holding, ψ will hold, i.e. φ until ψ).

O-formulas are defined to be those formulas that contain no positive occurrences of *inevitable* outside the scope of modal operators like BEL, GOAL, and INT. For φ an O-formula, we adopt Rao and Georgeff's axioms of *strong realism*:

$$\text{INT}(a, \varphi) \rightarrow \text{GOAL}(a, \varphi)$$

and

$$\text{GOAL}(a, \varphi) \rightarrow \text{BEL}(a, \varphi).$$

Social commitments

A social commitment between two agents is not as strong as a collective commitment among them (see the subsection "Collective commitments"), but stronger than an individual intention of one agent. If I am committed to you to do something, then I should have the *intention* to do that. Moreover, I commit to you only if you are *interested* in my fulfilling my intention. In order to formalize these two conditions for social commitments, we adopt the following definitions (inspired by (Castelfranchi 1995)) concerning commitments that one agent makes to the other:

$$\text{COMM}(a, b, \varphi) \rightarrow \text{INT}(a, \varphi) \wedge \text{GOAL}(b, \text{stl}(a, \varphi))$$

$$\text{COMM}(a, b, \alpha) \rightarrow \text{INT}(a, \alpha) \wedge \text{GOAL}(b, \text{done}(a, \alpha))$$

Thirdly, social commitment has a deontic aspect: Castelfranchi I *ought* to do it. We, in contrast, find that the strength of the obligation depends on the situation: are the agents involved *responsible* ones? Below we will give an axiom that characterizes responsible agents by relating social commitments and obligations. Moreover, the axiom reflects our view that the obligation is related to the current state of the agent's commitment: only as long as an agent's commitment is still valid, she ought to fulfill it. The situations in which agents can drop their previously established commitments will be discussed in the subsection "Cooperation during plan

execution". Thus, the deontic aspect of social commitments is expressed by the following axiom:

$$\begin{aligned} \text{COMM}(a, b, \text{inevitable } \Diamond\varphi) \rightarrow \\ \text{inevitable}(\text{OUGHT}(a, \varphi) \mathbf{U} \\ \neg\text{COMM}(a, b, \text{inevitable } \Diamond\varphi)). \end{aligned}$$

Here, OUGHT(a, φ) is taken to be a modal operator parametrized by the agent a , with as intended reading " a is obliged to make φ true".

There are many axiom systems and corresponding semantics in the literature on deontic logic (cf. (Åqvist 1984)). There are also systems in which agents have obligations not only towards propositions, but also with respect to actions (cf. (d'Alton, Meyer, & Wieringa 1993)). It is sufficient for the moment, though, to assume a standard propositional KD-type modal logic for the obligations of each agent. In the corresponding Kripke semantics, there are the usual accessibility relations R_a that lead from worlds w to worlds that are "optimal" for agent a in w . These accessibility relations are serial.

Commitment strategies

In (Rao & Georgeff 1991), agents are divided into three different kinds according to their *intention strategies*, which describe in what circumstances (if any) the agents drop their intentions. We define and investigate three kinds of agents according to the strength with which they maintain the stronger motivational attitude, namely their social commitments. The definitions are inspired by those in (Rao & Georgeff 1991), but in addition we include the social aspects of communication and coordination. We assume that the commitment strategies depend only on the agent and not on the goal to which she is committed, nor on the other agent to whom she is committed.

We also assume that each agent knows which commitment strategies are adopted by other agents in the group. This meta-knowledge ensures proper replanning and coordination (see the subsection "Cooperation during plan execution"). The key point is whether and in which circumstances the agent can drop a social commitment. If such a situation arises, the next question is how to deal with it responsibly.

The strongest commitment strategy is followed by the *blindly committed* agent, who maintains her commitments until she actually believes that they have been achieved. Formally,

$$\begin{aligned} \text{COMM}(a, b, \text{inevitable } \Diamond\varphi) \rightarrow \\ \text{inevitable}(\text{COMM}(a, b, \text{inevitable } \Diamond\varphi) \mathbf{U} \\ \text{BEL}(a, \varphi)). \end{aligned}$$

Single-minded agents may drop social commitments when they do not believe anymore that the commitment is realizable. However, as soon as the agent abandons a commitment, some communication and coordination with the other agent is needed. We state the following:

$$\begin{aligned} & \text{COMM}(a, b, \text{inevitable } \Diamond\varphi) \rightarrow \\ & \text{inevitable}(\text{COMM}(a, b, \text{inevitable } \Diamond\varphi) \text{ U} \\ & (\text{BEL}(a, \varphi) \vee \\ & (\neg\text{BEL}(a, \text{optional } \Diamond\varphi) \wedge \\ & \text{done}(\text{communicate}(a, b, \neg\text{BEL}(a, \text{optional } \Diamond\varphi))) \\ & \wedge \text{done}(\text{coordinate}(a, b, \varphi))))). \end{aligned}$$

For open-minded agents, the situation is similar as for single-minded ones, except that they can drop social commitments if they do not aim for the respective goal anymore. As in the case of single-minded agents, communication and coordination will be involved, as expressed by the axiom:

$$\begin{aligned} & \text{COMM}(a, b, \text{inevitable } \Diamond\varphi) \rightarrow \\ & \text{inevitable}(\text{COMM}(a, b, \text{inevitable } \Diamond\varphi) \text{ U} \\ & (\text{BEL}(a, \varphi) \vee \\ & (\neg\text{GOAL}(a, \text{optional } \Diamond\varphi) \wedge \\ & \text{done}(\text{communicate}(a, b, \neg\text{GOAL}(a, \text{optional } \Diamond\varphi))) \\ & \wedge \text{done}(\text{coordinate}(a, b, \varphi))))). \end{aligned}$$

There still remains the important problem of the consequences of an open-minded agent dropping a social commitment. We assume that he is allowed to do this after communicating and coordinating with his partner. This solution, however, seems not to be subtle enough. We agree that in some cases dropping a social commitment should be more difficult and should cause real consequences for an agent. In future we plan to elaborate on this problem in some more detail.

In these definitions, we assume that φ contains no positive occurrences of *inevitable*, so that *optional* $\Diamond\varphi$ is an O-formula, to which strong realism (see the subsection "Individual goals and intentions") applies.

As we consider agents' behaviour in a multi-agent system, in the section "Planning and acting in different groups", we will analyse the possibilities of cooperation between different agents within a group. It will appear, for example, that agents who seem most trustworthy at first sight, the blindly committed ones, are hard to cooperate with as soon as any kind of replanning is needed.

Collective motivational attitudes

Groups are created on the basis of *collective intentions*, which are defined in the subsection "Collective intentions". We abstract from the ways in which groups

are formed, and refer the interested reader to (Castelfranchi, Miceli, & Cesta 1992; Jennings 1993). We study the behaviour of such groups with respect to both collective intentions and collective commitments, which are defined in the subsection "Collective commitments". The creation of a collective commitment is based on the corresponding collective intention and hinges on the delegation of subtasks according to a plan. However, some agents in the group may not have delegated subtasks while still being involved in the collective intention and the collective commitment. The group exists as long as the collective intention among them exists.

When creating a collective commitment, the group as a whole is known, including the meta-knowledge about the members' commitment strategies. For simplicity we assume that the agent's commitment strategy persists during plan realization, after which it is allowed to change its strategy.

Collective intentions

In order to establish a collective intention among the members of a group, a necessary condition is that all members of the group have the associated individual intention, and that it is a common belief in the group that all members have this intention. However, this condition is certainly not sufficient. Imagine that two agents want to achieve the same goal but are in a competition in which both want to be the only one to achieve the goal. Suppose also that it is a common belief among the two agents that both want to achieve this goal. In this case the agents do not have a collective intention, even though the necessary condition stated above is fulfilled.

This example suggests adding an extra condition in a definition of a collective intention: all members in the group should *intend* the other members to have the associated individual intention (see also (Castelfranchi 1995)); and it should be a common belief in the group that this is so.

In order to formalize the above two conditions for collective intentions among a group G , let us recall that φ stands for a proposition and α for an action.

The formulas $\text{E-INT}_G(\varphi)$ and $\text{E-INT}_G(\alpha)$ ("everyone intends") are syntactically defined in a similar way as the operator for "everyone believes", namely by the following axioms:

$$\text{E-INT}_G(\varphi) \leftrightarrow \bigwedge_{i \in G} \text{INT}(i, \varphi)$$

and

$$\text{E-INT}_G(\alpha) \leftrightarrow \bigwedge_{i \in G} \text{INT}(i, \alpha).$$

Now we are ready to adopt the following axiom for collective intentions (the analogous one for actions α again holds as well):

$$\begin{aligned} \text{C-INT}_G(\varphi) \leftrightarrow & \text{E-INT}_G(\varphi) \wedge \text{C-BEL}_G(\text{E-INT}_G(\varphi)) \\ & \wedge \text{E-INT}_G(\text{E-INT}_G(\varphi)) \\ & \wedge \text{C-BEL}_G(\text{E-INT}_G(\text{E-INT}_G(\varphi))) \end{aligned}$$

Collective commitments

Finally we come to the strongest motivational attitude to be considered in this paper: collective commitment. At the point that a collective commitment to a proposition φ is established, the group of agents has decomposed the goal φ into a number of individual subgoals, a sequence $\langle \varphi_1, \dots, \varphi_n \rangle$ such that truth of φ follows from the truth of $\varphi_1 \wedge \dots \wedge \varphi_n$, and all φ_i are equivalent to formulas of the form *inevitable* $\Diamond\psi$.

These goals are delegated to the agents in the group. In this paper the planning process itself is not studied in detail. Formalization of plans, for example along the lines of (Rao, Georgeff. & Sonenberg 1992), and the influence of planning details on the establishment and maintenance of collective commitments in a group will be subject of further research.

Let us turn to a definition of collective commitment. A collective commitment among a group can only be established or maintained if the group has the associated collective intention.

In addition, for every one of the subgoals that together constitute the collective intention, there should be one agent in the group who is socially committed to at least one (mostly other) agent in the group to fulfil the subgoal; and it should be a common belief among the two agents that the social commitment has been made.

Moreover, there should be a common belief in the whole group that all subgoals have been adopted by committed members of the group.

The defining axiom below reflects all these characteristics.

$$\begin{aligned} \text{C-COMM}_{G,P}(\varphi) \leftrightarrow & \\ & (\text{C-INT}_G(\varphi) \wedge \\ & \bigwedge_{\varphi_i \in P} \bigvee_{a,b \in G} (\text{COMM}(a,b,\varphi_i) \wedge \\ & \text{C-BEL}_{a,b}(\text{COMM}(a,b,\varphi_i))) \wedge \\ & \text{C-BEL}_G\{ \bigwedge_{\varphi_i \in P} \bigvee_{a,b \in G} (\text{COMM}(a,b,\varphi_i) \wedge \\ & \text{C-BEL}_{a,b}(\text{COMM}(a,b,\varphi_i))) \} \end{aligned}$$

Planning and acting in different groups

In this section we present an analysis of cooperation of different kinds of agents within strictly cooperative homogenous and heterogenous groups. It will

be shown how the cooperative possibilities depend on the strength of the commitment strategy. We will also address the question of the scope of the meta-knowledge about commitment strategies and task delegation within the group.

Planning

In this section we claim that during the planning process, knowledge about the characteristic behavior of different kinds of agents is crucial. Let us first study how different commitment strategies influence a group, and then see how a group can make adequate planning choices based on this meta-knowledge. Let us stress that we consider agent and group behaviour in a dynamically changing environment, trying to discover how flexibly different kinds of agents react to unpredictable changes, and in what respects the agents differ in their reactions.

The strongest commitment strategy is followed by the blindly committed agent always working until his goals are achieved. Actually, his behaviour is completely narrow-minded and context-insensitive. In consequence, he is not really capable to cooperate with others: instead, the other members of the group are forced to control the situation.

The next essential question is: in what sense do single-minded and open-minded agents differ in their possibilities of reaction and cooperation? It seems that the difference between single-minded and open-minded agents is a matter of the degree of their autonomy: both kinds can replan on their own (usually in communication and/or coordination with others), but they differ in their possibilities of *adaptation*. A single-minded agent considers changing his goal (or plan in general) only when he has to, namely from the perspective of having lost his beliefs in the realizability of the plan (cf. the subsection "Commitment strategies"). Thus, in order to make his own decision about changing his delegated goals, a belief revision needs to be done first. An open-minded agent, on the other hand, is free in changing his goals even if he still believes in the realizability of his "old" goal (cf. the subsection "Commitment strategies"). Certainly, he is the most autonomous one: to decide about his goals he does not need to revise his beliefs. In other words, open-minded agents have the best possibilities of adaptation: they can immediately react to changes in the environment and in the group, drop their own delegated goal and instead choose a more adequate one.

These crucial characterological differences between agents predispose them to realize different kinds of subtasks. We suggest to take this into account both in planning (where task delegation hinges on it) as well as during plan execution. At the time of task division, groups look for an optimal plan in which tasks can be delegated in an adequate way. Thus, even the division

of the high-level goal φ into subtasks $\varphi_1 \dots \varphi_n$ is made according to the commitment strategies and the abilities of the agents constituting the group, in order to next delegate them to the proper agents.

For example, to blindly committed members of the group only tasks that are strongly believed to be achievable should be delegated. The single-minded agents can take subtasks that should be achieved as long as they are believed to be achievable (the iron-clad subtasks). The open-minded may take responsibility for all other subtasks because of their high adaptive skills.

For flexibility reasons we assume that during the planning stage some agents may not get delegated subtasks: instead, they will be able to pitch in when replanning is needed during plan execution. This is the phase that we turn to next.

Cooperation during plan execution

After planning within a group, the resulting plan is to be realized. Also at this stage the agents' commitment strategies are of importance. A preliminary analysis of the possibilities of cooperation in different groups will be presented.

We start from homogenous groups consisting of only blindly committed agents, always working until their goals are achieved. Only under the naive assumption that all delegated goals are achievable in a given situation, such a group can be successful.

More formally, suppose that a group G consists of blindly committed agents only, and realizes a plan P consisting of the following subgoals:

$inevitable \Diamond \psi_1, \dots, inevitable \Diamond \psi_n$. Suppose moreover that $C-COMM_{G,\rho}(\varphi)$ in such a way that for all $i \leq n$, $COMM(a_i, b_i, inevitable \Diamond \psi_i)$ (where the a_i, b_i are, not necessarily all different, agents in G ; that such a division exists follows directly from the definition of $C-COMM_{G,\rho}(\varphi)$). Then, because all a_i are blindly committed, it follows that for all $i \leq n$,

$inevitable(COMM(a, b, inevitable \Diamond \psi_i) \cup BEL(a, \psi_i))$. This means that if a blindly committed group is successful in reaching the goal of its collective commitment, its plan is carried out exactly as determined initially.

Because in these groups any kind of replanning (coordination) is impossible, as blindly committed agents totally ignore external circumstances, they seem to be hopeless in practice. The obvious reason for this situation is the lack of adaptive skills of blindly committed agents.

While analysing the behaviour of more interesting groups, we will sometimes distinguish the situations when only communication between agents is possible from the situations when also coordination is involved.

In heterogenous groups containing at least one blindly committed agent, others need to exploit their adap-

tive capabilities in order to assure proper plan realization. In fact, if it is not known from the start that delegated goals of blindly committed agents are achievable, the other agents have to be aware and prepared to re-plan in such a way that they realize the main collective intention without the help of those blindly committed members who work on unachievable goals. Note that even if task delegation was done in the proper way, some tasks could become unachievable because of changes in the environment or within the group.

The important question arises when to check the activity of blindly committed members of the group. The good moment for other agents to control how the blindly committed agent is doing on his goal is when either one of the others fails his delegated goal; one of the others depends on the result of the action of the blindly committed agent (in order to go on achieving his own goal); or when one of the agents has already achieved his own goal and has resources to possibly realize the next one.

Suppose that one single-minded agent does not believe that the delegated goal can optionally be achieved anymore and that he communicates this to others. Then the group as a whole has to replan in such a way that this agent does not have the same delegated goal anymore. Depending on circumstances, there are different possible strategies. The most obvious possibility is to give this delegated task to another agent who still believes in its achievability. When no convenient candidate is accessible, a belief revision, where some agents drop their belief in the achievability of the single-minded agent's controversial sub-goal, should be done. In both cases, in order to redivide the goals coordination is needed. However, because belief revision is a lengthy process, the first possibility should be preferred whenever feasible.

When no coordination between agents is possible (e.g. when they can communicate only), after dropping a delegated task by one of the agents, the open-minded agent is the only one capable to assure plan realization by adequate and autonomous reaction to changes in the environment and in the group. So, his presence in the group should be strongly recommended.

Otherwise, when coordination between agents can take place, single and open minded agents act analogously. In this case a single-minded agent who has already achieved his own goal can then go on taking another agent's dropped goal. It is clear that coordination strategies are easier to formulate if task delegation in the group has been done as outlined in the subsection "Planning".

Finally, the main outcome of our discussion is that agents should be equipped with meta-knowledge about commitment strategies and task delegation within the group. Next, we postulate to use this knowledge during planning and plan execution in different kinds of groups: the task delegation should be based on the commitment characters of agents involved.

In our discussion we adopted the assumption that the basis of group creation is a collective intention. Now we see the possibility to go a step further: in a multi-agent system one can create a group that satisfies the commitment properties appropriate to a given plan. After choosing the proper team, all agents should be equipped with the necessary motivational attitudes, knowledge, and meta-knowledge about the characterological characteristics of all agents in the group.

Discussion and conclusions

In this paper we have provided a formalization of social and collective motivational attitudes of individual agents and agents within strictly cooperative groups. Special attention was paid to the strongest attitude: collective commitment. Rao, Georgeff and Sonenberg (Rao, Georgeff, & Sonenberg 1992) consider some related issues with an emphasis on the ontology and semantics of social agents carrying out social plans. They also use a weaker definition of joint intention than ours. Our approach is inspired by Castelfranchi's preformal discussion (Castelfranchi 1995). By the way, a straightforward extension of our framework is to introduce social agents: in principle our definitions are applicable when considering a more generic concept of a collaborative agent.

In this paper we discuss collective plans from another point of view than (Rao, Georgeff, & Sonenberg 1992). At the stage of planning and plan realization we focus on the agents' commitment strategies and mutual dependencies in order to assure a possibly adequate realization of these two stages. In other words, we postulate to base the task delegation on the commitment characteristics of the agents involved in plan realization. To realize this idea we suggest to equip agents with extra meta-knowledge about commitment strategies and task delegation of other agents in the group.

On the basis of individual characteristics of particular agents, their mutual dependencies and other possibly complex criteria one can classify and investigate different types of groups, various types of cooperation, communication, negotiation etc. An interesting extension towards other than strictly cooperative groups will be a subject of our future research.

The results of our preliminary investigation may be of importance when extending the generic BDI-architecture in the multi-agent framework DESIRE so that it will incorporate collective motivational attitudes as well as individual and social ones (Dunin-Kępcicz & Treur 1995; Brazier *et al.* 1996a; 1996b).

Acknowledgements

This work is supported by the Polish KBN Grant 8T11C 03110. We would also like to thank the anonymous referees for their helpful comments.

References

- Åqvist, L. 1984. Deontic logic. In Gabbay, D., and Guenther, F., eds., *Handbook of Philosophical Logic*, volume III. Dordrecht: REIDEL. 605-714.
- Brazier, F.; Dunin-Kępcicz, B.; Jennings, N.; and Treur, J. 1996a. DESIRE: Modelling multi-agent systems in a compositional formal framework. *International Journal of Cooperative Information Systems*. M. Huhns, M. Singh, (eds.), special issue on Formal Methods in Cooperative Information Systems, to appear.
- Brazier, F.; Dunin-Kępcicz, B.; Treur, J.; and Verbrugge, R. 1996b. Beliefs, intentions and DESIRE. Technical report, Department of Mathematics and Computer Science, Vrije Universiteit, Amsterdam. To appear in the Proceedings of the Knowledge Acquisition Workshop KAW'96.
- Castelfranchi, C.; Miceli, M.; and Cesta, A. 1992. Dependence relations among autonomous agents. In Werner and Demazeau (1992).
- Castelfranchi, C. 1995. Commitments: From individual intentions to groups and organizations. In Lesser (1995), 41-48.
- d'Altan, P.; Meyer, J.-J. C.; and Wieringa, R. . 1993. An integrated framework for ought-to-be and ought-to-do constraints. Technical Report IR-342, Department of Mathematics and Computer Science, Vrije Universiteit, Amsterdam. To appear in the Journal of Artificial Intelligence and Law.
- Dunin-Kępcicz, B., and Radzikowska, A. 1995a. Actions with typical effects: Epistemic characterization of scenarios. In Lesser (1995), 445.
- Dunin-Kępcicz, B., and Radzikowska, A. 1995b. Epistemic approach to actions with typical effects. In *Proceedings ECSQARU'95*, 180-189.
- Dunin-Kępcicz, B., and Radzikowska, A. 1995c. Modelling nondeterministic actions with typical effects. In *Proceedings DIMAS'95*, 158-166.
- Dunin-Kępcicz, B., and Treur, J. 1995. Compositional formal specification of multi-agent systems. In Wooldridge, M., and Jennings, N., eds., *Intelligent Agents*, volume 890 of *Lecture Notes in Artificial Intelligence*. Berlin: Springer Verlag, 102-117.
- Fagin, R.; Halpern, J.; Moses, Y.; and Vardi, M. 1995. *Reasoning about Knowledge*. Cambridge, MA: MIT Press.
- Halpern, J., and Moses, Y. 1990. Knowledge and common knowledge in a distributed environment. *Journal of the ACM* 37:549-587.
- Jennings, N. 1993. Commitments and conventions: The foundation of coordination in multi-agent systems. *Knowledge Engineering Review* 3:223-250.
- Lesser, V., ed. 1995. *Proceedings First International Conference on Multi-Agent Systems*. San Francisco: AAAI-Press and MIT Press.

Rao, A., and Georgeff, M. 1991. Modeling rational agents within a BDI-architecture. In Fikes, R., and Sandewall, E., eds., *Proceedings of the Second Conference on Knowledge Representation and Reasoning*, 473-484. Morgan Kaufman.

Rao, A.; Georgeff, M.; and Sonenberg, E. 1992. Social plans: A preliminary report. In Werner and Demazeau (1992), 57-76.

Werner, E., and Demazeau, Y., eds. 1992. *Decentralized A.I.-3*. Amsterdam: Elsevier.