Agent theory for team formation by dialogue

Frank Dignum* Barbara Dunin-Kęplicz† Rineke Verbrugge‡

* Faculty of Mathematics and Computing Science, Technical University Eindhoven
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
dignum@win.tue.nl

† Institute of Informatics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
keplicz@mimuw.edu.pl

‡ Cognitive Science and Engineering, University of Groningen
Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands
rineke@tcw2.ppsw.rug.nl

Abstract. The process of cooperative problem solving can be divided into four stages. First, finding potential team members, then forming a team followed by constructing a plan for that team. Finally, the plan is executed by the team. Traditionally, protocols like the Contract Net protocol are used for performing the first two stages of the process. In an open environment however, there can be discussion among the agents in order to form a team that can achieve the collective intention of solving the problem. For these cases fixed protocols like contract net do not suffice. In this paper we present an architecture for agents that are able to discuss the team formation and subsequently work as a team member until the collective goal has been fulfilled. We also present a solution, using structured *dialogues*, with an emphasis on persuasion, that can be shown to lead to the required team formation. The dialogues are described formally using modal logics and speech acts.

1 Introduction

The area of Distributed Problem Solving (DPS) has been occupied for more than ten years already with solving complex problems by teams of agents. Although in this field the problem might be solved in a distributed way, the control of solving it usually lies with one agent. This means that the team of agents is either already available or can be created on demand by the controlling agent. Also, these teams of agents are collaborative by nature. Thus they are designed to participate in the team to solve the problem.

We are concerned with problems that have to be solved by a number of existing agents that are not designed to solve this particular problem together. In this case, getting the right team of agents and controlling them is of prime interest. In this setting Contract Net [17] is often proposed as a simple but effective and efficient way to distribute tasks over a number of agents in order to achieve a common goal. It basically makes use of the market mechanism of task demand and supply to match tasks with agents that are willing to perform them. The reason of the

success of this approach is the fact that it uses a fixed protocol with a limited number of steps and thus is easy to implement.
In our opinion, however, this functions only in cases where the market mechanism works, namely if several agents are willing to do the task (and these can compete), and if the tasks are all well described beforehand. We will concentrate on cases where these conditions are usually not met. Either because there is only one agent capable of performing a task and that one should be negotiated with or because the task cannot be described precisely at the beginning.
A good candidate to make conversation between agents during Cooperative Problem Solving (CPS) more flexible is the theory of Dialogue [20]. This theory gives rules for appropriate moves within different types of dialogues. The rules direct the dialogue without completely fixing the order of the moves. These moves themselves depend on particular stages of CPS. In fact, all of them are rather complex, both from the MAS and the AI perspective. Additionally, the cooperative process takes place in a dynamic and often unpredictable environment. Our intention is to present formal means allowing realisation of some relevant forms of dialogue. These dialogue types follow a formal theory of dialogue (see [20]) and speech acts (see [18]). In this way, it will be possible to prove that in given circumstances the dialogue results in a certain outcome. This is of prime importance if you want to construct a MAS for automated CPS.
We base our investigation on a formal model of teamwork. Thus, four stages of CPS are distinguished according to Wooldridge and Jennings' paper [21]. We define the stages somewhat differently, however (see [9] for a discussion). The first stage is *potential recognition* in which the agent that takes the initiative tries to find out which agents are potential candidates for achieving a given overall goal and how these can be combined in a team. The second stage is *team formation*. The result of this stage is a *collective intention* among a team to achieve the overall goal to solve the problem. This is the team that will try to actually achieve the goal. The third stage is *plan formation*. Here the team divides the goal into subtasks, associates these with actions and allocates these actions to team members. In terms of motivational attitudes the end result of this stage is a *collective commitment* to perform the social plan that realizes the goal. The last stage is *plan execution* in which the team members execute the allocated actions and monitor the appropriate colleagues. If necessary a *reconfiguration* of the plan can be constructed [9].
We concentrate on the first two stages of the process and show how the collective intention resulting from team formation is built up during the dialogues. It turns out that the main type of dialogue needed is persuasion. The running example in this paper is about team formation for achieving the following overall goal (further abbreviated as φ): "to arrange a trip of three weeks to Australia for a certain famous family; the trip should satisfy (specific) constraints on costs, times, places and activities". The initiative for teamwork is taken by travel agent a, who cannot arrange the whole trip on his own. The trip will be extensively publicized, so it has to be a success, even if circumstances change. Thus, he does not simply ask airline companies, hotels, and organizers of activities to deliver a number of fixed services. Instead, he believes that real teamwork, where all members are interested in the achievement of the overall goal, gives the best chances of a successful trip. This paper discusses the first two stages of cooperative problem solving only; plan formation and team action for this example will be described in further work.

The paper is structured in the following manner. A brief typology of dialogue is given in Section 2. Section 3 presents the logical background and briefly describes an agent architecture for the first two stages of CPS. Sections 4 and 5 investigate different dialogue types during potential recognition and team formation, respectively. Finally, in section 6 conclusions and further research are discussed. This paper is a revised and extended version of [5].

2 Conversations in CPS

In this section we will briefly discuss some theory to describe conversations.
Conversations are sequences of messages between two (or more) agents. These sequences can be completely fixed as is done by the use of a fixed protocol which states exactly which message should come next. Conversations can also be seen as completely free sequences of messages. In this case the agents can decide at any moment what will be the next message they send. We have already argued that fixed protocols are too rigid for the situation that we describe. However, complete freedom of choosing the next message would also be impractical. This would put a heavy burden on the agent that has to choose at each point in time which message it might send.
This is the reason that we choose for a form in between the two extremes sketched above. Dialogue theory structures conversations by means of a number of dialogue rules. These rules limit the number of possible responses at each point, while not completely fixing the sequence of messages. The agents speak in turn, for example asking questions and giving replies, and take into account, at each turn, what has occurred previously in the dialogue.
Krabbe and Walton [20] provide a typology of dialogue types between two agents, with an emphasis on the persuasion dialogue. They create a *normative model*, representing the ideal way reasonable, cooperative agents participate in the type of dialogue in question. For each type of dialogue, they formulate *an initial situation, a primary goal*, and *a set of rules*. Below, their typology is briefly explained and adapted to the CPS. In the course of communication among agents, there often occurs a shift from one type of dialogue to another, in particular *embedding* occurs when the second dialogue is functionally related to the first one.
A persuasion dialogue arises from a conflict of opinions. It may be that one agent believes φ while some others either believe a contrary proposition ψ (where $\varphi \wedge \psi$ is inconsistent) or just have doubt about φ. The goal of a persuasion dialogue is to resolve the conflict by verbal means, in such a way that a stable agreement results, corresponding to a collective informational attitude. In contrast to [20], we allow persuasion also with respect to motivational attitudes.
The initial situation of *negotiation* is a conflict of interests, together with a need for cooperation. The main goal is to make a deal. Thus, the selling and buying of goods and services, that is often described in the MAS literature, is only one of the many contexts where negotiation plays a role. Negotiation and persuasion are often not distinguished adequately.
The initial situation of *information seeking* occurs when one agent is ignorant about the truth of a certain proposition and seeks information from other agents.
Two other important types of dialogue are *inquiry* and *deliberation*, but they play a role mainly during the stage of plan formation, which will be considered in a forthcoming paper. The last type of dialogue, *eristics* (verbal fighting between agents), is not relevant as we focus on teamwork.

In general, Krabbe and Walton [20] are not interested in informational and motivational attitudes of agents involved in dialogue, if these attitudes are not communicated explicitly. In contrast to them, our goal is to make the whole process of dialogue among computational agents transparent. For this reason, at each step of team formation the agents' internal attitudes need to be established, and then updated and revised when necessary.

3 Logical background and agent architecture

We view individual beliefs, goals, and intentions as primitive notions and use well-known modal logics to formalize them. For the epistemic operator BEL, referring to agent's belief, the standard modal system $\mathrm{KD45}_n$ for n agents is used [13]. For the motivational operators GOAL and INT, referring to agent's goals and intentions, respectively, the axioms include the axiom system K_n. As to possible additional axioms, we assume that an agent's goals need not be consistent with each other. It chooses a limited number of its goals to be intentions, in such a way that consistency is preserved. This can be formulated by the well-known axiom D:

$$\neg\mathrm{INT}(i, \bot) \text{ for } i = 1, \ldots, n \quad \text{Intention Consistency Axiom}$$

Thus, for intentions (but not for goals, in contrast to [19]), the resulting system is KD_n.
As for Kripke semantics, formulas are evaluated with respect to pairs (M, w) of a model M and a world w, using binary accessibility relations B_i, D_i, I_i between such model-world pairs, corresponding to each agent i's beliefs, goals, and intentions in the standard way. The relations B_i and I_i are serial, as characterized by axiom D. Moreover, the B_i are also transitive and euclidean, corresponding to axioms 4 and 5 for positive and negative introspection, respectively.
In this paper, intentions, goals, and beliefs all refer to agents' attitudes towards *propositions* ψ. (In other work, motivational attitudes are also defined with respect to actions [8, 10].) We formalize the fact that agents are correct about their own intentions by the axiom $\mathrm{BEL}(i, \mathrm{INT}(i, \psi)) \rightarrow \mathrm{INT}(i, \psi)$, which characterizes the frames in which for all worlds u, v, if uI_iv, then there is a w such that uB_iw and wI_iv. We also find it useful to formalize the fact that the set of an agent's intentions is a subset of its goals, as expressed by the axiom $\mathrm{INT}(i, \psi) \rightarrow \mathrm{GOAL}(i, \psi)$, which characterizes the Kripke frames for which $D_i \subseteq I_i$.

3.1 Collective beliefs and collective intentions

To model teamwork, individual attitudes naturally do not suffice. In other work we discussed the pairwise notion of social commitments, as well as collective notions like collective belief, collective intention, and collective commitment [10]. In the first two stages of CPS, the essential notions are those of collective belief and collective intention.
In order to model group beliefs, a few axioms and rules need to be added to the System $\mathrm{KD45}_n$. First, the formula $\text{E-BEL}_G(\varphi)$ is meant to stand for "every agent in group G believes φ", axiomatized as follows:

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

A stronger operator is the one for collective belief, which is similar to the more usual common knowledge. $\text{C-BEL}_G(\varphi)$ is meant to be true if everyone in G believes φ, everyone in G believes that everyone in G believes φ, etc. (Note that even collective beliefs need not be true, so $\text{C-BEL}_G(\varphi)$ need not imply φ.) Collective belief is axiomatized by the following axiom and rule. Note that the rule may only be applied if the antecedent $\varphi \to \text{E-BEL}_G(\psi \wedge \varphi)$ is a theorem.

(2) $\text{C-BEL}_G(\varphi) \leftrightarrow \text{E-BEL}_G(\varphi \wedge \text{C-BEL}_G(\varphi))$
(3) From $\varphi \to \text{E-BEL}_G(\psi \wedge \varphi)$ infer $\varphi \to \text{C-BEL}_G(\psi)$ (Induction Rule)

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 [13].

The establishment of collective beliefs among a group is problematic. In [13, Chapter 11] it is shown that bilateral sending of messages does not suffice to determine collective belief if communication channels may be faulty. Some possible practical solutions are also given. We abstract from these problems by assuming collective beliefs can be achieved by global announcements.

The definition of *collective intention* is rather strong, because we focus on strictly cooperative groups. There, a necessary condition for a collective intention is that all members of the group have the associated individual intention $\text{INT}(i, \varphi)$. Moreover, to exclude the case of competition, all agents should *intend* all members to have the associated individual intention, as well as the intention that all members have the individual intention, and so on; we call such a mutual intention $\text{M-INT}_G(\varphi)$. Furthermore, all members of the group are aware of this mutual intention, that is, they have a collective belief about this ($\text{C-BEL}_G(\text{M-INT}_G(\varphi))$). In order to formalize the above two conditions, $\text{E-INT}_G(\varphi)$ (standing for "everyone intends") is defined by the following axiom, analogous to (1) above:

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

The mutual intention $\text{M-INT}_G(\varphi)$ is axiomatized by an axiom and rule analogous to (2) and (3):

(5) $\text{M-INT}_G(\varphi) \leftrightarrow \text{E-INT}_G(\varphi \wedge \text{M-INT}_G(\varphi))$
(6) From $\varphi \to \text{E-INT}_G(\psi \wedge \varphi)$ infer $\varphi \to \text{M-INT}_G(\psi)$ (Induction Rule)

The system resulting from adding (4), (5), and (6) to KD_n is called $KD_n^{\text{M-INT}_G}$, and it is sound and complete with respect to Kripke models where all n accessibility relations are serial (by a proof analogous to the one for common knowledge in [13]). Finally, the collective intention is defined by the following axiom:

(7) $\text{C-INT}_G(\varphi) \leftrightarrow (\text{M-INT}_G(\varphi) \wedge \text{C-BEL}_G(\text{M-INT}_G(\varphi)))$

Note that this definition is different from the one given in [8, 10]; the new definition will be extensively discussed and compared with others in a forthcoming paper. Let us remark that, even though $\text{C-INT}_G(\varphi)$ seems to be an infinite concept, a collective intention with respect to φ may be established in a finite number of steps: according to (6) and (7), it suffices that all agents in G intend $\varphi \wedge \text{M-INT}_G(\varphi)$ and that this fact is announced to the whole group by which a collective belief is established.

3.2 Agent architecture for team formation by dialogue

In order to ensure the proper realization of team formation by dialogue, we postulate that an agent architecture should contain a number of specific modules. The heart of the system is, as usual, the **reasoning** module. When realizing the consecutive stages leading ultimately to team formation, interaction with the **planning**, **communication**, and **social reasoning** modules is necessary. All these modules contain a number of specific *reasoning rules* which will be introduced formally in the sequel. Each rule refers to a specific aspect of the reasoning process; very often they can be viewed as rules bridging different modules.
The **reasoning** module contains rules creating agents' individual beliefs, intentions, and goals, as well as the collective informational and motivational attitudes established during potential recognition and team formation. The **communication** module contains rules leading to speech acts, but also some auxiliary rules or mechanisms organizing the *information seeking* and *persuasion* dialogues. In this paper we abstract from the latter, rather technical, ones. The **social reasoning** module contains rules allowing to reason about other agents, while the **planning** module will be used for preplanning during the potential recognition stage. In fact, it will be more busy during higher stages of CPS, i.e. when generating a social plan and possibly during the reconfiguration process (see [9, 11]).

4 Potential recognition

Potential recognition is about finding the set of agents that may participate in the formation of the team that tries to achieve the overall goal. These agents are grouped into several potential teams with whom further discussion will follow during *team formation*. For simplicity, we assume that one agent takes the initiative to realize the overall goal, which is given.

4.1 The end result of potential recognition

The first task of the *initiator* is to form a partial (abstract) plan for the achievement of the overall goal. On the basis of the (type of) subgoals that it recognizes it will determine which agents might be most suited to form the team. In order to determine this match the initiator tries to find out the properties of the agents, being interested in three aspects, namely their *abilities*, *opportunities*, and their *willingness* to participate in team formation.
The aspect of ability considers whether the agents can perform the right type of tasks. It does not depend on the situation, but may be viewed as an inherent property of the agent itself. The aspect of opportunity takes into account the possibilities of task performance in the present situation, involving resources and possibly other properties. The aspect of willingness considers their mental attitudes towards participating towards the overall goal. Very capable agents that do not want to do the job are of no use. The components of the agent's suitability are represented as follows (cf. [15] where formal definitions are given for ability and opportunity based on dynamic logic):

1. the individual ability of agent b to achieve a goal ψ is denoted by $able(b, \psi)$, meaning that ψ is in a pre-given set of agent b's abilities.

2. the resources available to agent b are reflected by the opportunity that agent b has to achieve ψ, denoted by $opp(b, \psi)$. This is true in a world and a state if there is a branch through that state such that at a later point on that branch, agent b has just seen to it that ψ.
3. the willingness of agent b to participate in team formation is denoted by $willing(b, \varphi)$. This stands for "agent b believes that at some branch, at some point in future, it intends φ".

4.2 Towards a potential of cooperation

In the previous subsection we have described what type of information the initiating agent tries to gather in order to start team formation. Now, we will describe how the information is collected, leading to the formal outcome.
The output at this stage is the "potential for cooperation" that the initiator a sees with respect to φ, denoted as $\mathrm{POTCOOP}(a, \varphi)$, meaning that φ is a goal of a ($\mathrm{GOAL}(a, \varphi)$), and there is a group G such that a believes that G can collectively achieve φ (C-$\mathrm{CAN}_G(\varphi)$) and are willing to participate in team formation; and either a cannot or doesn't desire to achieve φ in isolation. This is expressed by the following definitions (see [9] for more discussion):

$$\begin{aligned}&\mathrm{POTCOOP}(a, \varphi) \leftrightarrow \mathrm{GOAL}(a, \varphi) \wedge \\ &\exists G \subseteq T \mathrm{BEL}(a, \text{C-}\mathrm{CAN}_G(\varphi) \wedge \forall i \in G willing(i, \varphi)) \wedge \\ &(\neg \mathrm{CAN}(a, \varphi) \vee \neg \mathrm{GOAL}(a, stit(\varphi)))\end{aligned}$$

$$\mathrm{CAN}(a, \varphi) \leftrightarrow able(a, \varphi) \wedge opp(a, \varphi).$$

$\mathrm{POTCOOP}(a, \varphi)$ is derived from the information collected from the individual agents. To derive C-$\mathrm{CAN}_G(\varphi)$ the initiator compares the information obtained about the other agents against a partial abstract plan for the overall goal φ. For this purpose φ is split into a number of subgoals $\varphi_1, \ldots, \varphi_n$, which can be viewed as instrumental to the overall goal. Together they *realize* φ and are compared with the individual abilities and opportunities that the agents are believed to have:

$$\begin{aligned}&\text{C-}\mathrm{CAN}_G(\varphi) \leftrightarrow \exists \varphi_1, \ldots, \exists \varphi_n (realize(< \varphi_1, \ldots, \varphi_n >, \varphi) \wedge \\ &\forall i \leq n \exists j \in G(able(j, \varphi_i) \wedge opp(j, \varphi_i)))\end{aligned}$$

Here, $realize(< \varphi_1, \ldots, \varphi_n >, \varphi)$ intuitively means: "if $\varphi_1, \ldots, \varphi_n$ are achieved, then φ holds". Technically, one may need some extra-logical, context-dependent, reasoning to show the implication. In the reasoning module of the agent architecture, there is a formal rule corresponding to: "if $< \varphi_1, \ldots, \varphi_n >$ is a result of pre-planning to achieve φ, then $realize(< \varphi_1, \ldots, \varphi_n >, \varphi)$ holds".

4.3 Information seeking dialogue

Ultimately, the initiator has to form beliefs about the abilities, opportunities, and willingness of the individual agents in order to derive $\mathrm{POTCOOP}(a, \varphi)$. The possible strategies for organizing this information seeking part remain out of our interest. For example, a may first investigate the willingness of particular agents, and on this basis then ask the interested ones about their abilities and opportunities. Depending on the specific situation we deal with, another organization of this process may be more adequate. In any case, the questions in this stage form

part of an *information seeking* dialogue. This can be done by a asking every agent about its properties and the agent responding with the requested information.
Formally this can be expressed as follows, where ψ may stand for any of the aspects $opp(i, \varphi_i)$, $able(i, \varphi_i)$, $willing(i, \varphi)$, etc.

$$[\mathrm{REQ}_{a,i}(\textbf{ if } \psi \textbf{ then } \mathrm{ASS}_{i,a}(\psi) \textbf{ else } \mathrm{ASS}_{i,a}(\neg\psi))]$$
$$[\mathrm{ASS}_{i,a}(\psi)](\mathrm{TRUST}(a, i, \psi) \rightarrow \mathrm{BEL}(a, \psi))$$

The above formula is based on the formal theory on *speech acts* developed in [18] and [7]. It is expressed as a dynamic logic formula of the form $[\alpha_1][\alpha_2]\psi$, meaning that if α_1 is performed then always a situation arises such that if α_2 is performed then in the resulting state ψ will always hold. In the above case α_1 is the complex action $\mathrm{REQ}_{a,i}(\textbf{ if } \psi \textbf{ then } \mathrm{ASS}_{i,a}(\psi) \textbf{ else } \mathrm{ASS}_{i,a}(\neg\psi))$, where $\mathrm{REQ}_{a,i}(\alpha)$ stands for agent a requesting agent i to perform the action α. After this request i has three options.

1. It can simply ignore a and not answer at all.
2. It can state that it is not willing to divulge this information:
   $\mathrm{ASS}_{i,a}(\neg(\textbf{ if } \psi \textbf{ then } \mathrm{ASS}_{i,a}(\psi) \textbf{ else } \mathrm{ASS}_{i,a}(\neg\psi)))$.
3. It can state that it does not have enough information:
   $\mathrm{ASS}_{i,a}(\neg(\mathrm{BEL}(i, \psi) \wedge \neg\mathrm{BEL}(i, \neg\psi)))$.
4. It can either assert that ψ is the case (as described above) or that it is not, in which case a believes ψ is not true:

$$[\mathrm{ASS}_{i,a}(\neg\psi)](\mathrm{TRUST}(a, i, \neg\psi) \rightarrow \mathrm{BEL}(a, \neg\psi)).$$

Of course in case 2, agent a can already derive that i is not willing to achieve φ as part of a team; only in case 4 will a have a resulting belief about ψ.
An important notion here is whether a trusts the other agent's answer; we use $\mathrm{TRUST}(a, i, \psi)$ to mean "agent a trusts agent i with respect to proposition ψ". Trust is a complex concept that has been defined in many ways, from different perspectives (see [3] for some current work in this area). We will not try to define trust in any way in this paper, but simply use its intuitive meaning. We suppose that the module social reasoning contains rules giving a number of conditions (e.g. "a has observed that agent i is generally trustworthy") implying $\mathrm{TRUST}(a, i, \psi)$.
In an analogous way, an information seeking dialogue about all ingredients of $\mathrm{POTCOOP}(a, \varphi)$ takes place. In principle, the schema of all necessary questions may be rather complex. For example, when recognizing the ability to achieve a specific subgoal φ_i, agent a should repeat this question for all the subgoals it distinguished to every agent, but this solution is apparently not acceptable from the AI perspective. Of course it is more effective for agent a to ask each agent i to divulge all the abilities it has with respect to achieving this set of goals. These are, however, strategic considerations, not related to the theory of dialogue.
A next strategic point is about case 1. To avoid agent a waiting indefinitely for an answer, we assume that every speech act has an implicit deadline for reaction incorporated. After this deadline, the silent agent will not be considered as a potential team member anymore. The logical modeling of these types of deadlines is described in [6] and will not be pursued here.
Finally, the formal result of the potential recognition stage is that agent a has the belief that it is possible to form a team to achieve the overall goal or that it is not. In some applications, like scientific research projects, the outcome of this stage is communicated to the agents who replied positively. In the context of services like travel agencies this may not be necessary.

5 Team formation

At the stage of potential recognition, individual properties of agents were considered. These could play a role in different types of cooperation, where some services are exchanged but there is no interest in other agent's activities, and more specifically, there is no collective intention to achieve a goal as a team. At the stage of team formation, however, the conditions needed for teamwork in a strict sense are created. The main condition of teamwork is the presence of a collective intention, defined in subsection 3.1.

Note that this concept of teamwork requires agents that have a type of "social conscience". We do not consider a set of agents as a team if they cooperate by just achieving their own predefined part of a common goal. If agents are part of a team they should be interested in the performance of the other team members and willing to adjust their task on the basis of the needs of others.

At the beginning, the initiator has a sequence of groups in mind that could be suitable to form a team for achieving the goal. Although we do not go into details here, the organization of the groups into a sequence is important for the case that the most desirable group can not be formed.

All agents in these potential teams have expressed their willingness to participate towards the overall goal, but do not necessarily have the individual intention to contribute towards it yet. In this situation, the initiator tries to persuade them to take on the intention to achieve the overall goal as well as to act as a team with the other members. Note that task division and task allocation only come to the fore at the next stage, plan formation.

5.1 Persuasion dialogue

The goal of a persuasion dialogue is to establish a collective intention within a group G to reach the overall goal φ (C-$\mathrm{INT}_G(\varphi)$, see subsection 3.1). Axiom (7) makes evident that a crucial step for the initiator is to persuade all members of a potential team to take the overall goal as an individual intention. To establish the higher levels of the mutual intention, the initiator also persuades each member to take on the intention that all members of the potential team have the mutual intention, in order to strengthen cooperation from the start. It suffices if the initiator persuades all members of a potential team G to take on an individual intention towards φ ($\mathrm{INT}(i, \varphi)$) and the intention that there be a mutual intention among that team ($\mathrm{INT}(i, \text{M-INT}_G(\varphi))$). This results in $\mathrm{INT}(i, \varphi \wedge \text{M-INT}_G(\varphi))$ for all $i \in G$, or equivalently by axiom (4): $\text{E-INT}_G(\varphi \wedge \text{M-INT}_G(\varphi))$, which in turn implies by axiom (5) that $\text{M-INT}_G(\varphi)$. When all the individual motivational attitudes are established within the team, the initiator broadcasts the fact $\text{M-INT}_G(\varphi)$, by which the necessary collective belief $\text{C-BEL}_G(\text{M-INT}_G(\varphi))$ is established and the collective intention is in place.

This will be achieved during a persuasion dialogue, which according to [20] consists of three main stages: information exchange, rigorous persuasion and completion. In our case the information exchange already started in the potential recognition stage. Let us remind the reader that we extended the concept of persuasion to include also intentions, and not only beliefs. The final result of the team formation stage is reached when for one potential team all the persuasion dialogues have been concluded successfully.

Information exchange During the information exchange the agents make clear their initial stand with respect to the overall goal and to form part of a certain team to achieve it. These issues are expressed partly in the form of intentions and beliefs. Other beliefs supporting or related to the above issues might also be exchanged already. Only when a conflict arises about these issues a persuasion dialogue has to take place. In each persuasion there are two parties or roles; the proponent (P) and the opponent (O). In our case the proponent is the initiator and the opponent the other agent.

The stands the other agents take about the above issues are seen as its initial *concessions*. Concessions are beliefs and intentions that an agent takes on for the sake of argument, but need not be prepared to defend. The agents will also have private attitudes that may only become apparent later on during the dialogue. The stand of the initiator is seen as the initial thesis that it is prepared to defend during the dialogue. The initial conflict description consists of a set of O's initial concessions and P's initial thesis.

Rigorous persuasion During the rigorous persuasion stage the agents exchange arguments to challenge or support a thesis. The following rules can be used to govern these moves adapted from [20]:

1. Starting with O the two parties move alternately according to the rules of the game.
2. Each move consists of either a challenge, a question, a statement, a challenge or question accompanied by a statement, or a final remark.
3. The game is highly asymmetrical. All P's statements are assertions, and called *theses*, all O's statements are called *concessions*. P is doing the questioning and O all the challenging.
4. The initial move by O challenges P's initial thesis. It is P's goal to make O concede the thesis. P can do this by questioning O and thus bridge the gap between the initial concessions of O and the thesis, or by making an assertion to clinch the argument if acceptable.
5. Each move for O is to pertain to P's preceding move. If this was a question, then O has to answer it. If it was an assertion, then O has to challenge it.
6. Each party may give up, using the final remark $\mathrm{ASS}_{a,i}(quit)$ for the initiator, or $\mathrm{ASS}_{i,a}(\mathrm{INT}(i, \varphi \wedge \text{M-INT}_G(\varphi)))$.
   If O's concessions imply P's thesis, then P can end the dialogue by the final remark: $\mathrm{ASS}_{a,i}(won)$. In our system we assume to have the following rule: $[\mathrm{ASS}_{a,i}(won)]\mathrm{OBL}(\mathrm{ASS}_{i,a}(\mathrm{INT}(i, \varphi \wedge \text{M-INT}_G(\varphi))))$, which means that agent i is obliged to state that it has been persuaded and accepts its role in the team. This does not mean that i will actually make this assertion! Just that there is an obligation (according to the rules of the persuasion "game"). The modal operator OBL is taken from deontic logic [1].
7. All challenges have to follow logical rules. For example, a thesis $A \wedge B$ can be challenged by challenging one of the two conjuncts. For a complete set of rules for the propositional part of the logic we refer to [20].

In the completion stage the outcome is made explicit, such that the agents either have a collective belief and/or intention or they know that they differ in opinion.

Speech acts during persuasion In contrast to [20], we need to monitor the agent's informational and motivational attitudes during persuasion. We are concerned with assertions and challenges (wrt. informational attitudes), and concessions and requests (wrt. both informational and motivational attitudes).

As for assertions, after a speech act of the form $\mathrm{ASS}_{a,i}(B)$, agent i believes that the initiator believes that B:

$$[\mathrm{ASS}_{a,i}(B)]\mathrm{BEL}(i, \mathrm{BEL}(a, B))$$

Let us assume that i has only two rules for answering an assertion B. If i does not have a belief that is inconsistent with B then i will concede (similarly as in default logic). If, on the other hand, i does have a belief to the contrary it will challenge the assertion. Formally:

$$\neg\mathrm{BEL}(i, \neg B) \rightarrow \mathrm{DO}(i, \mathrm{CONCEDE}_{i,a}(B))$$

$$\mathrm{BEL}(i, \neg B) \rightarrow \mathrm{DO}(i, \mathrm{CHALLENGE}_{i,a}(B))$$

where the operator $\mathrm{DO}(i, \alpha)$ indicates that α is the next action performed by i.

The CONCEDE action with respect to informational attitudes is basically an assertion plus a possible mental update of the agent. Thus, it does not only assert the proposition but actually believes it as well, even if it did not believe the proposition beforehand. Suppose that i did not have a contrary belief, then i concedes B by the speech act $\mathrm{CONCEDE}_{i,a}(B)$. The effect of this speech act is similar to that of ASS, except that a can only assume that i believes the formula B during the dialogue and might retract it afterwards.

$$[\mathrm{CONCEDE}_{i,a}(B)]\mathrm{BEL}(a, \mathrm{BEL}(i, B))$$

The CHALLENGE with respect to informational attitudes, on the other hand, is a combination of a denial (assertion of a belief in the negation of the proposition) and a request to prove the proposition. The exact form of the challenge depends on the logical form of the assertion [20]. For this reason, the complete effects of this speech act are quite complex to describe fully. We will give an example challenge in the next subsection.

With respect to motivational attitudes, the situation is different. For example, initiator a requests i to take on an intention ψ by the following speech act:

$$\mathrm{REQ}_{a,i}(\mathrm{CONCEDE}_{i,a}(\mathrm{INT}(i, \psi))).$$

Again, i has only two rules for answering such a request. If i does not have an intention $\neg\psi$ (that is inconsistent with ψ) then i will concede. If, on the other hand i does have an intention to the contrary it will assert that it intends $\neg\psi$:

$$\neg\mathrm{INT}(i, \neg\psi) \rightarrow \mathrm{DO}(i, \mathrm{CONCEDE}_{i,a}(\mathrm{INT}(i, \psi)))$$

$$\mathrm{INT}(i, \neg\psi) \rightarrow \mathrm{DO}(i, \mathrm{ASS}_{i,a}(\mathrm{INT}(i, \neg\psi)))$$

For example, suppose that i did not have a contrary intention then i concedes by the speech act $\mathrm{CONCEDE}_{i,a}(\mathrm{INT}(i, \psi))$. The effect of this speech act is:

$$[\mathrm{CONCEDE}_{i,a}(\mathrm{INT}(i, \psi))]\mathrm{BEL}(a, \mathrm{INT}(i, \psi))$$

5.2 Team formation for the example

In the travel example, the initiator tries to persuade the other agents i in the potential team to take on the intention to achieve the overall goal of organizing the journey ($\mathrm{INT}(i, \varphi)$), but also with respect to doing this as a team with the other agents ($\mathrm{INT}(i, \text{M-INT}_G(\varphi))$). To this end, the initiator exploits the theory of intention formation.

Intentions are formed on the basis of beliefs and previously formed intentions of a higher abstraction level by a number of formal rules (see [4]). For example, the built-in intention can be to obey the law, or avoid punishment. The (instrumental) belief is that driving slower than the speed limit is instrumental for obeying the law, and is its preferred way to do so. Together with the rule the new intention of driving slower than the speed limit is derived.

The general intention generation rule is represented as follows:

(8) $\mathrm{INT}(i, \psi) \wedge \mathrm{BEL}(i, \mathrm{INSTR}(i, \chi, \psi)) \wedge \mathrm{PREFER}(i, \chi, \psi) \rightarrow \ \mathrm{INT}(i, \chi)$

It states that if an agent i has a intention ψ and it believes that χ is instrumental in achieving ψ and χ is its preferred way of achieving ψ, then it will have the intention to achieve χ. "χ is instrumental in achieving ψ" means that achieving χ gets the agent "closer" to ψ in some abstract sense. We do not define this relation any further, but leave it as primitive.

The PREFER relation is based on an agent's individual beliefs about the utility ordering between its goals, collected here into a finite set H. We abstract from the specific way in which the agent may compute the relative utilities, but see the literature about (qualitative) decision theory [2].

$$\mathrm{PREFER}(i, \chi, \psi) \leftrightarrow$$
$$\bigwedge_{\xi \in H} (\mathrm{BEL}(i, \mathrm{INSTR}(i, \xi, \psi)) \rightarrow \mathrm{BEL}(i, ut(i, \chi) \geq ut(i, \xi)))$$

The mechanism sketched in subsection 5.1 can be used in our setting during persuasion. In our example, the initiator a tries to get the other agent i to concede to higher level intentions, instrumental beliefs and preferences that together with (8) imply the intention to achieve the overall goal φ. To be more concrete, we could choose the higher level intention ψ to stand for "earn good money". Here follows an example move of the initiator:

$$\mathrm{ASS}_{a,i}(\forall j(\mathrm{INT}(j, \psi) \rightarrow \mathrm{INSTR}(j, \varphi, \psi))).$$

After this speech act agent i believes that the initiator believes that if an agent has the higher level intention to earn good money, then the overall intention φ is instrumental to this. Formally (see also [7]):

$$[\mathrm{ASS}_{a,i}(\forall j(\mathrm{INT}(j, \psi) \rightarrow \mathrm{INSTR}(j, \varphi, \psi)))]$$
$$\mathrm{BEL}(i, \mathrm{BEL}(a, \forall j(\mathrm{INT}(j, \psi) \rightarrow \mathrm{INSTR}(j, \varphi, \psi))))$$

According to the general rule about assertions there are two possibilities for i's answer. Let us assume that the positive case holds, i.e. i does not have a contrary belief, so it concedes: $\mathrm{CONCEDE}_{i,a}(\forall j(\mathrm{INT}(j, \psi) \rightarrow \mathrm{INSTR}(j, \varphi, \psi)))$. The effect of this speech act on agent a is given by the general rule:

$$[\mathrm{CONCEDE}_{i,a}(\forall j(\mathrm{INT}(j, \psi) \rightarrow \mathrm{INSTR}(j, \varphi, \psi)))]$$
$$\mathrm{BEL}(a, \mathrm{BEL}(i, \forall j(\mathrm{INT}(j, \psi) \rightarrow \mathrm{INSTR}(j, \varphi, \psi))))$$

Now the formula is believed by both a and i. Thus, the initiator's next aim in the persuasion will be to get i to intend ψ (earn good money) by the question:

$$\mathrm{REQ}_{a,i}(\mathrm{CONCEDE}_{i,a}(\mathrm{INT}(i,\psi))).$$

By the general rule, i is obliged to either concede it has the intention ψ (if it is consistent with its other intentions) or to assert that it intends its negation. After i's response, the initiator believes i's answer. Note that in the second case, it may be useful for a to embed a negotiation dialogue in the persuasion, in order to get i to revise some of its previous intentions. For the example, let us suppose that a is successful in persuading i.

When the initiator has persuaded agent i to take on high level intention ψ and to believe the instrumentality of φ with respect to ψ, it can go on to persuade the other that $\mathrm{PREFER}(i,\phi,\psi)$ by the speech act:

$$\mathrm{ASS}_{a,i}(\bigwedge_{\xi\in H}(\mathrm{BEL}(i,\mathrm{INSTR}(i,\xi,\psi))\rightarrow \mathrm{BEL}(i,ut(i,\varphi)\geq ut(i,\xi))))$$

To make the example more interesting, let us suppose that i does not yet prefer φ as a means to earn good money; instead it believes that χ, arranging some less complex holidays for another family, has a higher utility than φ. Thus i does not concede to the initiator's speech act, but instead counters with a challenge. According to the logical structure of the definition of PREFER, this challenge is a complex speech act consisting of three consecutive steps. First i asserts the negation of a's assertion, a conjunction of implications; then it concedes to the antecedent of the implication for the specific goal $\chi\in H$; and finally it requests a to present a proof that φ has a better utility for i than χ.

$$\begin{array}{l}\mathrm{CHALLENGE}_{i,a}\\ (\bigwedge_{\xi\in H}(\mathrm{BEL}(i,\mathrm{INSTR}(i,\xi,\psi))\rightarrow \mathrm{BEL}(i,ut(i,\varphi)\geq ut(i,\xi))))\equiv\\ \mathrm{ASS}_{i,a}(\neg(\bigwedge_{\xi\in H}(\mathrm{BEL}(i,\mathrm{INSTR}(i,\xi,\psi))\rightarrow \mathrm{BEL}(i,ut(i,\varphi)\geq ut(i,\xi))));\\ \mathrm{CONCEDE}_{i,a}(\mathrm{BEL}(i,\mathrm{INSTR}(i,\chi,\psi)));\\ \mathrm{REQ}_{i,a}(\mathrm{ASS}_{a,i}(\mathrm{PROOF}(ut(i,\varphi)\geq ut(i,\chi))))\end{array}$$

As a reply, a could prove that the utility of φ is in fact higher than that of χ, because it generates a lot of good publicity, which will be profitable for i in future – something of which i was not yet aware. Let us suppose that i is persuaded by the proof and indeed concedes to its new preference by the speech act: $\mathrm{CONCEDE}_{i,a}(\mathrm{PREFER}(i,\varphi,\psi))$.

All these concessions, together with the general intention formation rule and the fact that agents are correct about their intentions, then lead to $\mathrm{INT}(i,\varphi)$. For intentions with respect to cooperation with other potential team members, the process to persuade the agent to take on $\mathrm{INT}(i,\text{M-INT}_G(\varphi))$ is analogous.

6 Discussion and conclusions

In previous work [9] it was shown how all four stages of CPS result in specific motivational attitudes that can be formally described. In this paper, we have shown for the first two stages, potential recognition and team formation, which rules govern the dialogues by which the motivational attitudes are formed, and how to represent the moves within the dialogues by formalized speech acts.

It is clear that, even though the dialogues are governed by strict rules, the reasoning needed to find an appropriate move is highly complex. This implies that the agents also have to contain complex reasoning mechanisms in order to execute the dialogues. It means that, although the result is much more flexible and refined than using a protocol like Contract Net, the process is also more time consuming. For practical cases one should carefully consider what carries more weight and choose the method of team formation accordingly.
Related work can be found in [16], who also present an agent architecture and a representation for agent communication. In the discussion they note that their own "negotiation" in fact covers a number of Walton and Krabbe's different dialogue types. We find the more fine-grained typology to be very useful when designing agents for teamwork: one can use specific sets of rules governing each type of dialogue as well as the possible embeddings between the different types. Thus desired kinds of communication are allowed and harmful ones prevented, without completely fixing any protocol. Also [16] use multi-context logic whereas we stick to (multi-)modal logic.
The emphasis on pre-planning (here in the stage of potential recognition) and establishing appropriate collective attitudes for teamwork is shared with [14]. Nevertheless, the intentional component in their definition of collective plans is much weaker than our collective intention: Grosz and Kraus' agents involved in a collective plan have individual intentions towards the overall goal and a collective belief about these intentions; intentions with respect to the other agents play a part only at the level of individual sub-actions of the collective plan. We stress, however, that team members' intentions about their colleagues' motivation to achieve the overall goal play an important role in keeping the team on track even if their plan has to be changed radically due to a changing environment (see also [9]).
The first issue for further research is to give a complete set of formal rules for all the types of dialogue and indicate how these are implemented through formal speech acts. This would make it possible to extend the framework to the next stages of cooperative problem solving, namely plan formation and execution.
A second issue is the further investigation of several aspects of the internal reasoning of the agents. One example is the concept of giving a proof as defence of an assertion during the rigorous persuasion. Finally, it should be investigated how actual proofs can be constructed in an efficient way to prove that the end results of a dialogue are formed through the speech acts given the rules of the dialogue.

Acknowledgments

We would like to thank Alexandru Baltag, Magnus Boman, Wiebe van der Hoek, Chris Reed, and the anonymous referees for their useful comments. Barbara Dunin-Kęplicz' research was supported by ESPRIT under the Grant CRIT-2 No. 20288.

References

1. L. Aaqvist. Deontic logic. In: D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, Vol. III, Reidel, Dordrecht, 1984, pp. 605–714.
2. C. Boutilier. Toward a logic for qualitative decision theory. In: *Proceedings KR'94*, 1994, pp. 75-86.

3. C. Castelfranchi and Y.-H. Tan (eds.). *Trust and deception in virtual societies*, Kluwer, Dordrecht, 2000.
4. F. Dignum and R. Conte. Intentional agents and goal formation. In: M. Singh et. al.(eds.), *Intelligent Agents IV (LNAI 1365)*, Springer Verlag, 1998, pp. 231–244.
5. F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Dialogue in team formation: a formal approach. In: F. Dignum and B. Chaib-draa (eds.), *IJCAI Workshop on Agent Communication Languages*, Stockholm, 1999, pp. 39–50.
6. F. Dignum and R. Kuiper. Combining dynamic deontic logic and temporal logic for the specification of deadlines. In: Jr. R. Sprague (ed.), *Proceedings of thirtieth HICSS*, Wailea, Hawaii, 1997.
7. F. Dignum and H. Weigand. Communication and deontic logic. In: R. Wieringa and R. Feenstra (eds.), *Information Systems, Correctness and Reusability*, World Scientific, Singapore, 1995, pp. 242–260.
8. B. Dunin-Kęplicz and R. Verbrugge. Collective commitments. In: *Proc. Second International Conference on Multi-Agent Systems, ICMAS'96*, IEEE Computer Society Press, Kyoto, 1996, pp. 56–63.
9. B. Dunin-Kęplicz and R. Verbrugge. A Reconfiguration algorithm for distributed problem solving. To appear in: *Engineering Simulation: An International Journal of Electrical, Electronic and other Physical Systems*, special issue about distributed systems, 2000.
10. B. Dunin-Kęplicz and R. Verbrugge. Collective motivational attitudes in cooperative problem solving. In: V. Gorodetsky et al. (eds.), *Proceedings of The First International Workshop of Central and Eastern Europe on Multi-agent Systems (CEEMAS'99)*, St. Petersburg, 1999, pp. 22–41.
11. B. Dunin-Kęplicz and R. Verbrugge. The role of dialogue in Cooperative Problem Solving. To appear in: *Proceedings of the Workshop On Intelligent Agents for Computer Supported Cooperative Work*, Barcelona, 2000.
12. R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge (MA), 1995.
13. B.J. Grosz and S. Kraus. Collaborative plans for group action. *Artificial Intelligence* 86 (1996) pp. 269-357.
14. B. van Linder, W. van der Hoek, and J.-J. Ch. Meyer. Formalising abilities and opportunities of agents. *Fundamenta Informaticae* 34 (1998) pp. 53-101.
15. S. Parsons, C. Sierra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3) (1998), pp. 261-292.
16. T. Sandholm and V. Lesser. Issues in automated negotiation and electronic commerce: extending the contract net protocol. In: *Proceedings First International Conference on Multiagent Systems (ICMAS95)*, San Francisco, AAAI Press and MIT Press, 1995, pp. 328-335.
17. J.R. Searle and D. Vanderveken *Foundations of Illocutionary Logic*, Cambridge, Cambridge University Press, 1985.
18. A.S. Rao and M.P. Georgeff. Modeling rational agents within a BDI architecture. In: R. Fikes and E. Sandewall (eds.), *Proceedings of Knowledge Representation and Reasoning (KR&R-91)*, San Mateo, Morgan Kaufmann, 1991, pp.473-484.
19. D. Walton and E. Krabbe. *Commitment in Dialogue*, SUNY Press, Albany, 1995.
20. M. Wooldridge and N.R. Jennings. Cooperative Problem Solving. *Journal of Logic and Computation* 9 (4) (1999), pp. 563-592.