Fundamenta Informaticae 56 (2003) 329–371
IOS Press

Evolution of Collective Commitment during Teamwork

Barbara Dunin-Kęplicz*
Institute of Informatics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
and
Institute of Computer Science, Polish Academy of Sciences
Ordona 21, 01-237 Warsaw, Poland
e-mail: keplicz@mimuw.edu.pl

Rineke Verbrugge
Institute of Artificial Intelligence, University of Groningen
Grote Kruisstraat 2/1, 9712 TS Groningen
The Netherlands
e-mail: rineke@ai.rug.nl

Abstract. In this paper we aim to describe dynamic aspects of social and collective attitudes in teams of agents involved in Cooperative Problem Solving (CPS). Particular attention is given to the strongest motivational attitude, collective commitment, and its evolution during team action. First, building on our previous work, a logical framework is sketched in which a number of relevant social and collective attitudes is formalized, leading to the plan-based definition of collective commitments. Moreover, a dynamic logic component is added to this framework in order to capture the effects of the complex actions that are involved in the consecutive stages of CPS, namely potential recognition, team formation, plan formation and team action.

During team action, the collective commitment leads to the execution of agent-specific actions. A dynamic and unpredictable environment may, however, cause the failure of some of these actions, or present the agents with new opportunities. The abstract reconfiguration algorithm, presented in a previous paper, is designed to handle the re-planning needed in such situations in an efficient way. In this paper, the dynamic logic component of the logical framework addresses issues pertaining to adjustments in collective commitment during the reconfiguration process.

* Address for correspondence: Institute of Informatics, Warsaw University, Banacha 2, 02-097 Warsaw, Poland

1. Introduction

To set the stage, let us introduce some important notions from the field of multiagent systems, and provide some background about the complete story of which the present paper is a part (cf. [21]). In *multiagent systems* (MAS) one of the central issues is the study of how groups work, and how the technology enhancing group interaction can be implemented. From the distributed Artificial Intelligence perspective, multiagent systems are computational systems in which a collection of loosely-coupled autonomous agents interact in order to solve a given problem. As this problem is usually beyond the agents' individual capabilities, agents exploit their ability to *communicate*, *cooperate*, *coordinate*, and *negotiate* with one another. Apparently, the type of social interactions involved depends on circumstances and may vary from altruistic cooperation through to open conflict. A paradigmatic example of joint activity is *cooperative problem solving* (CPS) in which a group of autonomous agents choose to work together, both in advancement of their own goals as well as for the good of the system as a whole.

Some MAS are referred to as *intentional systems*. In such systems, in order to give a representation of the mental states and cognitive processes involved in a multiagent system, agents are represented as maintaining an intentional stance towards their environment. Such systems realize the *practical reasoning* paradigm ([4]) – the process of deciding, moment by moment, which action to perform in the furtherance of our goals. The best known and most influential are *belief-desire-intention systems*. BDI-agents are characterized by a "mental state" described in terms of *beliefs*, corresponding to the information the agent has about the environment; *desires*, representing options available to the agent, i.e. different states of affairs that the agent may choose to commit to; and *intentions* representing the chosen options. Ultimately, in our approach, intentions are viewed as an inspiration for a goal-directed activity, reflected in commitments. While beliefs are viewed as the agent's *informational* attitudes, desires or goals, intentions, and commitments refer to its *motivational* attitudes.

One of the vital aspects of BDI systems in which the dynamics is expressed is *teamwork*. In many recent BDI systems teamwork is modeled explicitly. The explicit model helps the team to monitor its performance and especially to re-plan based on the present situation. The dynamic and often unpredictable environment in which agents are acting, poses the problem that team members may fail to bring their tasks to a good end or new opportunities may appear. This leads to the so-called *reconfiguration problem*: when maintaining a collective intention during plan execution, it is crucial that agents re-plan properly and efficiently when the situation changes.

This reconfiguration problem has only recently come to be discussed (see [47], [17]). A generic solution of this problem in BDI systems is presented by us in [19], where the main contribution is the *reconfiguration algorithm* together with the discussion of an example application. We base our solution on the four-stage model of [53], containing the consecutive stages of *potential recognition*, *team formation*, *plan formation* and *team action*. When defining the levels we abstract from particular methods and algorithms meant to realize level-associated goals, but instead formulate their final results and associate them with appropriate individual, social, and collective motivational attitudes. Ultimately, the reconfiguration algorithm, showing the phases of construction, maintenance, and realization of collective commitment, is formulated in terms of these levels and their (complex) interplay.

The reconfiguration algorithm is a departure point to describe the dynamics of social and collective attitudes in a team of agents involved in CPS. In a formal specification of these notions in BDI-systems, different kinds of modal logics are exploited. Dynamic, temporal and epistemic logics are extensively used to describe the single agent case. Inevitably, social and collective aspects of CPS should be inves-

tigated and formalized, again, in a combination of different kinds of modal logics. In our approach, all individual motivational attitudes are viewed as primitive notions.

Starting from individual intentions, we first defined the notion of a *collective intention* for a team [21]. Together with individual and collective knowledge and belief, a collective intention constitutes a basis for preparing a plan (or a set of plans). Planning may be done in many different ways. Based on the resulting plan, we characterized the strongest motivational attitude, *collective commitment* of a team. We assume that bilateral aspects of a plan — obligations from one agent to another — are reflected in *social commitments*. Thus, collective commitment is defined on the basis of collective intention and social commitments. In other words, our approach to collective commitment is plan-based: the ongoing collective intention is split up into sub-actions, according to a given social plan. Next, the action allocation is reflected in social commitments between pairs of agents. In this paper, one definition of collective commitment, namely strong collective commitment, is given, based on a social plan. See [22] for a number of different kinds of collective commitments and a method for the system developer to calibrate the strength of the collective commitment to different environments, organizational structures and purposes.

Using this framework we aim to describe in this paper the maintenance of collective commitment during reconfiguration in the action execution stage.

The action execution stage, or team action, is the final stage of a BDI system life cycle. In case some action performance fails, the realization of the collective commitment of the team is threatened. It means that some effects of the previous stages of teamwork, which were sometimes realized in rather complex and expensive ways, may be wasted. However, in some cases a rather small correction of the overall plan, and of the collective commitment based on it, suffices to save the situation. For example, sometimes it is enough to reallocate some actions to different team members. If this cannot be done, a new plan may be established by means of a new task division slightly changing the existing one, etc. Often the necessary changes are insignificant, preferably saving most of the previously obtained results. The reconfiguration algorithm reflects a rigorous methodological approach to these changes: this way a sort of *evolution* of collective commitment during reconfiguration is shown in a dynamically changing environment. In this paper we will characterize the properties of this process using dynamic logic notation, which allows to precisely describe the results of complex actions involved in reconfiguration. The new contribution of this paper as compared to [19] is that the process of motivational and informational attitude change during reconfiguration is made transparent. The dynamic logic description will provide a basis for implementation of the system, as well as for formal verification methods.

Thus, we will adopt a computer science point of view, taking the perspective of a system developer, rather than the one of an agent. The properties describing system behavior are expressed in the formalism of dynamic logic and thus provide a kind of specification. This specification most of the time describes properties of actions that introduce or delete agents' attitudes, as well as properties of complex social actions that establish relevant properties of different stages of CPS. However, we will not come into details about how teamwork is to be realized. These procedures are rather complex and form a research subject by themselves. This application dependent problem is discussed elsewhere in more depth [19, 13, 12]. Instead we will focus on generic properties ensuring correct behaviour of the system as a whole. This enables a system designer to construct a program from an existing specification, even though this specification is rather complex.

All notions in this paper are formalized in a multi-modal logical framework with a well-defined Kripke semantics. Thus, a Computational Logic framework for specifying multiagent systems (MAS) involved in CPS is provided. For the collective intention part of this framework, soundness and complete-

ness have been proved [21]. For the full system, where the informational and motivational modalities are combined with dynamic ones, completeness remains to be investigated. The full system is known to be EXPTIME-hard, because it contains the logic of collective beliefs as a subsystem. Thus, in general it is not feasible to give automated proofs of desired properties, at least there is no single algorithm that performs well on all inputs. As with other modal logics, the better option is to develop a variety of different algorithms and heuristics, each performing well on a limited class of inputs. For example, it is known that restricting the number of propositional atoms to be used or the depth of modal nesting may reduce the complexity (cf. [29, 33, 27, 50]). Also, when considering specific applications it is possible to reduce some of the infinitary character of collective beliefs and intentions to more manageable proportions (cf. [23, Ch. 11]).

Usually, BDI-logics are based on a linear or branching temporal logic [42, 10], sometimes with some dynamic additions. We, in contrast, restrict ourselves to a dynamic, action-oriented formal framework. Apparently, a full specification of the system includes complex temporal aspects, such as persistence of certain properties over time until some given deadline. However, assuming a developer's perspective, we will not introduce these temporal elements into the logical framework. It is known that the combination of dynamic and temporal logic is extremely complex, especially in the presence of other modal operators (as is the case here). Therefore, instead of making the logical system even more intractable from the formal point of view, and much harder to understand, we decided to express temporal aspects in a procedural way. The method to implement them is left to the system developer, as it corresponds straightforwardly to the way the abstract reconfiguration algorithm is implemented for a particular application.

The paper is organized in the following way. In section 2, the logical language and semantics are introduced. Section 3 is devoted to Kripke models, dynamic logic for actions and social plans, and individual and collective beliefs. Then, in section 4, individual and social motivational attitudes are characterized, while section 5 investigates the collective motivational attitudes that come to the fore during teamwork, namely collective intentions and collective commitments. In section 6, the effects of individual agents' dropping their intentions and commitments are investigated. Section 7 gives a short overview of the four levels of CPS. The central section 8 presents in a multi-modal language how collective commitments evolve during reconfiguration. Finally, section 9 focuses on discussion and options for further research.

2. The language: formulas, individual actions and social plan expressions

We propose the use of multi-modal logics to formalize agents' informational and motivational attitudes as well as actions they perform and their effects. In CPS, both motivational and informational attitudes are considered on the following three levels: *individual*, *social* and *collective*.

2.1. The logical language

Individual actions and formulas are defined inductively, both with respect to a fixed finite set of agents. The basis of the induction is given in the following definition.

Definition 2.1. (Language)
The language is based on the following three sets:

- a denumerable set $\mathcal{P}$ of *propositional symbols*;

- a finite set $\mathcal{A}$ of *agents*, denoted by numerals $1, 2, \ldots, n$;
- a finite set $\mathcal{A}t$ of *atomic actions*, denoted by a or b.

In our framework most modalities relating agents' motivational attitudes appear in two forms: with respect to *propositions*, or 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 [14, 15, 16], etc.

A proposition reflects a particular state of affairs. The transition from a proposition that agents intend to bring about to an action realizing this is achieved by means-end analysis, which will be discussed in section 5.2.

The set of formulas (see definition 2.5) is defined in a double induction, together with the class of individual actions $\mathcal{A}c$, the class of complex social actions $\mathcal{C}o$ and the class of social plan expressions $\mathcal{S}p$ (see definitions 2.2, 2.3 and 2.4). The class $\mathcal{A}c$ is meant to refer to agents' individual actions; they are usually represented without naming the agents, except when other agents are involved such as in [**AC7**] below. The individual actions may be combined into group actions by the social plan expressions defined below.

Below, we give a particular choice of operators to be used when defining individual actions and social plan expressions. However, as actions and social plans are not the main subjects of this paper, in the sequel we hardly come into detail as to how particular individual actions and social plans are built up. Thus, another definition (e.g. without the iteration operation or without non-deterministic choice) may be used if more appropriate in a particular context.

Definition 2.2. (Individual actions)
The class $\mathcal{A}c$ of individual actions is defined inductively as follows:

AC1 each atomic action $a \in \mathcal{A}t$ is an individual action;

AC2 if $\varphi \in \mathcal{L}$, then $\texttt{confirm}\,\varphi$ is an individual action; (confirmation)

AC3 if $\alpha_1, \alpha_2 \in \mathcal{A}c$, then $\alpha_1; \alpha_2$ is an individual action; (sequential composition)

AC4 if $\alpha_1, \alpha_2 \in \mathcal{A}c$, then $\alpha_1 \cup \alpha_2$ is an individual action; (non-deterministic choice)

AC5 if $\alpha \in \mathcal{A}c$, then α^* is an individual action; (iteration)

AC6 if $\varphi \in \mathcal{L}$, then $\texttt{stit}(\varphi)$ is an individual action;

AC7 if $\varphi \in \mathcal{L}$, α is an individual action, $i, j \in \mathcal{A}$ and $G \subseteq \mathcal{A}$, then the following are individual actions:
$\texttt{announce}_G(i, \varphi)$, $\texttt{communicate}(i, j, \varphi)$, $\texttt{unintend}(i, \varphi)$,
$\texttt{commit}(i, j, \alpha)$, $\texttt{uncommit}(i, j, \alpha)$;

Here, in addition to the standard dynamic operators of [**AC1**] to [**AC5**], the operator `stit` of [**AC6**] stands for "sees to it that" or "brings it about that", and has been extensively treated in [2, 45]. The communicative actions $\texttt{announce}_G(i, \varphi)$ and $\texttt{communicate}(i, j, \varphi)$ and their role in creating belief changes in individuals and groups are treated in subsection 3.2.1. Finally, actions $\texttt{unintend}(i, \varphi)$, $\texttt{commit}(i, j, \alpha)$ and $\texttt{uncommit}(i, j, \alpha)$ refer to agents taking on and dropping motivational attitudes, as described in subsection 6.1. We do not add axioms for these special, application-dependent individual actions, because they do not obey any one generic axiom system.

The complex social actions defined below refer to the four stages of CPS, consecutively: potential recognition (`potential-recognition`), team formation (`team-formation`), plan generation (`plan-`

generation) and team action. The plan generation level in turn is divided into three consecutive substages, namely task division (task-division), means-end analysis (means-end-analysis), and action allocation (action-allocation). In this idealization, at the stages of CPS agents individually and collectively work on the creation, maintenance and realization of motivational attitudes on the individual, social and collective level. This rather complex process is described in detail in [19] and more concisely in section 7. Let us stress that these level-oriented complex social actions are deeply application-dependent, and by themselves may be considered as independent research project. For this reason, as well as for its complexity, it is impossible to fully describe them here. What is assumed here is that all stage-dependent actions are realized by the group as a whole.

Definition 2.3. (Complex social action)
The class $\mathcal{C}o$ of complex social actions is introduced as follows:

CO1 if φ is a formula, α is an individual action, $G \subseteq \mathcal{A}$, $\mathcal{G}$ a finite sequence of subsets of $\mathcal{A}$, σ a finite sequence of formulas, τ a finite sequence of individual actions, and P a social plan expression, then $\texttt{potential-recognition}(\varphi, \mathcal{G})$, $\texttt{team-formation}(\mathcal{G}, G)$, $\texttt{plan-generation}(\varphi, G, P)$, $\texttt{task-division}(\varphi, \sigma)$, means-end-analysis$(\sigma, \tau)$, $\texttt{action-allocation}(\tau, P)$, $\texttt{system-success}(\varphi)$, and $\texttt{system-failure}(\varphi)$, are complex social actions.

CO2 If β_1 and β_2 are complex social actions, then so is $\beta_1 \ ; \ \beta_2$.

Definition 2.4. (Social plan expressions)
The class $\mathcal{S}p$ of social plan expressions is defined inductively as follows:

SP1 If $\alpha \in \mathcal{A}c$ and $i \in \mathcal{A}$, then $\langle \alpha, i \rangle$ is a well-formed social plan expression;

SP2 If φ is a formula and $G \subseteq \mathcal{A}$, then $\texttt{stit}_G(\varphi)$ and $\texttt{confirm}(\varphi)$ are social plan expressions;

SP3 If α and β are social plan expressions, then $\langle \alpha;\beta \rangle$ (sequential composition) and $\langle \alpha \parallel \beta \rangle$ (parallelism) are social plan expressions.

A concrete example of a social plan expression will be given in subsection 5.2.1. The social plan $\texttt{confirm}(\varphi)$ (to test whether φ holds at the given world) is given here without group subscript, because the group does not influence the semantics, see section 3.1. It will be clear from the context whether confirm is used as an individual action or as a social plan expression.

The modalities appearing in the definition of formulas below are all explained later in the paper. See subsection 3.1 about dynamic modalities, subsection 3.2 about epistemic modalities, and sections 4 and 5 about individual, social and collective motivational modalities.

Definition 2.5. (Formulas)
We inductively define a set of formulas L as follows.

F1 each atomic proposition $p \in \mathcal{P}$ is a formula;

F2 if φ and ψ are formulas, then so are $\neg\varphi$ and $\varphi \wedge \psi$;

F3 if φ is a formula, $\alpha \in \mathcal{A}c$ is an individual action, $\beta \in \mathcal{C}o$ is a complex social action, $i, j \in \mathcal{A}$, $G \subseteq \mathcal{A}$, σ a finite sequence of formulas, τ a finite sequence of individual actions, and $P \in \mathcal{S}p$ a social plan expression, then the following are formulas:

epistemic modalities $\mathrm{BEL}(i,\varphi)$, $\text{E-BEL}_G(\varphi)$, $\text{C-BEL}_G(\varphi)$;

motivational modalities $\mathrm{GOAL}(i,\varphi)$, $\mathrm{GOAL}(i,\alpha)$, $\mathrm{INT}(i,\varphi)$, $\mathrm{INT}(i,\alpha)$, $\mathrm{COMM}(i,j,\varphi)$, $\mathrm{COMM}(i,j,\alpha)$, $\text{E-INT}_G(\varphi)$, $\text{E-INT}_G(\alpha)$, $\text{M-INT}_G(\varphi)$, $\text{M-INT}_G(\alpha)$, $\text{C-INT}_G(\varphi)$, $\text{C-INT}_G(\alpha)$, $\text{S-COMM}_{G,P}(\varphi)$, $\text{S-COMM}_{G,P}(\alpha)$;

temporal action modalities $done\text{-}ac(i,\alpha)$, $succ\text{-}ac(i,\alpha)$, $failed\text{-}ac(i,\alpha)$; $done\text{-}sp(G,P)$,$succ\text{-}sp(G,P)$, $failed\text{-}sp(G,P)$; $done\text{-}co(G,\beta)$, $succ\text{-}co(G,\beta)$, $failed\text{-}co(G,\beta)$, $do\text{-}ac(i,\alpha)$,$do\text{-}sp(G,P)$, $do\text{-}co(G,\beta)$;

abilities and opportunities $able(i,\alpha)$, $opp(i,\alpha)$;

dynamic modalities $[do(i,\alpha)]\varphi$, $[\beta]\varphi$, $[P]\varphi$;

level results $division(\varphi,\sigma)$, $means(\sigma,\tau)$, $allocation(\tau,P)$, $constitute(\varphi,P)$;

The level results in the above definition refer to the results of the three sub-stages of plan generation, namely task division, means-end analysis and action allocation. Plan generation is the third of the four stages of cooperative problem solving; they are all described in section 7. The predicate $constitute(\varphi,P)$ (for "P constitutes a correctly constructed social plan for realizing state of affairs φ") is defined from the three level results, see subsection 7.3.1.
The constructs $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined in the usual way.

3. Kripke models

Each Kripke model for the language defined in the previous section consists of a set of worlds, a set of accessibility relations between worlds, and a valuation of the propositional atoms, as follows. The definition also includes semantics for derived operators corresponding to abilities, opportunities, and performance of (individual or social) actions.

Definition 3.1. (Kripke model)
A Kripke model is a tuple
$\mathcal{M} = (W, \{B_i : i \in \mathcal{A}\}, \{G_i : i \in \mathcal{A}\}, \{I_i : i \in \mathcal{A}\}, \{R_{i,\alpha} : i \in \mathcal{A}, \alpha \in \mathcal{A}c\}, \{R_\beta : \beta \in \mathcal{C}o\}, \{R_P : P \in \mathcal{S}p\}, Val, abl, op, perfac, perfsp, perfco)$, such that

1. W is a set of possible worlds, or states;
2. For all $i \in \mathcal{A}$, it holds that $B_i, G_i, I_i \subseteq W \times W$. They stand for the accessibility relations for each agent w.r.t. beliefs, goals, and intentions, respectively[1].
3. For all $i \in \mathcal{A}$, $\alpha \in \mathcal{A}c$, $\beta \in \mathcal{C}o$ and $P \in \mathcal{S}p$, it holds that $R_{i,\alpha}, R_\beta, R_P \subseteq W \times W$. They stand for the dynamic accessibility relations[2].

[1]For example, $(w_1, w_2) \in B_i$ means that w_2 is an epistemic alternative for agent i in state w_1.
[2]For example, $(w_1, w_2) \in R_{i,\alpha}$ means that w_2 is a possible resulting state from w_1 by agent i executing action α.

4. $Val : \mathcal{P} \times W \rightarrow \{0,1\}$ is the function that assigns the truth values to propositional formulas in states.

5. $abl : \mathcal{A} \times \mathcal{A}c \rightarrow \{0,1\}$ is the ability function such that $abl(i,\alpha) = 1$ indicates that agent i is able to realize the action α. $\mathcal{M}, v \models able(i,\alpha) \Leftrightarrow abl(i,\alpha) = 1$.

6. $op : \mathcal{A} \times \mathcal{A}c \rightarrow (W \rightarrow \{0,1\})$ is the opportunity function such that $op(i,\alpha)(w) = 1$ indicates that agent i has the opportunity to realize action α in world w. $\mathcal{M}, v \models opp(i,\alpha) \Leftrightarrow op(i,\alpha)(v) = 1$;

7. $perfac : \mathcal{A} \times \mathcal{A}c \rightarrow (W \rightarrow \{0,1,2\})$ is the individual action performance function such that $perfac(i,\alpha)(w)$ indicates the result in world w of the performance of individual action α by agent i in world w; (here, 0 stands for failure, 1 for success, and 2 stands for "undefined", e.g. if w is not the endpoint of an $R_{(i,\alpha)}$ accessibily relation).

   - $\mathcal{M}, v \models succ\text{-}ac(i,\alpha) \Leftrightarrow perfac(i,\alpha)(v) = 1$;
   - $\mathcal{M}, v \models failed\text{-}ac(i,\alpha) \Leftrightarrow perfac(i,\alpha)(v) = 0$;
   - $\mathcal{M}, v \models done\text{-}ac(i,\alpha) \Leftrightarrow perfac(i,\alpha)(v) \in \{0,1\}$.

8. $perfco : 2^{\mathcal{A}} \times \mathcal{C}o \rightarrow (W \rightarrow \{0,1,2\})$ is the complex social action performance function such that $perfco(j,\beta)(w)$ indicates the result in world w of the performance of complex social action β by a group of agents j.

   - $\mathcal{M}, v \models succ\text{-}co(j,\beta) \Leftrightarrow perfco(j,\beta)(v) = 1$;
   - $\mathcal{M}, v \models failed\text{-}co(j,\beta) \Leftrightarrow perfco(j,\beta)(v) = 0$;
   - $\mathcal{M}, v \models done\text{-}co(j,\beta) \Leftrightarrow perfco(j,\beta)(v) \in \{0,1\}$.

9. $perfsp : 2^{\mathcal{A}} \times \mathcal{S}p \rightarrow (W \rightarrow \{0,1,2\})$ is the social plan performance function such that $perfsp(j, P)(w)$ indicates the result in world w of the performance of social plan P by a group of agents j.

   - $\mathcal{M}, v \models succ\text{-}sp(j,P) \Leftrightarrow perfasp(j,P)(v) = 1$;
   - $\mathcal{M}, v \models failed\text{-}sp(j,P) \Leftrightarrow perfsp(j,P)(v) = 0$;
   - $\mathcal{M}, v \models done\text{-}sp(j,P) \Leftrightarrow perfsp(j,P)(v) \in \{0,1\}$.

10. $nextac : \mathcal{A} \times \mathcal{A}c \rightarrow (W \rightarrow \{0,1\})$ is the next moment individual action function such that $nextac(i,\alpha)(w)$ indicates that in world w agent i will next perform action α. $\mathcal{M}, v \models do\text{-}ac(i, \alpha) \Leftrightarrow nextac(i,\alpha)(v) = 1$.

11. $nextco : 2^{\mathcal{A}} \times \mathcal{C}o \rightarrow (W \rightarrow \{0,1\})$ is the next moment complex social action performance function such that $nextco(j,\beta)(w)$ indicates that in world w the group of agents j will next start performing the complex social action β. $\mathcal{M}, v \models do\text{-}co(j,\beta) \Leftrightarrow nextco(j,\beta)(v) = 1$;

12. $nextsp : 2^{\mathcal{A}} \times \mathcal{S}p \rightarrow (W \rightarrow \{0,1\})$ is the next moment social plan performance function such that $nextsp(j,P)(w)$ indicates that in world w the group of agents j will next start performing social plan P. $\mathcal{M}, v \models do\text{-}sp(j,P) \Leftrightarrow nextsp(j,P)(v) = 1$.

The aspect of ability (cf. the abl-function) 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 (cf. the op-function) takes into account the possibilities of task performance in the present situation, involving resources and possibly other properties. Both abilities and opportunities are modeled in the above definition in a rather static way. It is possible to make a more refined definition, using a language that includes dynamic and/or temporal operators (see e.g. [5, 15, 38]). We have chosen not to do so here, because these concepts are not the main focus of this paper. We do assume that the functions are in accord with the construction of complex individual actions, for example, if an agent is able to realize $a; b$, then it is able to realize a.

Similarly, we have modeled action performance for individual actions, social plans and complex social actions by functions (*perfac*, *perfsp* and *perfco*), modeling whether a certain action has just been performed, and if so, whether it was successful. Finally the functions *nextac*, *nextsp* and *nextco* model whether a certain action will be executed next. Again, these functions are assumed to agree with the construction of complex actions, for example, if $perfac(i, a; b) = 1$, then $perfac(i, b) = 1$.

We use three-valued performance functions for actions, complex social actions and social plan expressions, because at many worlds it may be that the relevant action has not been performed at all. Of course one could also use partial functions here (where our value 2 is replaced by "undefined").

The truth conditions pertaining to the propositional part of the language $\mathcal{L}$ are the standard ones used in modal logics.

The derived operators above correspond in a natural way to the results of the ability, opportunity, performance and next execution functions. For example, $\mathcal{M}, v \models done\text{-}sp(j, P)$ is meant to be true if team j just executed the social plan P, as modelled by the performance function giving a value other than 2 (undefined), i.e. $perfsp(j, P)(v) \in \{0, 1\}$.

In the remainder of the paper we will mostly abbreviate all the above forms of success, failure and execution (past and future) for actions, complex actions, and social plans to simply $succ$, $failed$, $done$ and do.

The truth conditions for formulas with dynamic operators as main modality are given in subsection 3.1; for those with epistemic main operators, the truth definitions are given in subsection 3.2; finally, for those with motivational modalities as main operators, the definitions follow in section 4.

3.1. Dynamic logic for actions and social plans

In the semantics, the relations $R_{i,a}$ for atomic actions a are given. The other accessibility relations $R_{i,\alpha}$ for actions are built up from these as follows in the usual way[31]:

Definition 3.2. (Dynamic accessibility relations for actions)

- $(v, w) \in R_{i,\mathtt{confirm}(\varphi)} \Leftrightarrow (v = w$ and $\mathcal{M}, v \models \varphi)$;
- $(v, w) \in R_{i,\alpha_1;\alpha_2} \Leftrightarrow \exists u \in W[(v, u) \in R_{i,\alpha_1}$ and $(u, w) \in R_{i,\alpha_2}]$;
- $(v, w) \in R_{i,\alpha_1 \cup \alpha_2} \Leftrightarrow [(v, w) \in R_{i,\alpha_1}$ or $(v, w) \in R_{i,\alpha_2}]$;
- R_{i,α^*} is the reflexive transitive closure of $R_{i,\alpha}$.

In a similar way, the accessibility relations for social plan expressions and complex social actions are built up from those of individual actions in an appropriate way[31, 40]. We do not give the complete definition, but for example, we have:

- If $\alpha \in \mathcal{A}c$ and $i \in \mathcal{A}$, then $(v, w) \in R_{\langle \alpha, i \rangle} \Leftrightarrow (v, w) \in R_{i,\alpha}$;
- $(v, w) \in R_{\mathtt{confirm}(\varphi)} \Leftrightarrow (v = w$ and $\mathcal{M}, v \models \varphi)$;

Now we can define the valuations of complex formulas containing dynamic operators as main operator.

Definition 3.3. (Valuation for dynamic operators)
Let φ be a formula, $i \in \mathcal{A}$, $\alpha \in \mathcal{A}c$, $\beta \in \mathcal{C}o$, and $P \in \mathcal{S}p$.

actions $\mathcal{M}, v \models [do(i, \alpha)]\varphi \Leftrightarrow$ for all w with $(v, w) \in R_{i,\alpha}, \mathcal{M}, w \models \varphi$;

social plan expressions $\mathcal{M}, v \models [P]\varphi \Leftrightarrow$ for all w with $(v, w) \in R_P, \mathcal{M}, w \models \varphi$;

complex social actions $\mathcal{M}, v \models [\beta]\varphi \Leftrightarrow$ for all w with $(v, w) \in R_\beta, \mathcal{M}, w \models \varphi$.

For the dynamic logic of actions, we adapt the axiomatization PDL of propositional dynamic logic, as found in [26], see appendix I. The axiom system PDL is sound and complete with respect to Kripke models with only the dynamic accessibility relations $R_{i,\alpha}$ as defined above. Its decision problem is exponential time complete, as proved by [24].

One needs to add axioms for complex social actions and social plan expressions in an appropriate way, for example, for all $\mathcal{M}, w$:

$$\mathcal{M}, w \models [\mathtt{confirm}(\psi)]\chi \leftrightarrow (\psi \rightarrow \chi).$$

As this is not the main subject of this paper, and as the axiom systems depend on the domain in question, we will not include a full system here. However, for the $\|$-operator, one may use the appropriate axioms for concurrent dynamic logic as found in [31].

3.2. Beliefs

To represent beliefs, we take $\mathrm{BEL}(i, \varphi)$ to have as intended meaning "agent i believes proposition φ". In the semantics, BEL is defined as follows:

$$\mathcal{M}, w \models \mathrm{BEL}(i, \varphi) \text{ iff } t \models \varphi \text{ for all } t \text{ such that } (w, t) \in B_i.$$

One can define modal operators for group beliefs. The formula $\text{E-BEL}_G(\varphi)$ is meant to stand for "every agent in group G believes φ". Thus, $\mathcal{M}, w \models \text{E-BEL}_G(\varphi)$ iff for all $i \in G$, $\mathcal{M}, w \models \mathrm{BEL}(i, \varphi)$.

A traditional way of lifting single-agent concepts to multi-agent ones is through the use of *collective belief* $\text{C-BEL}_G(\varphi)$. This rather strong operator is similar to the more usual one of 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. Thus $\mathcal{M}, w \models \text{C-BEL}_G(\varphi)$ iff φ holds in all worlds reachable in one or more steps by B_i arrows for $i \in G$.

A standard system axiomatizing these belief operators for n agents 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 [23]. See appendix I for the axioms and rules. In the sequel, we will use the following standard properties of C-BEL$_G$ (see for example [23, exercise 3.11]).

Lemma 3.1.

- C-BEL$_G(\varphi \wedge \psi) \leftrightarrow$ C-BEL$_G(\varphi) \wedge$ C-BEL$_G(\psi)$
- C-BEL$_G(\varphi) \rightarrow$ C-BEL$_G$(C-BEL$_G(\varphi)$)

3.2.1. Belief changes through communication

Some of the ways in which individual beliefs can be generated are updating, revision, and contraction [49, 25]. The establishment of collective beliefs among a group is more problematic. In [30, 46] it is shown that bilateral sending of messages does not suffice to determine collective belief if communication channels may be faulty, or even if there is uncertainty whether message delivery may have been delayed. A good reference to the problems concerning collective belief and to their possible solutions is [23, Chapter 11]. In any case, it is generally agreed that collective belief is a good *abstraction tool* to study teamwork.

We assume that in our groups bilateral communication as well as a more general type of communication, e.g. by a kind of global announcement, can be achieved. Problems related to message delivery are disregarded in the rest of this paper. Given an agent i and an agent j, the action $\texttt{communicate}(i, j, \psi)$ stands for "agent i communicates to agent j that ψ holds". Next, given a group G and an agent $i \in G$, the action $\texttt{announce}_G(i, \psi)$ stands for "agent i announces to group G that ψ holds".

An important aspect involved in the process of communication is *trustworthiness*, addressing the question "whether agent j (the receiver) trusts agent i (the sender) with respect to proposition ψ". As trust is a rather complex concept, it may be defined in many ways, from different perspectives (see [8, 7] for some current work in this area).

We do not aim to define trust and trustworthiness in this paper, however some form of trust has to be adopted in CPS. It would be too much to assume that agents believe everything other agents communicate to them. For some propositions though, it is vital for the success of teamwork that agents who receive them adopt them as their own beliefs. For such propositions ψ, the following holds:

$$succ(\texttt{communicate}(i, j, \psi)) \rightarrow \text{BEL}(j, \psi)$$

$$succ(\texttt{announce}_G(i, \psi)) \rightarrow \text{C-BEL}_G(\psi).$$

In this paper, we assume that there is trust between agents with respect to all formulas communicated or announced to them that appear during the different stages of CPS.

4. Individual and social motivational attitudes

Practical reasoning involves two important processes: deciding *what* goals need to be achieved, and then *how* to achieve them. The former process is known as *deliberation*, the latter as *means-end reasoning*. In

$\mathrm{COMM}(i, j, \varphi)$	agent i commits to agent j to make φ true
$\mathrm{GOAL}(i, \varphi)$	agent i has as a goal that φ be true
$\mathrm{INT}(i, \varphi)$	agent i has the intention to make φ true
$\text{E-INT}_G(\varphi)$	every agent in group G has the individual intention to make φ true
$\text{M-INT}_G(\varphi)$	group G has the mutual intention to make φ true
$\text{C-INT}_G(\varphi)$	group G has the collective intention to make φ true
$\text{S-COMM}_{G,P}(\varphi)$	group G has strong commitment to make φ true by plan P

Table 1. Formulas and their intended meaning

the sequel we will discuss both these processes in the context of informational and motivational attitudes of the agents involved.

The key concept in the theory of practical reasoning is the one of *intention*, studied in-depth in [4]. Intentions form a rather special consistent subset of an agent's goals, that the agent wants to focus on for the time being. Speaking with Cohen and Levesque, intention consists of choice together with commitment, in a non-technical sense [10]. Notice that in our definitions, these two ingredients of intention are separated: intention is viewed as chosen goal, providing inspiration for a more concrete social (pairwise) commitment in the individual case, and a plan-based collective commitment in the collective case.

Thus intentions create a screen of admissibility for the agent's further, possibly long-term, deliberation. However, from time to time an agent's intentions should be reconsidered, for example because they will never be achieved, they are achieved already, or reasons originally supporting them hold no longer. This leads to the problem of balancing *pro-active*, (i.e. goal-directed) and *reactive* (i.e. event-driven) behavior.

In the presented approach we try to maintain this balance very carefully on the three different levels of teamwork: individual, social and collective. On the individual and social level the problem of persistence of both intentions and then commitments is first expressed in agent's *intention* and *commitment strategies*, addressing the question: *when and how can an agent responsibly drop its intention or commitment ?* The answer to this question is discussed in section 6, and more extensively in [17, 18]. The collective level is apparently much more complex. In our framework an agent's pro-activeness and re-activeness are implicitly or explicitly involved on consecutive stages of the reconfiguration algorithm [19]. The formal specification of these situations is given in section 8.

Our framework to describe motivational attitudes and related aspects is minimal in the sense that we aim to deal with concise necessary and sufficient conditions describing solely the *core aspects* of teamwork. Additional aspects appearing on the stage in specific cases may be addressed by refining the system and adding new axioms.

Table 1 gives a number of modal formulas appearing in this paper, together with their intended meanings. The symbol φ denotes a proposition, but all these formulas also appear with respect to an action α. Even though it may seem from the table as if the formulas have only an informal meaning (perhaps derived from so-called folk psychology), this is actually not the case. In fact, the individual motivational attitudes are primitive but are governed by axiom systems and corresponding semantics, while the social and collective motivational attitudes are defined by axioms in terms of the individual ones.

4.1. Individual goals and intentions

For the motivational operators GOAL and INT the axioms include the system K, which we adapt for n agents to K_n. In a BDI system, an agent's activity starts from goals. As the agent may have many different objectives, its goals need not be consistent with each other. Then, the agent chooses a limited number of its goals to be intentions. It is not the main focus of this paper to discuss how intentions are formed from a set of goals (but see [13, 11]).

But goals and their relation with intentions form an important part of BDI-theory, so goals are first-class citizens in our system. In any case, we assume that intentions are chosen in such a way that consistency is preserved. Thus for intentions we assume, as Rao and Georgeff do, that they should be consistent, see axiom $\mathbf{A6}_I$ in appendix I.

Rao and Georgeff also add an analogous axiom for the consistency of goals. However, it was argued above that an agent's goals are not necessarily consistent with each other. Thus, we adopt the basic system K_n for goals. Nevertheless, in the presented approach other choices may be adopted without consequences for the rest of the definitions in this paper. It is not hard to prove soundness and completeness of the basic axiom systems for goals and intentions with respect to suitable classes of models by a tableau method, and also give decidability results using a small model theorem.

As to interdependencies between the individual attitudes, we add five axioms (see appendix I), formalizing the properties that agents have positive and negative introspection about their individual motivational attitudes, as well as the property that every intention corresponds to a goal. The interdependence axioms correspond to structural conditions on Kripke models.

All axioms about individual motivational attitudes in appendix I are formulated with respect to formulas. However, companion axioms with respect to individual actions are also included in the system.

4.2. Social commitments

As [6] showed, it is important to distinguish between individual intentions, bilateral commitments, and collective motivational attitudes. A social commitment between two agents is not as strong as a collective commitment among them (see subsection 5.2), but stronger than an individual intention of one agent. If an agent commits to a second agent to do something, then the first agent should have the *intention* to do that. Moreover, the first agent commits to the second one only if the second one is *interested* in the first one fulfilling its intention. These two conditions are inspired by [6], but we find that for a social commitment to arise, a third condition is necessary, namely that the agents are aware about the situation, i.e. about their individual attitudes (cf. also [44] for an early discussion about the properties of promises). Such awareness, expressed in terms of collective belief, is generally achieved by communication. Here follows the defining axiom for social commitments with respect to propositions:

SC1

$$\mathrm{COMM}(i,j,\varphi) \leftrightarrow \mathrm{INT}(i,\varphi) \wedge \mathrm{GOAL}(j,\mathtt{stit}(i,\varphi)) \wedge$$

$$\text{C-BEL}_{\{i,j\}}(\mathrm{INT}(i,\varphi) \wedge \mathrm{GOAL}(j,\mathtt{stit}(i,\varphi)))$$

where $\mathtt{stit}(i,\varphi)$ means that agent i sees to it (takes care) that φ becomes true (see [45]).

Social commitments with respect to actions are defined by the axiom:

SC2

$$\mathrm{COMM}(i,j,\alpha) \leftrightarrow \mathrm{INT}(i,\alpha) \wedge \mathrm{GOAL}(j, done(i,\alpha)) \wedge$$
$$\text{C-BEL}_{\{i,j\}}(\mathrm{INT}(i,\alpha) \wedge \mathrm{GOAL}(j, done(i,\alpha)))$$

Social commitment obeys positive introspection, i.e.

$$\mathrm{COMM}(i,j,\varphi) \rightarrow \mathrm{BEL}(i, \mathrm{COMM}(i,j,\varphi)).$$

This follows from the awareness condition included in the defining axiom itself.

The complex social action $\texttt{commit}(i,j,\varphi)$ will not be defined further here. Informally speaking, it takes care a social commitment $\mathrm{COMM}(i,j,\varphi)$ is established. This is, however, achieved by a rather complex process, involving possibly complex communication (see [20]).

5. Collective motivational attitudes

After defining social commitment between two agents, we are ready to move to the collective level of cooperation. In our approach, teams are created on the basis of *collective intentions*, and exist as long as the collective intention between team members exists. A collective intention may be viewed as an inspiration for team activity, whereas the collective commitment reflects the concrete manner of achieving the intended goal by the team. This concrete manner is provided by planning. Thus, our approach to collective commitments is plan-based. However, some agents in the team may not have delegated actions while still being involved in the collective intention and the collective commitment.

Collective intention and collective commitment are not introduced as primitive modalities, with some restrictions on the semantic accessibility relations (as in e.g. [9]). We do give necessary and sufficient, but still minimal, conditions for such collective motivational attitudes to be present. In this way, we hope to make the behavior of a team easier to predict.

5.1. Collective intentions

In this paper, we focus on strictly cooperative teams, which makes the definition of collective intention rather strong. In such teams, a necessary condition for a collective intention $\text{C-INT}_G(\varphi)$ is that all members of the team G have the associated individual intention $\mathrm{INT}(i,\varphi)$ towards the overall goal φ. However, to exclude the case of competition, all agents should also *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 $\text{M-INT}_G(\varphi)$. Thus, $\text{M-INT}_G(\varphi)$ is meant to be true if everyone in G intends φ ($\text{E-INT}_G(\varphi)$), everyone in G intends that everyone in G intends φ ($\text{E-INT}_G(\text{E-INT}_G(\varphi))$), etc. Formalizing the above two conditions, $\text{E-INT}_G(\varphi)$ (standing for "everyone intends") corresponds to to the semantic condition that $\mathcal{M}, w \models \text{E-INT}_G(\varphi)$ iff for all $i \in G$, $\mathcal{M}, w \models \mathrm{INT}(i,\varphi)$. Then $\mathcal{M}, w \models \text{M-INT}_G(\varphi)$ iff φ holds in all worlds reachable in one or more steps by I_i arrows for $i \in G$.

The resulting system 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 in [21] which is analogous to the one for common knowledge in [23]).

The distinguishing features of collective intentions ($\text{C-INT}_G(\varphi)$) over and above mutual ones, is that all members of the team are aware of the mutual intention, that is, they have a collective belief about this

(C-BEL$_G$(M-INT$_G(\varphi)$). In [21], we introduce a formal definition which is extensively discussed and compared with alternatives such as joint intention theory and SharedPlans theory [37, 28, 51].

The above conditions are captured by the following axioms:

M1 E-INT$_G(\varphi) \leftrightarrow \bigwedge_{i \in G}$ INT(i, φ).

M2 M-INT$_G(\varphi) \leftrightarrow$ E-INT$_G(\varphi \wedge$ M-INT$_G(\varphi))$

M3 C-INT$_G(\varphi) \leftrightarrow$ M-INT$_G(\varphi) \wedge$ C-BEL$_G$(M-INT$_G(\varphi))$

RM1 From $\varphi \rightarrow$ E-INT$_G(\psi \wedge \varphi)$ infer $\varphi \rightarrow$ M-INT$_G(\psi)$ (Induction Rule)

Note that this definition of collective intention is stronger than the one given in our older work [17, 18]. Let us remark that, even though C-INT$_G(\varphi)$ seems to be an infinite concept, collective intentions may be established in practice in a finite number of steps: an initiator persuades all potential team members to adopt a mutual intention, and, if successful, announces that the mutual intention is established [12, 13].

In circumstances where communication is hampered but agents have to cooperate, they must sometimes make do with a less strong version of collective intention, which does not include collective belief about the mutual intention, but instead a mutual intention to establish it [21]. On the other hand, it is easy to see that once a collective intention is established, agents are aware of it:

Lemma 5.1.

$$\text{C-INT}_G(\varphi) \rightarrow \text{C-BEL}_G(\text{C-INT}_G(\varphi)).$$

Proof:
We give a semantic sketch, which can be translated to an axiomatic proof because of completeness: so suppose $\mathcal{M}, w \models$ C-INT$_G(\varphi)$, then by **M3**,

$$\mathcal{M}, w \models \text{C-BEL}_G(\text{M-INT}_G(\varphi)),$$

thus by the second part of lemma 3.1,

$$\mathcal{M}, w \models \text{C-BEL}_G(\text{C-BEL}_G(\text{M-INT}_G(\varphi))).$$

Combining these two we get

$$\mathcal{M}, w \models \text{C-BEL}_G(\text{M-INT}_G(\varphi)) \wedge \text{C-BEL}_G(\text{C-BEL}_G(\text{M-INT}_G(\varphi))),$$

so by the first part of lemma 3.1,

$$\mathcal{M}, w \models \text{C-BEL}_G(\text{M-INT}_G(\varphi) \wedge \text{C-BEL}_G(\text{M-INT}_G(\varphi))),$$

which is, by **M3**, equivalent to

$$\mathcal{M}, w \models \text{C-BEL}_G(\text{C-INT}_G(\varphi)).$$

□

5.2. Collective commitments

Inspired by Castelfranchi [6], we treat collective commitment as the strongest motivational attitude to be considered in teamwork. In our opinion a collective intention is a necessary but not sufficient condition for a collective commitment to be present. A collective commitment is based on a social plan.

5.2.1. Social plans

Let us give a simple **example** of a social plan (see definition 2.4). Consider a team consisting of three mathematicians t (the theorem prover), l (the lemma prover) and c (the proof checker) who have as collective intention to prove a new mathematical theorem. Suppose during planning they define two lemmas, which also still need to be proved, and the following complex individual actions: *provelemma1*, *provelemma2* (to prove lemma 1, respectively 2), *checklemma1*,*checklemma2* (to check a proof of lemma 1, respectively 2), *provetheorem* (prove the theorem from the conjunction of lemmas 1 and 2), *checktheorem* (to check the proof of the theorem from the lemmas). One possible social plan they can come up with is the following. First, the lemma prover, who proves lemmas 1 and 2 in succession, and the theorem prover, who proves the theorem from the two lemmas, work in parallel, and subsequently the proof checker checks their proofs in a fixed order, formally:

$$P = \langle\langle\langle provelemma1\,, l\rangle; \langle provelemma2\,, l\rangle\rangle \parallel \langle provetheorem\,, t\rangle\rangle;$$
$$\langle\langle\langle checklemma1\,, c\rangle; \langle checklemma2\,, c\rangle\rangle; \langle checktheorem\,, c\rangle\rangle\rangle$$

We will use this context as a running example in section 8.

Both the association of actions to members and the temporal structure are reflected in the recursive definition of a *social plan expression*, adapted from [43] and inspired by dynamic logic. The plans on which collective commitments are based are always represented as social plan expressions as defined in section 2, definition 2.4.

The last part of the definition introduces the temporal relations between the execution of actions. Paradigmatic aspects of CPS like negotiation, communication and coordination are all involved in planning. Note that the team members' characteristics, such as agents' abilities, opportunities, intention and commitment strategies and resources, may already play a role at the stage of task division. Let us stress however, that they are of the primary importance at earlier stages of CPS, especially at the potential recognition level. For the detailed discussion of this process see [12, 13].

We do not elaborate here on the ways by which the final social plan is constructed - this subject has been exhaustively discussed in AI literature (see e.g. [1]). For the complex process of dialogue that comes to the fore in plan generation, see [20]. The result of the whole planning process is a plan P, represented as a social plan expression and the predicate $constitute(\varphi, P)$ stating that a plan P ensures a proper realization of the goal φ. Thus, the successful realization of the plan P should lead to the achievement of the main goal φ:

CS

$$constitute(\varphi, P) \rightarrow [\texttt{confirm}(succ(P))]\varphi$$

The way the predicate $constitute(\varphi, P)$ is constructed will be discussed in subsection 7.3.1.

5.2.2. Strong collective commitment

Let us start from stressing the crucial role of collective intention when creating the group: the team is *based* on this attitude. In other words, no teamwork is considered without a collective intention among team members. After the group is constituted, another stage of CPS is started, leading ultimately to a *collective commitment* between the team members. In this section, we will give one definition of collective commitment, namely strong collective commitment: its power fully reflects the collective aspects of CPS.

In general, definitions of collective commitment are based on the social plan P and can be established or maintained if the group has the associated collective intention, if not stated explicitly otherwise. The social plan should result from the main goal by task division, means-end analysis, and action allocation, as reflected in $constitute(\varphi, P)$ (see subsection 7.3.1). Additionally, if the group is planning collectively, especially from first principles, in the end the plan is known to all members, as reflected in the conjunct $\text{C-BEL}_G(constitute(\varphi, P))$.

In [22], we present a sort of tuning machine allowing to define different versions of collective commitments, reflecting different aspects of CPS, and applicable in different situations. The definitions differ with respect to the *aspects* of teamwork of which the agents involved are aware, and the *kind* of awareness present within a team. In this way a kind of calibration mechanism is provided for the system developer to tune a version of collective commitment fitting the circumstances. Finally, we focused attention on a few exemplar versions of collective commitment resulting from instantiating the general tuning scheme, and sketched for which kinds of organization and application domains they are appropriate. Strong collective commitment formed one of these examples, and we believe it appropriate in many contexts.

A *strong collective commitment* ($\text{S-COMM}_{G,P}$) is based on collective planning: the whole team does it together, including negotiating and persuading each other who will do what. In addition to collective planning, for every one of the actions α that occur in social plan P, there should be one agent in the group who is socially committed to at least one (mostly other) agent in the group to fulfill the action. Moreover, even if there is no public awareness in the team about every single social commitment $(\text{COMM}(i, j, \alpha))$ that has been established about particular actions from the social plan, still the group as a whole believes that things are under control, i.e., that every part of the plan is within somebody's responsibility. These conditions are formalized in the defining axiom for strong collective commitments:

$$\begin{aligned}&\text{S-COMM}_{G,P}(\varphi) \leftrightarrow \text{C-INT}_G(\varphi) \wedge \\ &\quad constitute(\varphi, P) \wedge \text{C-BEL}_G(constitute(\varphi, P)) \wedge \\ &\quad \bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha) \wedge \text{C-BEL}_G(\bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha))\end{aligned}$$

Strong collective commitments are well-suited to model self-leading teams [3, 41].

Note that teams of agents have positive introspection about strong collective commitments among them, even if negative introspection does not follow from the defining axiom. Thus,

theorem: awareness of strong collective commitment

$$\text{S-COMM}_{G,P}(\varphi) \rightarrow \text{C-BEL}_G(\text{S-COMM}_{G,P}(\varphi)).$$

The proof is immediate from the definition and lemmas 5.1 and 3.1.

Remarks about collective commitment

The definition of strong collective commitment is not overloaded, and therefore easy to understand and to use. Some other approaches to collective commitments (see e.g. [37, 52]) introduce other aspects of collective attitudes, not treated here. For example, Wooldridge and Jennings consider triggers for commitment adoption formulated as preconditions [52]. If needed, these may be incorporated into our framework as well by adding an extra axiom. Note that in contrast to other approaches ([52],[37]), the collective commitment is not iron-clad: it may vary in order to adapt to changing circumstances, in such a way that the collective intention on which it is based can still be reached.

Our approach is especially strong when re-planning is needed. In contrast to [52], using our definition of collective commitment it is often sufficient to revise some of the pair-wise social commitments, instead of involving the entire team in the re-planning process (in the strong versions of the definition). This is a consequence of basing collective commitment on an explicitly represented plan, and of building it from pair-wise social commitments. In effect, if the new plan resulting from the analysis of the current situation within the team and the environment is as close as possible to the original one, the process of re-planning is maximally efficient. This reconfiguration problem was treated extensively in [19], where an abstract reconfiguration algorithm was presented. The next part of this paper contains a formal description of the situations globally treated in the algorithm.

6. Dynamic aspects of motivational attitudes

The previous sections recall the static theory of a BDI system built on individual and collective informational and motivational attitudes. In the rest of the paper we will treat the dynamics of systems situated in a changing and possibly unpredictable environment. We will focus on collective aspects of CPS. In this process, agents take on intentions during the process of intention formation or adoption [13, 11]. We leave this stage implicit in this paper. In the next stage, in order to maintain a good balance between goal-directed and event-driven aspects of an agent's behavior, its intentions and commitments should be reconsidered from time to time. In other words, they should persist, but for *how long?* The key point is whether and in which circumstances an agent can drop an intention or a social commitment. If such a situation arises, the next question is how to deal with it responsibly. To answer these questions three kinds of intention strategies (blind, single-minded and open-minded) may be defined, analogously to [42, 51], according to the strength with which agents maintain their intentions.

The strongest strategy is followed by the *blindly committed* agent, who maintains its commitments until it actually believes that they have been achieved. Single-minded agents may also drop social commitments when they do not believe anymore that the commitment is realizable. For open-minded agents, the situation is similar as for single-minded ones, except that they can also drop social commitments if they do not aim for the respective goal anymore. All three kinds of agents communicate with their partner after dropping a social commitment. As it is not the main subject of this paper, we do not give the formal definitions here, directing interested readers to [17, 18].

The phases of dropping intentions and commitments will be modeled explicitly by introducing the actions $\texttt{unintend}(i, \varphi)$, standing for "agent i drops its intention to achieve φ", and $\texttt{uncommit}(i, j, \alpha)$,

standing for "agent i drops its social commitment towards agent j to do α". Two important aspects of these actions need to be addressed:

1. What are the exact preconditions for the actions `unintend` and `uncommit`?
2. What are the consequences of the actions `unintend` and `uncommit`?

Temporal logic is best suited to represent possible answers for question 1, whereas dynamic logic is better suited to formalize consequences of actions, which is needed to answer question 2. This will be briefly addressed in the next subsection.

The problem of the preconditions of the actions `unintend` and `uncommit` is more complex as it deals with the dynamics of CPS. In other words, we can deal with a variety of reasons to change agents' individual, social and collective motivational attitudes. For individual intentions and social commitments they are, at least partly, recognized and specified in intention and commitment strategies ([17, 18]). The problem of persistence of collective intention is briefly discussed in subsection 8.1. Finally, the changes in collective commitment, based on the collective intention in question, lead to an evolution of this group attitude. This part is extensively discussed and formally proved to be correct in section 8. More precisely, the evolution of collective commitment, treated in detail in the reconfiguration algorithm, is formally expressed in dynamic logic, leading to a high level specification of a real computer system.

The reconfiguration procedure is based on the generally recognized four level model of teamwork. Before describing particular cases of reconfiguration, we need to make sure that all the levels are properly specified, and then constructed. As a sort of idealization of this process, we introduce level-associated actions, viewed as *complex social actions* (see subsection 3.1 for a formal definition and subsection 7 for the specification of their behaviour). These rather complex actions are highly context- and application-dependent and need to be tailored for a specific system. As they do not obey any generic axiomatization, we do not give an axiom system characterizing them. Instead we formulate in the extended language of dynamic logic (see subsection 3.1) rather straightforward high-level properties, to be ensured by the system developer. These propositions occurring in the sequel are thus meant as *semantic validities*.

On the other hand, for social plan expressions (see subsection 3.1 for a definition), some axiom system could be easily built, but as it is not the main focus of this paper we have refrained from this.

6.1. Dropping individual intentions and social commitments

The action $\texttt{unintend}(i, \varphi)$ stands for "agent i drops its intention to achieve φ". As a partial answer to the third question posed above, we assume that the following general fact holds, namely if agent i intends φ and it is possible to perform $\texttt{unintend}(i, \varphi)$, then after its performance, the agent does not have the intention anymore. The circumstances under which $\texttt{unintend}(i, \varphi)$ cannot be performed are application-dependent (related to question 1 above), but for example in a situation where $\mathrm{INT}(i, \varphi)$ does not hold, $\texttt{unintend}(i, \varphi)$ is prevented; such a situation is explicitly excluded in the formal fact:

UI1

$$\mathrm{INT}(i, \varphi) \wedge \langle \texttt{unintend}(i, \varphi) \rangle \top \rightarrow [\texttt{unintend}(i, \varphi)] \neg \mathrm{INT}(i, \varphi)$$

One of the consequences of `unintend` is given by the following, which is semantically implied by the fact above and the negative intention introspection axiom $\mathbf{A8}_{IB}$.

UI2

$$\mathrm{INT}(i,\varphi) \wedge \langle \mathtt{unintend}(i,\varphi)\rangle \top \rightarrow [\mathtt{unintend}(i,\varphi)]\mathrm{BEL}(i, \neg \mathrm{INT}(i,\varphi))$$

Analogously, we give some general consequences related to the action of dropping a commitment, addressing question 2 above. The action $\mathtt{uncommit}(i,j,\alpha)$ stands for "agent i drops social commitment towards agent j to do α". The following fact holds:

UC1

$$\mathrm{COMM}(i,j,\alpha) \wedge \langle \mathtt{uncommit}(i,j,\alpha)\rangle \top \rightarrow$$
$$[\mathtt{uncommit}(i,j,\alpha)] \neg \mathrm{COMM}(i,j,\alpha)$$

One of the consequences of `uncommit` is given by the following restricted axiom of negative introspection; note that, because we do not have negative introspection for social commitments in general (cf. section 4.2), this does not follow from **S3**:

UC2

$$\mathrm{COMM}(i,j,\alpha) \wedge \langle \mathtt{uncommit}(i,j,\alpha)\rangle \top \rightarrow$$
$$[\mathtt{uncommit}(i,j,\alpha)] \mathrm{BEL}\,(i, \neg \mathrm{COMM}(i,j,\alpha))$$

The above facts are postulates to be ensured by system developer. Individual uncommit actions are also discussed in a dynamic logic framework in [34, 32].

7. The four levels of CPS

In our approach to CPS, collective intention is considered as an inspiration to a goal-directed activity expressed in terms of collective commitment. We assume that in a dynamic system collective commitment may evolve in order to ensure the proper realization of collective intention of the group. Thus, in the first place, one needs to guarantee that collective intention will last long enough (see the previous section). Next, one should monitor the construction, maintenance, and realization, i.e. an *evolution* of collective commitments in a dynamic system. We adopted the four-stage model of [53], containing the consecutive stages of *potential recognition*, *team formation*, *plan formation* and *team action*. The key point is to bind the appropriate individual, social and collective attitudes to these stages. However, especially with respect to collective intentions and collective commitments, our analysis differs from the one in [53]. The processes of potential recognition and team formation have been extensively discussed in [12, 13]. For more about the levels and the role of dialogue, see [20].

Now we specify a formal system realizing the above-mentioned consecutive stages. The stages are rather complex, needing extensive communication (especially if planning is done from first principles), discussed elsewhere [13, 20]. As a sort of idealization, we assume that the levels are realized by complex level-associated actions, called `potential-recognition`, `team-formation`, `task-division`, `means-end-analysis`, and `action-allocation`. These application-dependent actions will not be further refined here.

Even though the three stages of potential recognition, team formation and plan formation have been extensively discussed in the multiagent systems and artificial intelligence literature, the important phase of collective team action has received relatively little attention. The requirements of a constantly changing environment lead to the *reconfiguration problem*: when maintaining a collective commitment and its

constituent collective intention, social commitments, and individual intentions during plan execution, it is crucial that agents re-plan properly and efficiently when some members do not fulfill their delegated actions or are presented with new opportunities. The solution of this problem leads to the *reconfiguration algorithm* formulated by us in terms of the abstract levels and their (complex) interplay. See [19] for a presentation of the algorithm which is also repeated in appendix II.

This chapter deals with the evolution of collective commitments during reconfiguration — we will build a formal system based on this abstract algorithm. Thus, for all four levels, both the positive case (when the level-associated action succeeds) and the negative case (when this action fails) will be specified, and treated accordingly. Again, all these level-oriented postulates should be ensured by a system developer.

7.1. The potential recognition level

Analogous to [53], we consider CPS to begin when some agent in a multi-agent environment recognizes the potential for cooperative action in order to reach its goal. The input of this stage is an agent a, a goal φ plus a finite set $T \subseteq \mathcal{A}$ of agents from which a potential team may be formed. The output at this stage is the "potential for cooperation" ($\mathrm{POTCOOP}(\varphi, a)$) that agent a sees with respect to φ, meaning that φ is a goal of a, and there is a team G such that a believes that G can collectively achieve φ and are willing to participate in team formation; and either a cannot or doesn't desire to achieve φ in isolation. As this problem is not closely related to the subject of this paper, we refer the interested reader to [19] for a formal definition and extensive discussion.

Let us assume that potential recognition is realized by a complex action `potential-recognition`. Thus, in case of successful performance of this action by agent a we have:

Ps

$$succ(\texttt{potential-recognition}(\varphi, a)) \rightarrow \mathrm{POTCOOP}(\varphi, a)$$

However, the failure of `potential-recognition` action, meaning that agent a doesn't see any potential of cooperation w.r.t. φ, leads to the failure of the system.

Pf

$$failed(\texttt{potential-recognition}(\varphi, a)) \rightarrow do(\texttt{system-failure}(\varphi))$$

This uses the notation for results of actions inspired by dynamic logic, and stands for: after potential recognition has failed, the action $\texttt{system-failure}(\varphi)$ is done. The $\texttt{system-failure}(\varphi)$ and $\texttt{system-success}(\varphi)$ are realized by complex actions which will not be refined here. Their proper realization should be ensured by a system developer.

7.2. The team formation level

Suppose that agent a sees the potential for cooperation to achieve φ. Somewhat different from [53], we find that during the team formation stage agent a attempts to establish in some team G the *collective intention* $\text{C-INT}_G(\varphi)$ to make φ true. The input of this stage is agent a, a formula φ and sequence of potential teams as output by the potential recognition stage. The successful outcome of this stage is one team G from the sequence together with a collective intention among G to achieve φ, which includes corresponding individual intentions of all team members. Let us assume that team formation is realized by a complex action `team-formation`. Thus, in case of its successful performance we have:

Ts

$$succ(\texttt{team-formation}(\varphi, a, G)) \rightarrow \text{C-INT}_G(\varphi)$$

However, the failure of execution of the `team-formation` action, meaning that the collective intention w.r.t. φ cannot be established among any of the teams from a sequence chosen during the `potential-recognition` action, requires a return to the potential recognition stage to construct a new sequence of potential teams:

Tf

$$failed(\texttt{team-formation}(\varphi, a, G)) \rightarrow do(\texttt{potential-recognition}(\varphi, a))$$

7.3. The plan generation level

The input of this stage is a team G together with its collective intention $\text{C-INT}_G(\varphi)$. The successful outcome is a strong collective commitment $\text{S-COMM}_{G,P}(\varphi)$ of the team G based on the social plan P. Because of our strong notion of collective commitment, the successful outcome of plan generation is somewhat different than in [53].

When building a collective commitment in the group we always assume that the team is established, and the collective intention is in place. Then, a planning process starts. The intended result of this phase is the establishment of $constitute(\varphi, P)$, meaning informally that planning has been done correctly. Then, a collective team activity leading ultimately to the establishment of $\text{S-COMM}_{G,P}(\varphi)$ takes place.

7.3.1. Construction of $constitute(\varphi, P)$

We see planning as a three-step process.

The first step is *task division* or decomposition, in which the question is addressed how to decompose a complex task φ into (possibly also complex) subgoals. We assume that this phase is realized by a complex action `task-division` and its result is described by a predicate $division(\varphi, \sigma)$ standing for "the sequence σ is a result of task decomposition of goal φ into subgoals". Here, σ is a finite sequence of propositions standing for goals, for example $\sigma = \langle \varphi_1, \ldots, \varphi_n \rangle$. Thus, after successful realization of this stage, we have $division(\varphi, \sigma)$. This is formalized using dynamic logic as follows:

Ds

$$succ(\texttt{task-division}(\varphi, \sigma)) \rightarrow division(\varphi, \sigma)$$

Otherwise, we have

Df

$$failed(\texttt{task-division}(\varphi, \sigma)) \rightarrow \neg division(\varphi, \sigma)$$

The social action $\texttt{task-division}(\varphi, \sigma)$, as well as the ones corresponding to other levels, will not be decomposed further in this paper. In fact, they are rather complex group level actions, depending on the context and the application domain, as well as communication and coordination protocols between agents. For the two first stages of potential recognition and team formation, these actions have been further refined for a relatively flexible communication protocol based on dialogue theory [13]. In the present general context, we are mainly interested in the results of the level-associated actions.

Next follows the phase of *means-end analysis* determining means realizing ends, i.e. actions realizing particular subgoals. Let us stress that for any subgoal there may be many (possibly complex) actions realizing it. Again, we assume that this phase is realized by a complex action means-end-analysis, and its result is described by a predicate $means(\sigma, \tau)$ standing for "the action sequence τ is a result of means-end analysis for the subgoal sequence σ". Here, τ is a finite sequence of actions $\in \mathcal{A}\rfloor$, for example $\tau = \langle \alpha_1, \ldots, \alpha_n \rangle$. This is a generalization of standard means-end analysis, which is performed for a single goal at a time. Note that to each subgoal in σ, a number of actions may be associated, so that σ and τ may have different lengths. Thus, the result of successful realization of this stage is $means(\sigma, \tau)$, or formally in the dynamic logic format:

Ms

$$succ(\text{means-end-analysis}(\sigma, \tau)) \rightarrow means(\sigma, \tau)$$

Otherwise, we have

Mf

$$failed(\text{means-end-analysis}(\sigma, \tau)) \rightarrow \neg means(\sigma, \tau)$$

This step is followed by *action allocation,* in which the actions resulting from means-end analysis are given to team members. It is realized by a complex action *allocation.* This results first in pairs $\langle \alpha, i \rangle$ of action α and an agent i. To make allocation complete, the temporal structure among pairs $\langle \alpha, i \rangle$ should be established. This process of constructing a social plan is described by the predicate $allocation(\tau, P)$ standing for "P is a social plan resulting from allocation of a sequence of actions τ to interested team members". Thus, the result of successful realization of this stage may again be represented formally by:

As

$$succ(\texttt{action-allocation}(\tau, P)) \rightarrow allocation(\tau, P)$$

Otherwise, we have

Af

$$failed(\texttt{action-allocation}(\tau, P)) \rightarrow \neg allocation(\tau, P)$$

The predicate $constitute(\varphi, P)$ informally stands for "P is a correctly constructed social plan to achieve φ", with as formal definition:

C0

$$constitute(\varphi, P) \leftrightarrow$$
$$\exists\sigma\exists\tau[division(\varphi, \sigma) \wedge means(\sigma, \tau) \wedge allocation(\tau, P)).$$

Note that in the predicates $division$, $means$ and $allocation$, it does not matter that the lengths of the subgoal sequence and the action sequence are not fixed in advance; one can always code finite sequences in such a way that their length may be recovered from the code.

The overall planning phase is considered as a three-step process consisting of the complex action `task-division;means-end-analysis;action-allocation`. In case of successful performance of this action a correct plan is constructed:

C1

$$succ(\texttt{task-division}(\varphi, \sigma); \text{means-end-analysis}(\sigma, \tau);$$
$$\texttt{action-allocation}(\tau, P)) \rightarrow constitute(\varphi, P)$$

The case of failure of plan formation will now be considered more in detail, looking carefully at which step the failure actually takes place. Thus, the failure of `task-division`, meaning that no task division for φ was found, requires a return to team formation in order to establish a collective intention in the chosen new team. It may be viewed as reconfiguration of the team together with revision of the collective intention and the respective individual attitudes.

Dd

$$failed(\texttt{task-division}(\varphi, \sigma)) \rightarrow do(\texttt{team-formation}(\varphi, a, G'))$$

The failure of means-end-analysis, meaning that there are no available means to realize some subgoals from a goal sequence σ, requires a return to the task division level in order to create a new sequence of subgoals.

Md

$$failed(\text{means-end-analysis}(\sigma, \tau)) \rightarrow do(\texttt{task-division}(\varphi, \sigma'))$$

The failure of `action-allocation`, meaning that some of the previously established actions cannot be allocated to agents in G, requires a return to means-end analysis for new means that could be allocated to members of the current team.

Ad

$$failed(\texttt{action-allocation}(\tau, P)) \rightarrow do(\text{means-end-analysis}(\sigma, \tau'))$$

In the two above situations, when backtracking is considered, some partial results of earlier stages already established may be reused to achieve $constitute(\varphi, P')$ for a new social plan P'. This way a sort of system conservativity is maintained (see [19] for a more detailed discussion). Thus, it is assumed that the following holds:

C2

$$division(\varphi, \sigma) \wedge means(\sigma, \tau) \wedge succ(\texttt{action-allocation}(\tau, P') \rightarrow constitute(\varphi, P')$$

C3

$$division(\varphi, \sigma) \wedge succ(\text{means-end-analysis}(\sigma, \tau'); \texttt{action-allocation}(\tau', P')) \rightarrow constitute(\varphi, P')$$

In fact, **C2** follows directly from **C0** and **As**.

7.3.2. Construction of $\mathrm{S\text{-}COMM}_{G,P}(\varphi)$ by communication

The goal of this part of plan generation is to establish a collective commitment in a team G towards an overall goal φ. The method of achieving this, including different types of communicative acts and/or communication protocols, depends on the type of commitment in a team (as discussed in subsection 5.2). As we focus on strictly cooperative teams, in the sequel we will consider $\mathrm{S\text{-}COMM}_{G,P}$. In this case, three phases of communication will be necessary. Here we will comment solely on formal results of this process without coming into details about communication, which is not the main subject of this paper. Usually, such a complex process requires rather compound communicative acts.

The communication starts when (i) $\text{C-INT}_G(\varphi)$ and (ii) $constitute(\varphi, P)$ are in place as a result of previous stages. When considering $\text{S-COMM}_{G,P}(\varphi)$, the (iii) $\text{C-BEL}_G(constitute(\varphi, P))$ has to be established in the first place. After the social plan is communicated, all agents from a team need to socially commit to carry out their respective actions, and to communicate about these social commitments in order to establish pairwise mutual beliefs about them, leading finally to $\text{COMM}(i, j, \alpha)$ for all actions α in the plan, and then to a collective belief that all relevant commitments have been made.

The successive phases to establish these collective beliefs may be handled by different protocols. By means of the chosen protocol first

(iv) $\bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha)$

needs to be in place, after which a new phase of communicative acts should lead to the establishment of

(v) $\text{C-BEL}_G(\bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha))$.

This way, after the appropriate sequence of communicative acts, together with (i), (ii), and (iii), a strong collective commitment $\text{S-COMM}_{G,P}(\varphi)$ in a team G, based on social plan P towards a goal φ, holds.

Let `construction` be the application-dependent complex social action establishing (iv) and (v), obeying the following postulate:

CTR

$$\text{C-INT}_G(\varphi) \wedge constitute(\varphi, P) \wedge succ(\texttt{construction}(\varphi, G, P)) \rightarrow \text{S-COMM}_{G,P}(\varphi)$$

This information exchange concludes the *collective* part of plan generation, and the team is ready to start *team action*. Section 8 treats what happens in a dynamic environment, where some actions from the social plan fail.

7.3.3. Frame axioms for plan generation

We assume that the system developer takes care that all complex actions executed during plan generation, when carried out in the appropriate order, do not disturb the partial planning results created previously. For example, the result of task division stays intact during subsequent means-end analysis, action allocation and construction. Thus, we have the following axiom schemas:

FR1

$$succ(\texttt{action-allocation}(\tau, P); \texttt{construction}(\varphi, G, P)) \rightarrow allocation(\tau, P)$$

FR2

$$succ(\text{means-end-analysis}(\sigma, \tau); \texttt{action-allocation}(\tau, P); \texttt{construction}(\varphi, G, P)) \rightarrow means(\sigma, \tau) \wedge allocation(\tau, P)$$

FR3

$$succ(\texttt{task-division}(\varphi,\sigma);\text{means-end-analysis}(\sigma,\tau);$$
$$\texttt{action-allocation}(\tau,P);\texttt{construction}(\varphi,G,P))$$
$$\rightarrow division(\varphi,\sigma) \wedge means(\sigma,\tau) \wedge allocation(\tau,P)$$

The above axioms do not form a complete listing of all frame axioms adopted in our approach, but just the examples needed to reason about the complex social actions taking place during plan generation.

7.4. The team action level: reconfiguration

During team action or plan execution, all team members start executing their adequate agent-specific actions from the $\text{S-COMM}_{G,P}(\varphi)$. In terms of motivational attitudes, team action amounts to the maintenance of *social commitments* and associated *individual intentions*. The successful outcome of this stage is that all actions making up the social plan P have been carried out by the agents who were socially committed to do them, and that by the success of their actions the overall goal φ has been achieved. In the more common non-perfect case disturbances require reconfiguration of the system at some point. As this may happen at any moment, the team action level will be referred to as the *reconfiguration process*.

The successful realization of team action finishes the evolution of the team and its motivational attitudes. Before this expected situation takes place, all aspects of evolution are treated in the reconfiguration process. During this phase communication, including all types of dialogue, cooperation, and coordination take place.

8. Evolution of commitments during reconfiguration

Even though the definition of collective commitment is intuitive, its complexity calls for a rigorous maintenance of all motivational attitudes involved in CPS. In a dynamic environment, team members may fail to bring their tasks to a good end or new opportunities may appear. The explicit model of teamwork helps the team to monitor its performance, and to replan based on the present situation. In [19] we discuss how the abstract reconfiguration algorithm helps to do this in an effective way. Here, in contrast, we concentrate on the maintenance of the collective commitments during reconfiguration.

The collective commitment is the attitude needed to start the plan realization. Once the process is underway, the collective commitment may evolve, so that the collective intention (which naturally persists during plan realization) is finally achieved, if possible. More precisely, the evolution of collective commitment may be connected with the evolution of collective intention, in the sense that then team involved in collective intention may evolve, though the overall goal remains the same.

This section is built on the *reconfiguration algorithm*, which is described in [19]. It distinguishes several cases that can occur during plan execution. These cases are based on an analysis of different kinds of failure and success of individual actions, i.e. on strictly technical aspects of team action. In this sense the presented analysis is universal: it does not take into account the strenght of a commitment between team members, meaning that is independent on a kind of collective commitment we deal with. Agents' awareness of the situation they remain involved in is left aside. Thus, when introducing this aspect of CPS, the analysis should be extended. Here we aim solely at formulating a *minimal* set of properties ensuring a recofiguration process to be correct.

To properly deal with the variety of situations the reason of disturbances has to be recognized. We say that the execution of an action α fails for an *objective reason* R, denoted as $objective_G(R, \alpha)$, if R implies logically that α is not realizable by anybody in the present team G in the current state of the world, formally, that $\neg\exists i \in G(able(i, \alpha) \wedge opp(i, \alpha))$. The needed information may be achieved in different ways, depending on the problem solving domain and the organizational structure of the team. One possible way is by information seeking (for example by the initiator) about team members' abilities and current opportunities, followed by a general announcement. This communication may be formalized similarly to the information seeking phase for the potential recognition stage in [13, 12], where the needed speech acts and their effects on the communicators' belief states are formalized. In this context, the objectivity is of course an idealization.

Moreover, the complex actions `system-failure`(φ) and `system-success`(φ) are introduced. Again, like in the case of the level-associated actions, they are context-dependent and will not be decomposed here. In terms of motivational attitudes, they are interpreted as the failure and the success of collective intention, respectively. The failure of collective intention is equivalent to the failure of the reconfiguration algorithm.

In short, during plan execution a number of different cases is treated by the reconfiguration algorithm, all of them leading to changes in the agents' attitudes. It may be helpful to keep in mind the analogue of backtracking. In the successful case, all agents successfully perform the actions to which they socially committed, leading to `system-success` (φ) (see case 1 below). Otherwise, the unsuccessful case 2 is split into a number of subcases, according to the reasons of failure and the possibility of action reallocation:

- A new action allocation succeeds (case 2a)
- A new action allocation fails, and
  - A failed action blocks achieving the collective intention (case 2b), or
  - No failed action blocks achieving the collective intention, and
    - a new means-end analysis, followed by action allocation, succeeds (case 2c), or
    - a new means-end analysis, followed by action allocation, fails, and
      - a new task division, followed by means-end analysis and action allocation, succeeds (case 2d), or
      - a new task division, followed by means-end analysis and action allocation, fails.

Thus, if some actions failed but no action failed for an objective reason, first a new `action-allocation` is attempted. This new action allocation, in which failed actions are assigned to other team members, may be successful (case 2a). If `action-allocation` fails, on the other hand, we consider two cases: either the failed action was necessary for achieving the collective intention; i.e. there is an objective reason R for the failure of the action, that implies that the overall goal φ will never be achieved by the current team, leading to `system-failure` (φ) (case 2b).

If the overall goal is not blocked in this way, a new means-end-analysis is attempted for the current sequence of subgoals of φ. This new means-end analysis, followed by a new `action-allocation`, may be successful (case 2c). If it fails, however, a new `task-division` is attempted for the overall goal of the system, followed by means-end-analysis and `action-allocation`. This may be successful (case 2d). If not, on the other hand, a return is made to the level of *team formation* in order to establish a new

team realizing the overall goal φ of the system (see 7.2). Now we are ready to treat the cases of evolution of the collective commitment according to the reconfiguration algorithm.

In the proofs, we will make use of properties for complex actions, that follow immediately from correct construction of the function $perfco$ and the definition of `confirm`:

Proposition 8.1. In all Kripke models $\mathcal{M}$ and worlds w, and complex actions β_1, β_2, we have

$$\mathcal{M}, w \models succ(\beta_1; \beta_2) \rightarrow succ(\beta_2) \text{ and}$$

$$\mathcal{M}, w \models [\texttt{confirm}(\psi)]\chi \leftrightarrow (\psi \rightarrow \chi)$$

The second of these has already been given in subsection 3.1.

We will illustrate all cases with the theorem-proving example that first appeared in subsection 5.2.1. Here follow some more details:

Running example The team $G = \{t, l, c\}$ have created a collective intention to overall goal φ="theorem T has been proved". During task division they created the sequence of subgoals $\sigma = \langle \sigma_1, \sigma_2 \rangle$, with σ_1 = "lemmas relevant for T have been proved" and σ_2= "theorem T has been proved from lemmas". During means-end analysis, complex actions have been found to achieve these subgoals, namely the sequence $\tau = \langle$*provelemma1, provelemma2, checklemma1,checklemma2,provetheorem, checktheorem*$\rangle$. During action allocation the team divided these actions and created a temporal structure, resulting in social plan P:

$$P = \langle\langle\langle provelemma1\,, l\rangle; \langle provelemma2\,, l\rangle\rangle \parallel \langle provetheorem\,, t\rangle\rangle;$$
$$\langle\langle\langle checklemma1\,, c\rangle; \langle checklemma2\,, c\rangle\rangle; \langle checktheorem, c\rangle\rangle\rangle$$

They collectively made sure that their plan was correct ($constitute(\varphi, P)$) and publically established pairwise social commitments:

$$\mathrm{COMM}(l, t, provelemma1\,) \wedge \mathrm{COMM}(l, t, provelemma2\,) \wedge$$
$$\mathrm{COMM}(t, l, provetheorem) \wedge \mathrm{COMM}(c, l, checklemma1\,) \wedge$$
$$\mathrm{COMM}(c, l, checklemma2) \wedge \mathrm{COMM}(c, t, checktheorem)$$

Case 1: The successful case

In this case, everything goes right after the establishment of the collective commitment. Thus at the level of action execution, all agents carry out the actions making up social plan P according to the given temporal structure and the action allocation.

property: the successful case

If a collective commitment $\text{S-COMM}_{G,P}(\varphi)$ holds and P has just been successfully executed, then φ holds. In other words, for all Kripke models $\mathcal{M}$ in which the teamwork axioms hold, and all worlds w,

$$\mathcal{M}, w \models \text{S-COMM}_{G,P}(\varphi) \rightarrow [\texttt{confirm}(succ(P))]\varphi$$

Proof Suppose $\mathcal{M}, w \models \text{S-COMM}_{G,P}(\varphi)$. Then, using the definition of strong collective commitment, $\mathcal{M}, w \models \mathit{constitute}(\varphi, P)$. Finally, by axiom **CS**,
$\mathcal{M}, w \models [\texttt{confirm}(\mathit{succ}(P))]\varphi$.

The example In this case, l has proved the two lemmas, t has proved the theorem from the two lemmas, and c has found all the proofs to be correct. Indeed, after such a successful plan execution, the overall goal has been achieved: theorem T has been proved.

Case 2: An action failed

In the next four subcases, at a certain moment during action execution an action fails. We treat the change of collective commitment according to the differentiation of reasons for failure given in the reconfiguration algorithm. However, independently on the reasons of failure the "old" collective commitment has to be dropped, because the social commitments with respect to the failed actions from $\text{S-COMM}_{G,P}(\varphi)$ do not exist anymore. We do not come into details with respect to the needed `unintend` and `uncommit` actions to be executed by agents involved in such a situation.

After an action failed, the situation is not a priori hopeless: the collective commitment may evolve. We divide this case into four subcases of varying difficulty.

Case 2a: Reallocation possible

If there are other agents that can realize the previously failed actions, that is, action reallocation is possible, a new plan P' can be established, as well as a new collective commitment based on it. This is expressed by the property.

property: reallocation possible Suppose that there is an $(i, \alpha) \in P$ such that $\mathit{failed}(i, \alpha)$ but no failed action failed for an objective reason, then for the current action sequence τ and a new social plan P' we have for all Kripke models $\mathcal{M}$ in which the teamwork axioms hold, and all worlds w,

$$\text{C-INT}_G(\varphi) \wedge \mathit{division}(\varphi, \sigma) \wedge \mathit{means}(\sigma, \tau) \rightarrow$$

$$[\texttt{confirm}(\mathit{succ}(\texttt{action-allocation}(\tau, P'); \texttt{construction}(\varphi, G, P')))]$$

$$\text{S-COMM}_{G,P'}(\varphi)$$

Proof Suppose $\mathcal{M}, w \models \text{C-INT}_G(\varphi) \wedge \mathit{division}(\varphi, \sigma) \wedge \mathit{means}(\sigma, \tau)$. Now by the second property in proposition 8.1, it suffices to show that if

$$\mathcal{M}, w \models \mathit{succ}(\texttt{action-allocation}(\tau, P'); \texttt{construction}(\varphi, G, P')),$$

then $\mathcal{M}, w \models \text{S-COMM}_{G,P'}(\varphi)$; so suppose

$$\mathcal{M}, w \models \mathit{succeeded}(\texttt{action-allocation}(\tau, P'); \texttt{construction}(\varphi, G, P')).$$

It immediately follows by axiom **FR1** that $\mathcal{M}, w \models \mathit{allocation}(\tau, P')$. Combined with $\mathcal{M}, w \models \mathit{division}(\varphi, \sigma) \wedge \mathit{means}(\sigma, \tau)$ this implies by axiom **C0** that $\mathcal{M}, w \models \mathit{constitute}(\varphi, P')$. On the other hand, by the first property of proposition 8.1 we derive $\mathcal{M}, w \models \mathit{succ}(\texttt{construction}(\varphi, G, P'))$ from

$$\mathcal{M}, w \models \mathit{succ}(\texttt{action-allocation}(\tau, P'); \texttt{construction}(\varphi, G, P')).$$

Thus $\mathcal{M}, w \models \text{C-INT}_G(\varphi) \wedge constitute(\varphi, P') \wedge succ(\texttt{construction}(\varphi, G, P')$ holds, so by postulate **CTR** we conclude $\mathcal{M}, w \models \text{C-COMM}_{G,P'}(\varphi)$, as desired.

The example Suppose that l does not succeed in proving lemma 1, and in fact believes that he cannot as he misses some knowledge about elliptic curves, which t does have. After t communicates that she will pitch in for l, $\text{COMM}(l, t, provelemma1)$ (and thus the old collective commitment) is dropped, and a new social plan is devised:

$$P = \langle\langle\langle provelemma2, l\rangle \parallel \langle\langle provelemma1, t\rangle; \langle provetheorem, t\rangle\rangle\rangle;$$
$$\langle\langle\langle checklemma1, c\rangle; \langle checklemma2, c\rangle\rangle; \langle checktheorem, c\rangle\rangle\rangle$$

Finally, a new strong collective commitment is constructed, containing the social commitment $\text{COMM}(t, l, provelemma1)$.

Case 2b: Some failed action blocks the goal

In this case at least some action α that was *necessary* for achieving the goal failed for an objective reason R in a strong sense, where R implies that nobody will ever succeed in executing α. This is the most serious negative case, generally leading to system-failure. [3]

The example Suppose that, while checking t's proof of the theorem from the lemmas, c discovers that not only the proof is wrong, but that there is a counterexample to the theorem itself. Indeed, this reason for failure of checking the proof blocks the overall goal, and the disillusioned team disbands.

Case 2c: New means-end analysis possible

When action reallocation is not possible, but no failed action blocks the overall goal, this means that for every relevant social plan P', allocation with respect to the current action sequence τ fails. In this situation, the old collective commitment is dropped again, but its evolution is possible if a new means-end analysis yields new actions realizing the failed subgoals, followed by a new allocation. This is expressed by the following property.

[3] To formalize this case and prove consequences, a more extended language is needed than the dynamic one used here. In this paper, we focus on highlighting the dynamic aspects of teamwork as expressed by the results of team actions, taking into account that the extended system combining temporal logic with dynamic multi-agent modal logics, is bound to be highly intractable. Thus, we give solely some hints to how this may be done. The Kripke model is extended with a discrete temporal structure branching towards the future (as in **CTL** and **CTL***) and the language is extended with operators **E** for "in future on some branch through the present point" (with its dual **A**), **P** for "somewhere in the past" and $\Diamond$ for "in future somewhere on this branch" (with its dual $\Box$). Thus, at a moment where action α has never succeeded before, j just failed executing it, and no agent will ever achieve it we have
$\mathcal{M}, w, t \models \neg\exists i \mathrm{P} succ(i, \alpha) \wedge failed(j, \alpha) \wedge \neg\exists i(\mathbf{E}\Diamond(succ(i, \alpha)))$ (1).
We then define "α is necessary for achieving φ" formally as
$\mathcal{M}, w, t \models \neg\exists i \mathrm{P}(succ(i, \alpha)) \rightarrow \neg\varphi$ (2).

It follows by temporal logic from (1) and (2) that $\mathcal{M}, w, t \models \Box\neg\varphi$, i.e. φ will never hold. Thus,if it is discovered that a failed action blocks the overall goal in the above way, the system fails and neither a collective intention nor an evolved collective commitment towards it will be established.

property: new means-end analysis possible Suppose that there is an $(i, \alpha) \in P$ such that $failed(i, \alpha)$ and no failed α blocks φ. Then for the current goal sequence σ and action sequence τ, and for every social plan P', there are τ' and P'' such that the following holds for all Kripke models $\mathcal{M}$ in which the teamwork axioms hold, and all worlds w:

$$\text{C-INT}_G(\varphi) \wedge division(\varphi, \sigma) \rightarrow$$
$$[\texttt{confirm}(failed(\texttt{action-allocation}(\tau, P')))]$$
$$[\texttt{confirm}(succ(\text{means-end-analysis}(\sigma, \tau');$$
$$\texttt{action-allocation}(\tau', P''); \texttt{construction}(\varphi, G, P'')))]$$
$$\text{S-COMM}_{G,P''}(\varphi).$$

Proof Suppose $\mathcal{M}, w \models \text{C-INT}_G(\varphi) \wedge division(\varphi, \sigma)$. Now by the second property in proposition 8.1, it suffices to show that if

$$\mathcal{M}, w \models succ(\text{means-end-analysis}(\sigma, \tau'); \texttt{action-allocation}(\tau', P'');$$
$$\texttt{construction}(\varphi, G, P'')),$$

then $\mathcal{M}, w \models \text{S-COMM}_{G,P''}(\varphi)$; so suppose

$$\mathcal{M}, w \models succ(\text{means-end-analysis}(\sigma, \tau'); \texttt{action-allocation}(\tau', P'');$$
$$\texttt{construction}(\varphi, G, P''))$$

It immediately follows by axiom **FR2** that $\mathcal{M}, w \models means(\sigma, \tau') \wedge allocation(\tau', P'')$. When combined with $\mathcal{M}, w \models division(\varphi, \sigma)$ this implies by axiom **C0** that
$\mathcal{M}, w \models constitute(\varphi, P'')$.

On the other hand, by the first property of proposition 8.1 we derive

$$\mathcal{M}, w \models succ(\texttt{construction}(\varphi, G, P''))$$

and, exactly as in case 2a, we derive $\mathcal{M}, w \models \text{S-COMM}_{G,P''}(\varphi)$ by **CTR**.

The example As in case 2a, suppose that l does not succeed in proving lemma 1, but now t and c do not believe they can prove it, either. The team does a new means-end analysis based on the old subgoal sequence, and comes up with some other lemmas (say 3, 4 and 5) that together hopefully imply the theorem. This gives rise to a new action sequence $\tau' = \langle$*provelemma3, provelemma4, provelemma5, checklemma3, checklemma4, checklemma5, provetheorem, checktheorem* $\rangle$. They allocate the actions in a similar way as before, creating a social plan P'', for example:

$$P'' = \langle\langle\langle\langle provelemma3, l\rangle; \langle provelemma4, l\rangle\rangle;$$
$$\langle provelemma5, l\rangle\rangle \parallel \langle provetheorem, t\rangle\rangle;$$
$$\langle\langle\langle\langle checklemma3, c\rangle; \langle checklemma4, c\rangle\rangle;$$
$$\langle checklemma5, c\rangle\rangle; \langle checktheorem, c\rangle\rangle\rangle$$

Finally, by public communication among them they establish new social commitments, resulting in a new strong collective commitment.

Case 2d: New task division possible

When neither action reallocation, nor a new means-end analysis is possible for the failed actions, this means that for the current τ, allocation with respect to τ fails to deliver any social plan P'; and then means-end analysis with respect to the current σ fails to deliver any action sequence τ'.

But the evolution of the collective commitment is still possible, if task division for the overall goal (φ) is executed, in order to establish a new goal sequence σ', followed by new rounds of means-end analysis, establishing a new action sequence τ'', and allocation, to create a new social plan P''. The following property describes the result.

property: new task division possible Suppose that there is an $(i, \alpha) \in P$ such that $failed(i, \alpha)$ and no failed α blocks φ. Then for the current goal sequence σ and action sequence τ, and for every social plan P' and action sequence τ', there are σ', τ'' and P'' such that:

$$\text{C-INT}_G(\varphi) \rightarrow$$

$$[\texttt{confirm}(\mathit{failed}(\texttt{action-allocation}(\tau, P')))]$$

$$[\texttt{confirm}(\mathit{failed}(\text{means-end-analysis}(\sigma, \tau')))]$$

$$[\texttt{confirm}(\mathit{succ}(\texttt{task-division}(\varphi, \sigma'); \text{means-end-analysis}(\sigma', \tau'');$$

$$\texttt{action-allocation}(\tau'', P''); \texttt{construction}(\varphi, G, P'')))]$$

$$\text{S-COMM}_{G,P''}(\varphi).$$

Proof Suppose $\mathcal{M}, w \models \text{C-INT}_G(\varphi)$. By the second property in proposition 8.1, it suffices to show that if

$$\mathcal{M}, w \models \mathit{succ}(\texttt{task-division}(\varphi, \sigma'); \text{means-end-analysis}(\sigma', \tau'');$$

$$\texttt{action-allocation}(\tau'', P''); \texttt{construction}(\varphi, G, P'')))$$

then $\mathcal{M}, w \models \text{S-COMM}_{G,P''}(\varphi)$; so suppose

$$\mathcal{M}, w \models \mathit{succ}(\texttt{task-division}(\varphi, \sigma'); \text{means-end-analysis}(\sigma', \tau'');$$

$$\texttt{action-allocation}(\tau'', P''); \texttt{construction}(\varphi, G, P'')).$$

It immediately follows by axiom **FR3** that $\mathcal{M}, w \models \mathit{division}(\varphi, \sigma') \wedge \mathit{means}(\sigma', \tau'') \wedge \mathit{allocation}(\tau'', P'')$. This implies by axiom **C0** that $\mathcal{M}, w \models \mathit{constitute}(\varphi, P'')$.

On the other hand, by the first property of proposition 8.1 we derive

$$\mathcal{M}, w \models \mathit{succ}(\texttt{construction}(\varphi, G, P''))$$

and, exactly as in case 2a, we conclude $\mathcal{M}, w \models \text{S-COMM}_{G,P''}(\varphi)$ by **CTR**.

The example Suppose after the theorem has been divided into lemmas several times, and all these times it turned out to be impossible for the team to prove some essential lemma. They may conclude that they

cannot prove the theorem by defining lemmas to be proved. Then they may come up with a completely different task division, e.g. $\sigma' = \langle \sigma_3, \sigma_4 \rangle$ where σ_3="a theorem 'isomorphic' to T has been found in a different area of mathematics" and σ_4= "a suitable translation between the two contexts has been defined". Now means-end analysis and action allocation will result in a social plan P'' very different from P; we do not come into details.

If task division is not successful, the story of the current team is completed in this way and a return to team formation is made. The contribution of this section was to show how the collective commitment evolves during the execution of the reconfiguration algorithm.

8.1. Persistence of the collective intention during teamwork

The previous section treated the evolution of collective commitment within a fixed team of agents. The agents could exchange their individual actions and /or create new social plans, as the basis for a new collective commitment, as long as the group was consolidated on the basis of collective intention.

The problem of persistence of individual and collective motivational attitudes is rather complex as it needs a careful coordination of the agent's personal and team perspective. For example, when an agent succeeded in its action, it may drop its social commitment towards this action. On the other hand, it remains involved in the team effort regarding:

- its own social commitment(s) wrt. other actions;
- monitoring the agents who have committed to it;
- the collective belief about all actions being committed to, as well as about $constitute(\varphi, P)$;
- the collective intention to achieve φ.

Let us recall that any team is created on the basis of collective intention, and conversely, when the collective intention exists no longer, the group may desintegrate. Therefore, the individual agents carry a special responsibility to protect the collective intention, and thus to refrain from dropping their corresponding individual intention to the overall goal if it is not absolutely necessary.

The continuing existence of collective intention is a necessary condition for any collective commitment among the team to persist. On the other hand, due to dynamic circumstances, social commitments may naturally change as results of agents' individual decisions, based on their individual commitment strategies. These changes lead to the evolution, but not to dropping any collective commitment, as long as the current team exists. If these possibilities are exploited but the team cannot work for the common goal anymore, the team must desintegrate. Then, the old collective intention is dropped. This inevitably leads to dropping of the associated collective commitment. Then, according to the reconfiguration algorithm, a new team is created, a collective intention with respect to overall goal φ within this team is established, and this way the process of *plan formation* is restarted. The return to other stages is not treated in this paper, as we are concerned with the maintenance of collective commitments, not with the attitudes associated with potential recognition and team formation.

9. Discussion and conclusions

This paper falls within a larger research program, an important part of which presents a *static* characterization of CPS with collective commitment as a central notion [21], constituting a complete theory of motivational attitudes in teamwork. This theory is built incrementally starting from individual intentions, through social commitments, leading ultimately to collective level to motivational attitudes, namely collective intentions and collective commitments. All these notions play a crucial role in a practical reasoning process. As they are defined in multi-modal logics their semantics is clear and well defined; this also enables to express many subtle aspects of CPS like various connections between agents.

The present paper, on the other hand, deals with the *dynamic* use of the above theory in a dynamic and often unpredictable environment. Discussions of reconfiguration in the MAS literature (see e.g. [47]) do not give specific attention to the dynamics of attitude revision in a team. In this paper, we have started to fill this gap. Our definitions of collective commitment ensure efficiency of reconfiguration in two ways. Firstly, different from [53], the motivational attitudes occurring in the definition of collective commitments are defined in a non-recursive way, allowing straightforward revision. Secondly, because only social commitments to individual actions appear, it is often sufficient to revise some of these and not to involve the whole team in replanning. Thus, the definitions of collective commitment have pragmatic power: agents can take the whole process of building and revising collective commitments into their own hands. The dynamic language allows to precisely describe the results of all relevant complex actions needed during reconfiguration. Let us stress the novelty of using dynamic logic to describe the dynamics of collective attitudes in BDI-systems.

The static definitions and dynamic properties given express solely vital aspects of CPS, leaving room for case-specific extensions. Within this scope both parts of the theory can be viewed as a set of *teamwork axioms*. They constitute a definition of motivational attitudes in BDI systems, as well as a specification of their evolution in a dynamic environment. This way they may serve a system designer as a high-level specification of a complete system. The next step may be the application of formal verification methods to verify the behavior of the system. Recently, at Institute of Informatics of Warsaw University a platform DORCAS enabling a user to build BDI systems containing collective motivational attitudes has been created. Moreover, an interesting instantiation of the reconfiguration algorithm dealing with emergency situations on a boat is already implemented. Based on this implementation a paper containing a working example of evolution of collective commitment is under preparation. Another example of reconfiguration in a team of researchers proving a new theorem was presented in [19].

The presented analysis of dynamic aspects of social and collective attitudes in teams of agents in CPS assumes a rather high level of idealization: solely strictly cooperative teams are considered. This leads to a strong definition of collective intention, based on agents' mutual intentions, and then to a plan-based collective commitment. Even though in [22] we introduced a general tuning mechanism to calibrate the strength of collective commitments fitting to a variety of circumstances, an essential ingredient of these definitions — agents' awareness — is formalized by means of a strong notion of common belief. We agree with Cristiano Castelfranchi that after investigating and formalizing the basic case of strictly cooperative teams, there is time to relax some of the strong assumptions underlying this research in order to take a closer look on weaker and more distributed forms of cooperation.

Also, this normal modal framework, like any logic based on standard Kripke semantics, suffers from well-known problems related to logical omniscience. Because of the necessitation rule, agents are supposed to know and intend all tautologies; and because of the distribution axiom, they are supposed to

know all logical consequences of their knowledge, and to intend all logical consequences of their intentions. This is clearly unrealistic. For epistemic logic, several solutions to the logical omniscience problem have been proposed, mostly based on non-normal modal logics (see [23, Ch. 9] and [39, Ch. 2] for overviews). Similar solutions have been proposed for individual intentions (see [35]). For future research, we plan to design a non-normal multi-modal logic suitable to solve logical omniscience problems for our framework characterizing those motivational attitudes essential for teamwork, namely collective intentions, and social and collective commitments.

10. Acknowledgements

We would like to thank Cristiano Castelfranchi for fruitful discussions about this work. We have also found remarks of the anonymous referee very valuable.

This work is supported by the Polish KBN Grant 7T11C 006 20.

11. Appendix I: Logical system for the modal operators

11.1. General axiom and rule

P1 All instantiations of propositional tautologies;

PR1 From φ and $\varphi \to \psi$, derive ψ; (Modus Ponens)

11.2. Axioms and rules for dynamic logic

P2 $[do(i,\alpha)](\varphi \to \psi) \to ([do(i,\alpha)]\varphi \to [do(i,\alpha)]\psi)$; (Dynamic Distribution)

P3 $[do(i,\texttt{confirm}(\varphi))]\psi \leftrightarrow (\varphi \to \psi)$;

P4 $[do(i,\alpha_1;\alpha_2)]\varphi \leftrightarrow [do(i,\alpha_1)][do(i,\alpha_2)]\varphi$;

P5 $[do(i,\alpha_1 \cup \alpha_2)]\varphi \leftrightarrow ([do(i,\alpha_1)]\varphi \wedge [do(i,\alpha_2)]\varphi$;

P6 $[do(i,\alpha^*)]\varphi \to \varphi \wedge [do(i,\alpha)][do(i,\alpha^*)]\varphi$; (Mix)

P7 $(\varphi \wedge [do(i,\alpha^*)](\varphi \to [do(i,\alpha)]\varphi)) \to [do(i,\alpha^*)](\varphi)$; (Induction)

PR2 From φ, derive $[do(i,\alpha)]\varphi$. (Dynamic Necessitation)

11.3. Axioms and rules for individual and collective belief

A2 $\mathrm{BEL}(i,\varphi) \wedge \mathrm{BEL}(i,\varphi \to \psi) \to \mathrm{BEL}(i,\psi)$ (Belief Distribution)

A4 $\mathrm{BEL}(i,\varphi) \to \mathrm{BEL}(i,\mathrm{BEL}(i,\varphi))$ (Positive Introspection)

A5 $\neg\mathrm{BEL}(i,\varphi) \to \mathrm{BEL}(i,\neg\mathrm{BEL}(i,\varphi))$ (Negative Introspection)

A6 $\neg\mathrm{BEL}(i,\bot)$ (Consistency)

C1 $\text{E-BEL}_G(\varphi) \leftrightarrow \bigwedge_{i\in G} \mathrm{BEL}(i,\varphi)$

C2 $\text{C-BEL}_G(\varphi) \leftrightarrow \text{E-BEL}_G(\varphi \wedge \text{C-BEL}_G(\varphi))$

RC1 From $\varphi \to \text{E-BEL}_G(\psi \wedge \varphi)$ infer $\varphi \to \text{C-BEL}_G(\psi)$ (Induction Rule)

R2 From φ infer $\mathrm{BEL}(i,\varphi)$ (Belief Generalization)

11.4. Axioms for individual motivational operators

A2$_D$ $\mathrm{GOAL}(i,\varphi) \wedge \mathrm{GOAL}(i,\varphi \to \psi) \to \mathrm{GOAL}(i,\psi)$ (Goal Distribution)

A2$_I$ $\mathrm{INT}(i,\varphi) \wedge \mathrm{INT}(i,\varphi \to \psi) \to \mathrm{INT}(i,\psi)$ (Intention Distribution)

R2$_D$ From φ infer $\mathrm{GOAL}(i,\varphi)$ (Goal Generalization)

R2$_I$ From φ infer $\mathrm{INT}(i,\varphi)$ (Intention Generalization)

A6$_I$ $\neg\mathrm{INT}(i,\bot)$ for $i=1,\ldots,n$ (Intention Consistency Axiom)

Interdependencies between intentions and other attitudes

A7$_{DB}$ $\mathrm{GOAL}(i,\varphi)\rightarrow\mathrm{BEL}(i,\mathrm{GOAL}(i,\varphi))$ (Positive Introspection for Goals)

A7$_{IB}$ $\mathrm{INT}(i,\varphi)\rightarrow\mathrm{BEL}(i,\mathrm{INT}(i,\varphi))$ (Positive Introspection for Intentions)

A8$_{DB}$ $\neg\mathrm{GOAL}(i,\varphi)\rightarrow\mathrm{BEL}(i,\neg\mathrm{GOAL}(i,\varphi))$ (Negative Introspection for Goals)

A8$_{IB}$ $\neg\mathrm{INT}(i,\varphi)\rightarrow\mathrm{BEL}(i,\neg\mathrm{INT}(i,\varphi))$ (Negative Introspection for Intentions)

A9$_{ID}$ $\mathrm{INT}(i,\varphi)\rightarrow\mathrm{GOAL}(i,\varphi)$ (Intention implies goal)

11.5. Axioms for social commitments

Here follows the defining axiom for social commitments with respect to propositions:

SC1

$$\mathrm{COMM}(i,j,\varphi)\leftrightarrow\mathrm{INT}(i,\varphi)\wedge\mathrm{GOAL}(j,\mathtt{stit}(i,\varphi))\wedge$$

$$\text{C-BEL}_{\{i,j\}}(\mathrm{INT}(i,\varphi)\wedge\mathrm{GOAL}(j,\mathtt{stit}(i,\varphi)))$$

where $\mathtt{stit}(i,\varphi)$ means that agent i sees to it (takes care) that φ becomes true (see [45]).

Social commitments with respect to actions are defined by the axiom:

SC2

$$\mathrm{COMM}(i,j,\alpha)\leftrightarrow\mathrm{INT}(i,\alpha)\wedge\mathrm{GOAL}(j,\mathit{done}(i,\alpha))\wedge$$

$$\text{C-BEL}_{\{i,j\}}(\mathrm{INT}(i,\alpha)\wedge\mathrm{GOAL}(j,\mathit{done}(i,\alpha)))$$

11.6. Axioms and rule for mutual and collective intentions

M1 $\text{E-INT}_G(\varphi)\leftrightarrow\bigwedge_{i\in G}\mathrm{INT}(i,\varphi)$.

M2 $\text{M-INT}_G(\varphi)\leftrightarrow\text{E-INT}_G(\varphi\wedge\text{M-INT}_G(\varphi))$

M3 $\text{C-INT}_G(\varphi)\leftrightarrow\text{M-INT}_G(\varphi)\wedge\text{C-BEL}_G(\text{M-INT}_G(\varphi))$

RM1 From $\varphi\rightarrow\text{E-INT}_G(\psi\wedge\varphi)$ infer $\varphi\rightarrow\text{M-INT}_G(\psi)$ (Induction Rule)

11.7. Axiom for strong collective commitments

CC

$$\begin{aligned}\text{S-COMM}_{G,P}(\varphi)\leftrightarrow{}&\text{C-INT}_G(\varphi)\wedge\\&\mathit{constitute}(\varphi,P)\wedge\text{C-BEL}_G(\mathit{constitute}(\varphi,P))\wedge\\&\bigwedge_{\alpha\in P}\bigvee_{i,j\in G}\mathrm{COMM}(i,j,\alpha)\wedge\text{C-BEL}_G(\bigwedge_{\alpha\in P}\bigvee_{i,j\in G}\mathrm{COMM}(i,j,\alpha))\end{aligned}$$

12. Appendix II: The reconfiguration algorithm

This appendix contains a refined version of an abstract reconfiguration algorithm originally presented in [19]. As the algorithm is an abstract one, it is meant to be instantiated for each specific application. In particular, the abstract level-associated actions, including belief revision and motivational attitudes revision, are rather complex. A global structure of the algorithm is based on backtracking search. In the generic case, when not using any domain-dependent information, context-dependent improvements as obtained in informed search methods cannot be made. However, forward checking whether an objectively failed action blocks the overall is performed. Some attitudes (like POTCOOP), as not relevant here, are not defined in this paper. For their definition and discussion see [19].

Reconfiguration algorithm:

```
begin
{input of the system:
φ - a goal of agent a;
a - the agent that is input to the system and that will recognize potential;
T - a set of agents from whom potential teams are selected}
A: potential-recognition (φ, a);
   {input: φ, a, T}
   if not (potential-recognition-succeeded) then
      {a does not see any potential for cooperation with respect to goal φ}
      system-failure (φ);
      STOP
   else
      {potential recognition succeeded - output: H - a sequence of teams H =(G_1, ..., G_n)
      with the potential to realize φ}
      initialization-of-motivational-attitudes(φ,a);
   fi;
   {the attitude POTCOOP(φ, a) is established}
B: team-formation (φ, a, G);
   {input: φ, a, H}
   if not (team-formation-succeeded) then
      {C-INT_G(φ) cannot be established among any of the teams from H;
      return to the potential recognition level for a to construct a new sequence of potential
      teams for which it sees cooperation potential w.r.t. φ}
      goto A
   else
      {team formation succeeded - output G: a set of agents aiming to realize φ;
      Suppose G is the first unused G_i from H for which the collective intention towards φ
      can be established}
      motivational-attitudes-assignment (φ, G);
   fi;
   {the collective intention C-INT_G(φ) is established here}
```

C: task-division (φ, σ);
{input: φ, G}
if not (task-division-succeeded) then
{task division failed; return to the team-formation level in order to attempt the first unused G_{i+1} from H to establish collective intention towards φ}
goto B
fi;
{task division succeeded - output: σ - a sequence of subgoals together realizing φ; i.e. the first part of a social plan P}
D: means-end analysis (σ, τ);
{input: φ, G, σ}
if not (means-end-analysis-succeeded) then
{means-end analysis failed; return to the task-division level}
goto C
fi;
{means-end analysis succeeded - output: τ - a sequence of actions together realizing goal sequence σ; i.e. the second part of a social plan P}
E: action-allocation (τ, P);
{input: φ, G, τ}
if not (action-allocation-succeeded) then
{action allocation failed; return to the means-end analysis level}
goto D
fi;
{action allocation succeeded - output: a social plan P corr. to action sequence τ}
motivational-attitude-assignment (φ, G, τ, P);
{the collective commitment S-COMM$_{G,P}(\varphi)$ is established (including corresponding social commitments to actions from τ)}
{plan execution part starts here}
F: plan-execution (φ, G, τ, P);
{input φ, G, τ, P}
if plan-execution-succeeded then
{all actions that constitute the plan P are successfully executed; agents' beliefs and motivational attitudes need to be revised to reflect that φ is achieved as well as agents' actions from τ}
belief-revision (φ, G, τ);
motivational-attitudes-revision (φ, G, τ, P);
system-success(φ);
STOP
elseif
{some action execution failed: differentiation of reasons for failure}
F1: if subjective-reason-for-failure($\varphi, P, G, \tau, \tau_1, \tau_2$) and
not (objective-reason-for-failure($\varphi, P, G, \tau, \tau_1, \tau_2$)) then
{τ_1: the sequence of actions from τ that have been successfully achieved thus far;
τ_2: the sequence of pairs (A, X) of actions that failed plus reasons for failure,

```
            where X=ob or X=sub}
            {for every action that failed, it failed for a subjective reason, i.e.
            ∀A ∈ τ ∀X((A, X) ∈ τ2 → X = sub)}
            belief-revision (G, τ, τ1, τ2));
            if action-reallocation-possible then
               {for every action that failed for a subjective reason, there is another team
               member believing it can achieve it, i.e. ∀A ∈ τ ∀M1, M2 ∈ G((A, sub) ∈ τ2
               ∧COMM(M1, M2, A) → ∃M3 ∈ G(M1 ≠ M3 ∧ BEL(M3, can(M3, A))));
               before an attempt at action reallocation based on P, τ1 and τ2 is made,
               a revision of motivational attitudes is needed}
               motivational-attitudes-revision (φ, P, G, τ);
               goto E
            elseif
               {no action reallocation is possible: a new means-end analysis is needed}
               goto D
            fi
         fi
         {there are also objective reasons for failure, i.e. ∃A ∈ τ((A, ob) ∈ τ2)}
         F2:  belief-revision (G, τ, τ1, τ2);
         {now it should be investigated whether the objective reason for failure blocks
         achieving the overall goal φ}
         if blocked(φ) then
            {the objective reason for failure blocks achieving the goal φ}
            system failure (φ);
            STOP
         else
            {the objective reason for failure does not block achieving the goal φ;
            a new means-end analysis is needed, return to the means-end analysis level}
            goto D
         fi
      fi
end
```

References

[1] Allen, J., Hendler, J., Tate, J., Eds.: *Readings in Planning*, Morgan Kaufman, Los Altos (CA), 1990.

[2] Belnap, N., Perloff, M.: Seeing to it that: A Canonical Form for Agentives, *Theoria*, **54**, 1988, 175–199.

[3] Beyerlin et al., M., Ed.: *Theories of Self-managing Work Teams*, JAI Press, Greenwich (CN), 1994.

[4] Bratman, M.: *Intention, Plans, and Practical Reason*, Harvard University Press, Cambridge (MA), 1987.

[5] Brown, M.: On the Logic of Ability, *Journal of Philosophical Logic*, **17**, 1988, 1–26.

[6] Castelfranchi, C.: Commitments: From Individual Intentions to Groups and Organizations, in: Lesser [36], 41–48.

[7] Castelfranchi, C.: The Social Nature of Information and the Role of Trust, *International Journal of Cooperative Information Systems*, **11(3)**, 2002, 381–403.

[8] Castelfranchi, C., Tan, Y.-H., Eds.: *Trust and Deception in Virtual Societies*, Kluwer, Dordrecht, 2001.

[9] Cavedon, L., Rao, A., Tidhar, G.: Social and Individual Commitment (Preliminary Report), in: *Intelligent Agent Systems: Theoretical and Practical Issues* (L. Cavedon, A. Rao, W. Wobcke, Eds.), vol. 1209 of *LNAI*, Springer Verlag, Berlin, 1997, 152–163.

[10] Cohen, P., Levesque, H.: Intention is Choice with Commitment, *Artificial Intelligence*, **42**, 1990, 213–261.

[11] Dignum, F., Conte, R.: Intentional Agents and Goal Formation: Extended Abstract, *Preproceedings Fourth International Workshop on Agent Theories, Architectures and Languages* (M. Singh, A. Rao, M. Wooldridge, Eds.), Providence, Rhode Island, 1997.

[12] Dignum, F., Dunin-Kęplicz, B., Verbrugge, R.: Agent Theory for Team Formation by Dialogue, *Intelligent Agents VII: Agent Theories, Architectures and Languages* (C. Castelfranchi, Y. Lesperance, Eds.), 1986, Springer Verlag, Berlin, 2001.

[13] Dignum, F., Dunin-Kęplicz, B., Verbrugge, R.: Creating Collective Intention through Dialogue, *Logic Journal of the IGPL*, **9**, 2001, 145–158.

[14] Dunin-Kęplicz, B., Radzikowska, A.: Actions with Typical Effects: Epistemic Characterization of Scenarios, in: Lesser [36], page 445.

[15] Dunin-Kęplicz, B., Radzikowska, A.: Epistemic Approach to Actions with Typical Effects, *Proceedings ECSQARU'95*, Fribourg, 1995.

[16] Dunin-Kęplicz, B., Radzikowska, A.: Modelling Nondeterminstic Actions with Typical Effects, *Proceedings DIMAS'95*, Cracow, 1995.

[17] Dunin-Kęplicz, B., Verbrugge, R.: Collective Commitments, in: Tokoro [48], 56–63.

[18] Dunin-Kęplicz, B., Verbrugge, R.: Collective motivational attitudes in cooperative problem solving, *Proceedings of the First International Workshop of Eastern and Central Europe on Multi-agent Systems (CEEMAS'99)* (V. Gorodetsky, Ed.), St. Petersburg, 1999.

[19] Dunin-Kęplicz, B., Verbrugge, R.: A Reconfiguration Algorithm for Distributed Problem Solving, *Electronic Modeling*, **22**, 2000, 68 – 86.

[20] Dunin-Kęplicz, B., Verbrugge, R.: The Role of Dialogue in Collective Problem Solving, *Proceedings of the Fifth International Symposium on the Logical Formalization of Commonsense Reasoning (Commonsense 2001)* (E. Davis, J. McCarthy, L. Morgenstern, R. Reiter, Eds.), New York, 2001.

[21] Dunin-Kęplicz, B., Verbrugge, R.: Collective Intentions, *Fundamenta Informaticae*, **51(3)**, 2002, 271–295.

[22] Dunin-Kęplicz, B., Verbrugge, R.: Calibrating collective commitments, *Proceedings of The 3rd International Central and Eastern European Conference on Multi-Agent Systems* (J. V.Marik, M. Pechoucek, Eds.), 2691, Springer Verlag, Berlin, 2003.

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

[24] Fischer, M., Ladner, R.: Propositional dynamic logic of regular programs, *Journal of Computer and System Sciences*, **18(2)**, 1979, 194–211.

[25] Gerbrandy, J., Groeneveld, W.: Reasoning about Information Change, *Journal of Logic, Language and Information*, **6(2)**, 1997, 147–169.

[26] Goldblatt, R.: *Logics of Time and Computation*, Number 7 in CSLI Lecture Notes, Center for Studies in Language and Information, Palo Alto (CA), 1992.

[27] Graedel, E.: Why is Modal Logic so Robustly Decidable?, *Bulletin of the EATCS*, **68**, 1999, 90–103.

[28] Grosz, B., Kraus, S.: Collaborative Plans for Complex Group Action, *Artificial Intelligence*, **86(2)**, 1996, 269–357.

[29] Halpern, J.: The Effect of Bounding the Number of Primitive Propositions and the Depth of Nesting on the Complexity of Modal Logic, *Artificial Intelligence*, **75**, 1995, 361–372.

[30] Halpern, J., Moses, Y.: Knowledge and Common Knowledge in a Distributed Environment, *Journal of the ACM*, **37**, 1990, 549–587.

[31] Harel, D., Kozen, D., Tiuryn, J.: *Dynamic Logic*, MIT Press, Cambridge, MA, 2000.

[32] van der Hoek, W., van Linder, B., Meyer, J.-J. C.: An Integrated Modal Approach to Rational Agents, in: *Foundations of Rational Agency* (A. Rao, M. Wooldridge, Eds.), Kluwer, Dordrecht, 1999, 37–75.

[33] Hustadt, U., Schmidt, R.: On Evaluating Decision Procedures for Modal Logics, *Proceedings IJCAI'97* (M. Pollack, Ed.), Morgan Kauffman, Los Angeles (CA), 1997.

[34] J-J. Ch. Meyer, W. v. d. H., van Linder, B.: A Logical Approach to the Dynamics of Commitments, *Artificial Intelligence*, **113**(1–2), 1999, 1–41.

[35] Konolige, K., Pollack, M. E.: A representationalist theory of intention, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)* (R. Bajcsy, Ed.), Morgan Kaufmann publishers Inc.: San Mateo, CA, USA, Chambéry, France, 1993.

[36] Lesser, V., Ed.: *Proceedings First International Conference on Multi-Agent Systems*, AAAI-Press and MIT Press, San Francisco, 1995.

[37] Levesque, H., Cohen, P., Nunes, J.: On acting together, *Proceedings Eighth National Conference on AI (AAAI90)*, AAAI-Press and MIT Press, Menlo Park (CA), Cambridge (MA), 1990.

[38] van Linder, B., van der Hoek, W., Meyer, J.-J. C.: Formalising Abilities and Opportunities of Agents, *Fundamenta Informaticae*, **34**, 1998, 53–101.

[39] Meyer, J.-J. C., van der Hoek, W.: *Epistemic Logic for AI and Theoretical Computer Science*, Cambridge University Press, Cambridge, 1995.

[40] Peleg, D.: Concurrent Dynamic Logic, *Journal of the ACM*, **34(2)**, 1987, 450–479.

[41] Purser, R., Cabana, S.: *The Self-managing Organization*, Free Press, New York (NY), 1998.

[42] Rao, A., Georgeff, M.: Modeling Rational Agents within a BDI-architecture, *Proceedings of the Second Conference on Knowledge Representation and Reasoning* (R. Fikes, E. Sandewall, Eds.), Morgan Kaufman, 1991.

[43] Rao, A., Georgeff, M., Sonenberg, E.: Social Plans: A Preliminary Report, *Decentralized A.I.-3* (E. Werner, Y. Demazeau, Eds.), Elsevier, Amsterdam, 1992.

[44] Searle, J. R.: *Speech Acts*, Cambridge University Press, Cambridge, 1969.

[45] Segerberg, K.: Bringing it About, *Journal of Philosophical Logic*, **18**, 1989, 327–347.

[46] Stulp, F., Verbrugge, R.: A knowledge-based Algorithm for the Internet Protocol TCP, *Bulletin of Economic Research*, **54**(1), 2002, 69–94.

[47] Tambe, M.: Teamwork in real-world, dynamic environments, in: Tokoro [48], 361–368.

[48] Tokoro, M., Ed.: *Proceedings Second International Conference on Multi-Agent Systems*, AAAI-Press, Menlo Park (CA), 1996.

[49] van Linder, B., van der Hoek, W., Meyer, J.-J. C.: Actions that Make you Change Your Mind, in: *Knowledge and Belief in Philosophy and Artificial Intelligence* (A. Laux, H. Wansing, Eds.), Akedemie Verlag, Berlin, 1995, 103–146.

[50] Vardi, M.: Why is Modal Logic so Robustly Decidable?, *DIMACS Series on Discrete Mathematics and Theoretical Computer Science*, **31**, 1997, 149–184.

[51] Wooldridge, M.: *Reasoning About Rational Agents*, MIT Press, Cambridge, MA, 2000.

[52] Wooldridge, M., Jennings, N.: Towards a Theory of Collective Problem Solving, in: *Distributed Software Agents and Applications* (J. Perram, J. Muller, Eds.), vol. 1069 of *LNAI*, Springer Verlag, Berlin, 1996, 40–53.

[53] Wooldridge, M., Jennings, N.: Cooperative Problem Solving, *Journal of Logic and Computation*, **9**, 1999, 563–592.