5

Creating Common Beliefs in Rescue Situations

Barbara Dunin-Kęplicz[1,2] and Rineke Verbrugge[3]

[1] Institute of Informatics, Warsaw University, Banacha 2,
02-097 Warsaw, Poland
keplicz@mimuw.edu.pl
[2] Institute of Computer Science, Polish Academy of Sciences, Ordona 21,
01-237 Warsaw, Poland
[3] Institute of Artificial Intelligence, University of Groningen, Grote Kruisstraat 2/1,
9712 TS Groningen, The Netherlands
rineke@ai.rug.nl

Summary. In some rescue or emergency situations, agents may act individually or on the basis of minimal coordination, while in others, full-fledged teamwork provides the only means for the rescue action to succeed. In such dynamic and often unpredictable situations agents' *awareness* about their involvement becomes, on the one hand, crucial, but one can expect that it is only *beliefs* that can be obtained by means of communication and reasoning. A suitable level of communication should be naturally tuned to the circumstances. Thus in some situations individual belief may suffice, while in others everybody in a group should believe a fact or even the strongest notion of *common belief* is the relevant one.

Even though common knowledge cannot in general be established by communication, in this paper we present a procedure for establishing common beliefs in rescue situations by *minimal* communication. Because the low-level part of the procedure involves file transmission (e.g. by TCP or alternating-bit protocol), next to a general assumption on trust some additional assumptions on communication channels are needed. If in the considered situation communication is hampered to such an extent that establishing a common belief is not possible, creating a special kind of *mutual intention* (defined by us in other papers) within a rescue team may be of help.

5.1 Introduction

Looking at emergency situations in their complexity, a rather powerful knowledge-based system is needed to cope with them in dynamic and often unpredicatble environment. In emergencies, coordination and cooperation are on the one hand vital, and on the other side more difficult to achieve than in normal circumstances. To make the situation even more complex, time is critical for rescues to succeed, and communication is often hampered. Also, usually expertise from different fields is needed. Multiagent systems exactly fit the bill: they deliver means for organizing complex, sometimes spectacular interactions among different, physically and/or logically distributed knowledge based entities [1]:

> A MAS can be defined as a loosely coupled network of problem solvers that work together to solve problems that are beyond the individual capabilities or knowledge of each problem solver.

This paper is concerned with a specific kind of MAS, namely a team. A *team* is a group in which the agents are restricted to having a common goal of some sort in which team-members typically cooperate and assist each other in achieving their common goal. Rescuing people from a crisis or emergency situation is a complex example of such a common goal.

Emergency situations may be classified along different lines. It is not our purpose to provide a detailed classification here, but an important dimension of classification is along the need for teamwork. A central joint mental attitude addressed in teamwork is *collective intention*. We agree with [2] that:

> Joint intention by a team does not consist merely of simultaneous and coordinated individual actions; to act together, a team must be aware of and care about the status of the group effort as a whole.

In some rescue situations, agents may act individually or on the basis of minimal coordination, while in others, full-fledged teamwork, based on a collective intention, provides the only means for the rescue action to succeed.

MAS can be organized using different paradigms or metaphors. For teamwork, BDI (Beliefs, Desires, Intentions) systems form a proper paradigm. Thus, some multiagent systems may be viewed as intentional systems implementing practical reasoning — the everyday process of deciding, step by step, which action to perform next. This model of agency originates from Michael Bratman's theory of human rational choice and action [3]. His theory is based on a complex interplay of informational and motivational aspects, constituting together a belief-desire-intention model of rational agency. Intuitively, an agent's *beliefs* correspond to information the agent has about the environment, including other agents. An agent's *desires* or goals represent states of affairs (options) that the agent would choose. Finally, an agent's *intentions* represent a special subset of its desires, namely the options that it has indeed chosen to achieve. The decision process of a BDI agent leads to the construction of agent's *commitment*, leading directly to action execution.

The BDI model of agency comprises beliefs referring to agent's informational attitudes, intentions and then commitments referring to its motivational attitudes. The theory of informational attitudes has been formalized in terms of epistemic logic as in [4, 5]. As regards motivational attitudes, the situation is much more complex. In Cooperative Problem Solving (henceforth CPS), a group as a whole needs to act in a coherent pre-planned way, presenting a unified *collective* motivational attitude. This attitude, while staying in accordance with individual attitudes of group members, should have a higher priority than individual ones. Thus, from the perspective of CPS these attitudes are considered on three levels: *individual*, *social* (bilateral), and *collective*.

When analyzing rescue situations from the viewpoint of BDI systems, one of the first purposes is to define the scope and strength of motivational and informational attitudes needed for successful team action. These determine the strength and scope of the necessary communication. In [6], [7], we give a generic method for the system developer to tune the type of collective commitment to the application in question, the organizational structure of the group or institution, and to the environment, especially to its communicative possibilities.

In this paper, however, the essential question is in what terms to define the *communication* necessary for teamwork in rescue situations. *Knowledge*, which always corresponds to the facts and can be justified by a formal proof or less rigorous argumentation, is the strongest and therefore preferred informational attitude. The strongest notion of knowledge in a group is *common knowledge*, which is the basis of all conventions and the preferred basis of coordination. Halpern and Moses proved that common knowledge of certain facts is on the one hand necessary for coordination in well-known standard examples, while on the other side, it cannot be established by communication if there is any uncertainty about the communication channel [4].

In practice in MAS, agents do with belief instead of knowledge for at least the following reasons. First, in MAS perception provides the main background for beliefs. In a dynamic unpredictable environment the natural limits of perception may give rise to false beliefs or to beliefs that, while true, still cannot be fully justified by the agent. Second, communication channels may be of uncertain quality, so that even if a trustworthy sender knows a certain fact, the receiver may only believe it. Common belief is the notion of group belief which is constructed in a similar way as common knowledge. Thus, even though it puts less constraints on the communication environment that common knowledge, it is still logically highly complex.

For efficiency reasons it is often important to minimize the level of communication among agents. This level should be tuned to the circumstances under consideration. Thus in some situations individual belief may suffice, while in others everybody in a group should believe a fact and again in the others the strongest notion of common belief is needed. In this paper we aim to present a method for establishing common beliefs in rescue situations by *minimal* communication. If in the considered situation communication is hampered to such an extent that establishing a common belief is not possible, we attempt some alternative solutions.

The paper is structured in the following manner. In section 5.2, a short reminder is given about individual and group notions of knowledge and belief, and the difficulty to achieve common belief in certain circumstances. Then, a procedure for creating common beliefs is introduced in section 5.3, which also discusses the assumptions on the environment and the agents that are needed for the procedure to be effective. Section 5.4 presents three case studies of rescue situations where various collective attitudes enabling appropriate teamwork are established, tuned to the communicative possibilities of the environment. Finally, section 5.5 discusses related work and provides some ideas about future research.

5.2 Knowledge and belief in groups

In multiagent systems, agents' *awareness* of the situation they are involved in is a necessary ingredient. Awareness in MAS is understood as a reduction of the general meaning of this notion to the state of an agent's beliefs (or knowledge when possible) about itself, about other agents as well as about the state of the environment, including the situation they are involved in. Assuming such a scope of this notion, different epistemic logics can be used when modelling agents' awareness. This awareness may be expressed in terms of any informational (individual or collective) attitude fitting given circumstances. In rescue situations, when the situation is usually rather complex and hard to predict, one can expect that only beliefs can be obtained.

5.2.1 Individual and common beliefs

To represent beliefs, we adopt a standard $KD45_n$-system for n agents as explained in [4], where we take $\mathrm{BEL}(i, \varphi)$ to have as intended meaning "agent i believes proposition φ". A stronger notion than the one for belief is knowledge, often called "justified true belief". The usual axiom system for individual knowledge within a group is $S5_n$, i.e. a version of $KD45_n$ where the consistency axiom is replaced by the (stronger) truth axiom $\mathrm{KNOW}(i, \varphi) \rightarrow \varphi$. We do not define knowledge in terms of belief. Definitions occurring in the MAS-literature (such as $\mathrm{KNOW}(i, \varphi) \leftrightarrow \varphi \wedge \mathrm{BEL}(i, \varphi)$, i.e. knowledge is true belief) have been shown to be overly simplistic [8]. An example where knowledge is stronger than true belief is one where an agent i believes φ to be true for a certain (unjustified) reason, and φ is in fact true (so $\varphi \wedge \mathrm{BEL}(i, \varphi)$ holds). For example, let φ be "Alex is falling into the water", and suppose that agent i sees John falling into the water, mistakenly taking him to be Alex; but that Alex is falling into the water as well, unseen by i. In this case one should *not* conclude that the agent *knows* that Alex is falling into the water.

In the sequel we will use belief instead of knowledge, because agents, due to uncertainty and problems with perception and communication in their dynamic and possibly unpredictable environment, usually do not attain knowledge of the facts relevant for teamwork, even if they are *logically* perfect reasoners.

For the need of teamwork one can define modal operators for group beliefs, in particular $\text{E-BEL}_G(\varphi)$ is meant to stand for "every agent in group G believes φ". The stronger operator *common belief* $\text{C-BEL}_G(\varphi)$ is similar to the one of common knowledge: everyone believes φ, everyone believes that everyone believes φ, and so on, ad infinitum. This is formalized as follows:

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

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

Axiom **C2** is often called the fixed-point axiom, showing how $\text{C-BEL}_G(\varphi)$ can be viewed as fixed point of the function $f(x) = \text{E-BEL}_G(\varphi \wedge x)$. Soundness of rule

R2 is proved by induction, to show from the antecedent $\models \varphi \rightarrow \text{E-BEL}_G(\psi \wedge \varphi)$ that $\models \varphi \rightarrow \text{E-BEL}_G^k(\psi \wedge \varphi)$; thus it is named Induction Rule. Alternative way of formalizing common belief is given by Meyer and van der Hoek in [5].

In contrast to common knowledge, which is always sure, common belief need not be truthful, thus in some situations $\text{C-BEL}_G(\varphi)$ may even become a common illusion. The axiom system governing individual and common belief is called $KD45_n^C$ (see [4, 9, 5] for more about these logics). We do not give details about the semantics here, but only a reminder that in a possible world w, agent i knows φ ($\text{KNOW}(i, \varphi)$) if and only if φ is true in all worlds v that are knowledge-accessible for i from world w, and similarly for belief $\text{BEL}(i, \varphi)$, where the belief-accessible worlds are checked for truth of φ.

5.2.2 Degrees of belief in a group

It is well-known that for teamwork, as well as coordination, it often does not suffice that a group of agents all believe or know a certain proposition ($\text{E-BEL}_G(\psi)$ or $\text{E-KNOW}_G(\psi)$), but they should commonly believe or know it ($\text{C-BEL}_G(\psi)$ or $\text{C-KNOW}_G(\psi)$). An example is formed by collective actions where the success of each individual agent is vital to the result, for example, lifting a heavy object together or coordinated attack. It has been proved that for such an attack to succeed, the starting time of the attack must be common belief (even common knowledge) for the generals involved [4].

Parikh has introduced a hierarchy of levels of knowledge between individual knowledge and common knowledge and, together with Krasucki, proved a number of interesting mathematical properties. It turns out that, due to the lack of the truth axiom, the similarly defined hierarchy between individual belief and common belief is structurally different from the knowledge hierarchy [10].

One advantage of common belief over "everybody believes" is that if C-BEL_G holds for ψ, then C-BEL_G also holds for all logical consequences of ψ. The same is true for common knowledge. Thus, agents reason in a similar way from ψ and commonly believe in this similar reasoning and the final conclusions. In short, common knowledge and common belief are hard to achieve, but easy to understand.

In cases in which only $\text{E-BEL}_G(\psi)$ has been established, it is much more difficult for agents to maintain a model of the other team members with respect to ψ and its consequences. However, establishing $\text{E-BEL}_G(\psi)$ places much less constraints on the communication medium than $\text{C-BEL}_G(\psi)$ does. Thus, the system developer's decision about the level k of group belief ($\text{E-BEL}_G^k(\psi)$) to be established, hinges on determining a good balance between communication and reasoning for a particular application.

5.2.3 Difficulties in attaining common knowledge

Halpern and Moses [11] proved a surprising result in the eighties: under some very natural assumptions, namely that processors do not change their local states simultaneously, common knowledge does not increase over a run (sequence of time steps)

in a distributed system. The well-known example of the two generals who do not manage to reach common knowledge about the time of attack, even if a messenger brings any number of acknowledgments back and forth, is an example of this result. If there is any uncertainty about the messenger making it to the other general, even about whether he may be delayed, common knowledge cannot be reached [4]. In rescue situations, there is almost always uncertainty about messages reaching the other party.

Note that Halpern and Moses' result does not carry over to common belief. Their proof hinges on the fact that if processors do not change their local states simultaneously, then any two global states in a sequence of time-steps are accessible to each other by a sequence of knowledge-accessibility relations. This is in turn based on the fact that other global states with the same local state are always knowledge-accessible for a processor, a fact that need not hold for belief-accessibility.

5.3 A procedure for creating common beliefs

Even though common knowledge cannot in general be established by communication, we will show that common belief can. In this context, it turns out to be an advantage that belief, in contrast to knowledge, need not be true. Thus, Halpern and Moses' impossibility results about the growth of common knowledge do not carry over to common belief. Even stronger, it is possible to give a procedure that can, under some assumptions, establish common beliefs.

Usually in MAS literature, it is assumed as a simplification that public announcements are always successful: announcements reach all group members, and in the end their content is commonly believed by the group. Such an assumption takes for granted that the communication medium is perfect and that no messages are lost, which is not the case in practice. In this paper, we relax this assumption on a perfect communication medium.

In this section, we will informally present a procedure for creating a common belief in a group, essentially by one initiator broadcasting an appropriate message to all agents in the group.

5.3.1 The procedure for creating common beliefs

Suppose an initiator a wants to establish C-BEL$_G(\varphi)$ within a fixed group $G = \{1, \ldots, n\}$, where $a \in G$. Informally and from a higher-level view, the procedure works by the initiator a sending messages as follows:

1. a sends the message φ to agents $\{1, \ldots, n\}$ in an interleaved fashion, where each separate message is sent from a to i using the alternating-bit protocol or TCP;
2. then in the same way, a sends the message C-BEL$_G(\varphi)$ to agents $\{1, \ldots, n\}$;
3. recipients send acknowledgements of bits received (as by the alternating-bit protocol and TCP) but need not acknowledge the receipt of the full message.

Finally, all agents believe the messages they receive from a, and a believes them as well. Thus, after all agents have received the messages, we will have BEL$(i, \varphi \wedge$ C-BEL$_G(\varphi))$ for all $i \leq n$, thus by axiom **C1**, we have E-BEL$_G(\varphi \wedge$ C-BEL$_G(\varphi))$, which by axiom **C2** is equivalent to C-BEL$_G(\varphi)$, as desired.

Notice that the reason that this procedure can establish common belief, whereas common knowledge can never be established, is exactly that common beliefs need not be true. Thus, initiator a may believe and utter $\varphi \wedge$ C-BEL$_G(\varphi)$ even if C-BEL$_G(\varphi)$ has not yet, in fact, be established. Thus, if φ is in fact true, $\varphi \wedge$ C-BEL$_G(\varphi)$ is a prime example of the belief-analogue of a "successful formula" as defined in dynamic epistemic logic, namely a formula that comes to be commonly believed by being publicly announced [12].

On a lower level, the procedure is built on a well-known protocol for file transmission. We give a short reminder here.

5.3.2 A file transmission protocol

There are two processors, let us say a sender S and a receiver R. The goal is for S to read a tape $X = \langle x_0, x_1, \ldots \rangle$, and to send all the inputs it has read to R over a communication channel. R in turn writes down everything it reads on an output tape Y. Unfortunately, the channel is not trustworthy: there is no guarantee that all messages arrive. On the other hand, if an agent repeats sending a certain message long enough, an instance of it will arrive eventually. This property is called *fairness*. Now one needs a protocol that satisfies the following two constraints, provided that fairness holds:

- *safety:* at any moment, Y is a prefix of X;
- *liveness:* every x_i will eventually be written on Y.

In the knowledge-based protocol below, $K_S(x_i)$ means that S knows that the i-the element of X is equal to x_i.

PROTOCOL FOR S:

```
S1 i :=0
S2 while true do
S3     begin read x_i;
S4         send x_i until K_S K_R(x_i);
S5         send "K_S K_R(x_i)" until K_S K_R K_S K_R(x_i)
S6         i := i + 1
S7     end
```

PROTOCOL FOR R:

$R1$ when $K_R(x_0)$ set i :=0
$R2$ while true do
$R3$ begin write x_i;
$R4$ send "$K_R(x_i)$" until $K_R K_S K_R(x_i)$;
$R5$ send "$K_R K_S K_R(x_i)$" until $K_R(x_{i+1})$
$R6$ i := i + 1
$R7$ end

Suitable implementations of this protocol have been proved to be correct for fair environments in which errors like deletion, mutation or insertion may occur, or even any combination of two of them, but not all three [13]. As a final remark, let us note that it is possible to rewrite the protocol without using any knowledge operators. The result is known as the 'alternating-bit protocol'. The often used Internet protocol TCP is a variation of this protocol, where transmission does not work bit by bit, but window by window, where the size of the window may be adapted to the available bandwidth [14].

5.3.3 The effectiveness of the procedure under some assumptions

There are three different kinds of assumptions that work together to make the procedure above effective.

Assumptions about the communication channels

Because the low-level part of the procedure involves file transmission from a to the other agents in G by the alternating-bit protocol or TCP, the assumptions of the chosen protocol need to hold, namely fairness, and presence of no more than two types of error (see subsection 5.3.2). Also, we make the usual that the sender a's state records all data elements it has read and acknowledgements received, and that all other receiver agents in G record all data elements they have written. These assumptions take care that, after a finite time, all agents have received the complete message $\varphi \wedge \text{C-BEL}_G(\varphi)$.

Assumptions on trust

Whenever communication between agents appears, the question of trust is inevitably involved. Though this paper is not meant to be yet another voice in the discussion about trust in commonsense reasoning, in order to make communication and reasoning based on it more context-sensitive, it is useful to distinguish different notions or levels of trust. For example, an agent can trust the other completely: ($TRUST(j, a)$ for j trusts a), or partially (e.g. $TRUST_\psi(j, a)$ for j trusts a w.r.t. formula ψ). See [15] for interesting discussions about trust in MAS.

It seems that for public announcements, the speaker's assertions are believed by the hearers as long as trust is present. Thus, after agent a asserts ψ to agent j in such

a context and j has received the message, we have (see [16]):
$TRUST_{\psi}(j, a) \rightarrow \mathrm{BEL}(j, \psi)$.

For the procedure of subsection 5.3.1 to work, an assumption is needed of partial trust of all agents in G with respect to agent a and the formulas φ and $\text{C-BEL}_G(\varphi)$. In particular, agent a needs to believe φ himself.

In order to create the necessary trust, it helps to make the communication procedure commonly known or believed in advance.

Assumptions on persistence of beliefs

As is often implicitly assumed in studies of communication protocol, we also assume that the recipients of a's messages do not drop their beliefs about φ again during the process of the communication procedure.

Proof sketch of effectiveness

When proving properties of knowledge-based communication protocols, it is commonly agreed to use a semantics of interpreted systems representing the behavior of a number of processors over time (see [13, 5]). We give a short review here.

At each point in time, each of the processors in a distributed system (agent in our case) is in some *local state*. All of these local states, together with the environment's state, form the system's *global state* at that point in time. These global states will be represented as possible worlds in a Kripke model. Thus, if one represents the global state as a vector of the local states, a system consisting of n processors $\{1, \ldots, n\}$ in environment e may be in global state $s = (s_e, s_1, \ldots, s_n)$; in an asynchronous environment, a local state may be represented as the sequence of distinct observations of the processor.

The state of the environment consists of those aspects of the distributed system that are relevant to an analysis of the problem at hand but that are not part of the local states of the processors. The accessibility relations are defined according to the following informal description of "knowledge" of a processor. The processor j "knows" φ if in every other global state which has the same local state as processor j i.e. is *knowledge-accessible* from it, the formula φ holds. In particular each processor knows its own local state. In general, the belief-accessibility relation forms a subset of the knowledge-accessibility relation, corresponding to the axiom $\mathrm{KNOW}(j, \psi) \rightarrow \mathrm{BEL}(j, \psi)$.

A *run* is a sequence of global states, which may be viewed as running through time. Time is taken as isomorphic to natural numbers, or a finite part of them.

Step 1 of the procedure

Let us first look at the low-level protocol from section 5.3.2, by which message φ is sent, bit by bit, from a to all other agents $j \in G$. This solves the sequence transmission problem in communication media where at most two kinds of errors

occur. Formally, this can be proved using the semantics of interpreted systems I (i.e. sets of runs) that are consistent with the knowledge-based protocol [13]:

Theorem 1. *Let I be an interpreted system consistent with the knowledge-based protocol given in section 5.3.2. Then every run of I has the safety property and every fair run of I has the liveness property.*

Intuitively, safety is obvious since each receiver $j \in G$ writes a data element only if it knows its value, for, by coding, all errors are detected. Assuming fairness, one can show that a eventually sends j every bit of the representation of φ, thus that every message eventually arrives and is written by the receiver. Halpern and Zack shown in [13] that the sender a establishes explicit depth 4 knowledge of the form $K_a K_j K_a K_j(x_i)$ of data element x_i before sending the next data element:

Theorem 2. *Let R be any set of runs where:*

- *the environment allows for only two kinds of errors;*
- *the safety property holds (so that at any moment the sequence Y of data elements received by j is a prefix of the infinite sequence X of data elements on a's input tape);*
- *a's state records all data elements that it has read and all acknowledgements that it has received;*
- *j's state records all the data elements it has written, for each $j \in G$, $j \neq a$.*

Then for all runs in R we have:
If a stores message "$K_j K_a K_j(x_i)$", then for all moments from then on, it holds that $K_a K_j K_a K_j(x_i)$.

In particular, after j has received the acknowledgement on the last bit of φ, we have for all moments from then on, $K_a K_j K_a K_j(x_i)$ for all bits x_i of φ. Thus at the end of step 1, for all agents $j \neq a$ in G:
KNOW(j,"the formula φ has been received"),
and a knows this in turn. To conclude from this that $\mathrm{BEL}(j, \varphi)$, we use the assumption on trust, namely that $TRUST_{\varphi}(j, a)$ holds everywhere in the run, from which $\mathrm{BEL}(j, \varphi)$ and thus $\text{E-BEL}_G(\varphi)$ follow immediately. By the assumption on persistence of beliefs, these beliefs about φ remain valid throughout the next part of the procedure, step 2.

Step 2 of the procedure
Similarly as in step 1, by the end of step 2, for all agents $j \neq a$ in G we have:
KNOW(j,"the formula $\text{C-BEL}_G(\varphi)$ has been received"),
and a knows this in turn. By assumption $TRUST_{\text{C-BEL}_G(\varphi)}(j, a)$ this leads to $\mathrm{BEL}(j, \text{C-BEL}_G(\varphi))$ for all $j \in G$, thus by the assumption on persistence of beliefs on φ, $\text{E-BEL}_G(\varphi \wedge \text{C-BEL}_G(\varphi))$, which by axiom **C2** is equivalent to $\text{C-BEL}_G(\varphi)$, as desired.

In a full proof, assumptions such as $TRUST_{\varphi}(j, a)$ should be formally operationalized.

5.4 Creating collective attitudes in rescue situations

The procedure for creating common beliefs is typically suited for situations where coordination (and thus a common belief) is needed, but there is no time to waste on communication among group members. In such a case the kind of fixed one-to-many communication (like announcement) from an initiator to the rest of a group, as exemplified by this procedure, is efficient and effective. Situations where one-to-many communication is possible, but where other members in G cannot easily reach each other, are common in rescue or emergency situations (e.g. alarm telephone hotlines).

5.4.1 Case study I: Creating common belief in a rescue situation

Let us consider a situation from the real world in which a variant of the procedure of subsection 5.3.1 would work well. Let a be the operator of an alarm telephone line SOS, that people can call in emergency situations where lives are in danger. Let b be a witness who sees that a house is on fire and assumes that there are people inside, and calls the SOS line. Now b provides a with information about the place and the nature of the disaster, represented by a formula ψ. In her turn, a calls the ambulance service A, the police P and the fire department F, giving all of them essentially the information $\psi \wedge \text{C-BEL}_G(\psi)$ (where $G = \{a, A, P, F\}$), by checking that the others receive all information accurately, and conveying that she is giving exactly the same information to the others.

In this example the role of each participant in the rescue process is clearly identified and well defined. If the rescue procedure goes well along some well established procedures there may be no need for extensive coordination between members of different services. After the common belief about the disaster is established, commonly known rescue procedures suffice for coordinated action.

However, usually life is more complex and rescue procedure requires more advanced forms of teamwork, that can adapt to a dynamic environment presenting unexpected changes. The success of the complex rescue action will depend on the successful establishment of collective motivational attitudes within the rescue team. The first step towards a goal-directed activity is creating a collective intention within a group.

5.4.2 The notion of collective intention

As multiagent systems consist of independent autonomous entities, the problem of an adequate organizational structure as well as the problem of predicting the behaviour of other agents becomes of special importance in rescue situations. When considering the rationale behind the behaviour of others, social theories about group behaviour come to the fore. In these theories collective intention towards an *action* or a *state of affairs* is a first-class citizen.

The first phase of our research concerned investigation on the sound and complete logical systems modelling a notion of *mutual intention*. Let us remind the reader of our characterization of mutual and collective intentions in cooperative teams [17].

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 $\text{INT}(i, \varphi)$ towards the overall goal φ. However, to exclude cases of competition and adversarial action, 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. Thus mutual intention is built in a way analogical to common belief.

In order to model a collective intention, the notion of agents' *awareness*, expressed in terms of any informational attitude fitting given circumstances, needs to be considered. This includes also different degrees of agents' beliefs, as discussed in section 5.2.2. This way a sort of tuning mechanism, allowing the system developer to tune not only the scope of agents' awareness, but also its strength, is provided. In [6], [7] this idea was applied in the context of collective commitment. Here, when defining collective intention, agents' awareness is expressed in the strongest way, that is in terms of common belief.

Thus, the distinguishing feature 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 common belief about this ($\text{C-BEL}_G(\text{M-INT}_G(\varphi))$. In [17], we introduce a formal definition which is extensively discussed and compared with alternatives. Above conditions are captured by the following axioms:

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

Even though $\text{C-INT}_G(\varphi)$ 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 [18]. Here we give a short description of the process, the announcement part of which may be viewed as an instantiation of the general procedure for establishing common beliefs presented in subsection 5.3.1.

5.4.3 Case study II: Creating collective intention in a rescue situation

Usually, for teamwork to emerge, it is not sufficient to share information about the present situation as does team G in Case study I: one also needs a *collective intention* to solve the problem as a team, especially if no fixed commonly known procedures are at hand.

Axiom **M3** makes evident that, in establishing a collective intention, a crucial step for the initiator is to persuade all members of a potential team to take the overall goal as an individual intention. To establish the higher levels of the mutual intention, the initiator also persuades each member to take on the intention that all members of the potential team have the mutual intention, in order to strengthen cooperation from the start. It suffices if the initiator persuades all members of a potential team G to take on an individual intention towards φ ($\mathrm{INT}(i, \varphi)$) and the intention that there be a mutual intention among that team ($\mathrm{INT}(i, \text{M-INT}_G(\varphi))$. This results in $\mathrm{INT}(i, \varphi \wedge \text{M-INT}_G(\varphi))$ for all $i \in G$, or equivalently by axiom **M1**: $\text{E-INT}_G(\varphi \wedge \text{M-INT}_G(\varphi))$, which in turn implies by axiom **M2** that $\text{M-INT}_G(\varphi)$. When all the individual motivational attitudes are established within the team, the initiator broadcasts the fact $\text{M-INT}_G(\varphi) \wedge \text{C-BEL}_G(\text{M-INT}_G(\varphi))$ by the general procedure described previously, by which the necessary common belief $\text{C-BEL}_G(\text{M-INT}_G(\varphi))$ is established and the collective intention is in place.

Thus, in the **example** case of the fire introduced in Case study I, the telephone operator a may establish a collective intention $\text{C-INT}_G(\varphi)$ among G using the above-described procedure, where φ stands for "all people in the house have been rescued and have received appropriate medical treatment". This collective intention is a trigger for more concrete planning how to achieve φ and establishing a collective commitment within the team [6], but we do not treat this further step here. The collective intention allows the team to monitor its progress towards the main goal. It also acts as a kind of glue of the team, enabling appropriate re-planning when the circumstances unexpectedly change. In [19, 20] we describe a generic algorithm for effective and efficient reconfiguration, and show how, while collective intentions persist as long as possible and needed, collective commitments evolve in appropriate ways.

5.4.4 Case study III: Creating a collective pre-attitude without communication

In some situations, for example time-critical ones, communication may be impossible, so that the low-level part of the procedure of subsection 5.3.1 does not work. In such cases there is no immediate way of establishing a common belief, except if a rescue protocol has been common knowledge (or common belief) from the start. The latter is the case for people who go on a canoeing trip with two boats: then there is common knowledge about exactly what they need to do in case one of the two boats capsizes. Let us consider less lucky situations, without pre-knowledge or communication.

We have argued in [17] that teamwork may tentatively start even if the collective intention as defined in subsection 5.4.3 has not yet been established, especially in circumstances where it has not been possible (yet) to establish a common belief among the team about their mutual intention. For such situations, as a start for teamwork we defined a notion that is somewhat stronger than the mutual intention defined in axiom **M2**: in the axiom **M2′** below, even though a common belief about the mutual intention has not been established in actual fact, all members of the group intend it to be established.

Consider, for **example**, a situation in which a person c has disappeared under the ice and two potential helpers a and b are in the neighbourhood; they do not know each other, and there is no clearly marked initiator among them. Suppose further that, at this point in time, communication among them is not possible, for example because of strong wind. Perception is possible in a limited way: they can see the other one move, but cannot distinguish facial expressions. Both have the individual intention to help (thus $\mathrm{INT}(a, \varphi)$ and $\mathrm{INT}(b, \varphi)$, i.e. $\text{E-BEL}_G(\varphi)$, where $G = a, b$ and φ stands for "c has been rescued"). Moreover, in general two persons are needed for a successful rescue, and this is a commonly believed fact ($\text{C-BEL}_G(\psi)$, where ψ stands for "at least two persons are needed to achieve φ"). As there are no other potential helpers around, a and b believe that they need to act together. Thus, we may expect that a mutual intention $\text{M-INT}_G(\varphi)$ is already established. Both agents may even form an individual belief about the mutual intention being established, so at this point there may be $\text{M-INT}_G(\varphi) \wedge \text{E-BEL}_G(\text{M-INT}_G(\varphi))$. However, communication being limited, the common belief about the mutual intention ($\text{C-BEL}_G(\text{M-INT}_G(\varphi))$) cannot be established; for this reason, the standard collective intention $\text{C-INT}_G(\varphi)$ does not hold. On the other hand, time is critical, so *some* team-like attitude needs to be established. In this situation, it is justified that goal-directed activity may be based on a revised notion of mutual intention.

In order to build a proper collective commitment, leading to team action, from the present attitude $\text{M-INT}_G(\varphi)$, the common belief is necessary. For example in the rescue situation, such a common belief enables co-ordination needed for mouth-on-mouth breathing and heart massage. Both agents believe this: they believe that if φ is ever achieved, a collective intention $\text{C-INT}_G(\varphi)$ has been established before. Thus, even if communication is severely restricted at present, they still try to establish a team together, and do both *intend* that the common belief about the mutual intention be established to make real teamwork possible.

Thus, the alternative mutual intention $\text{M-INT}'_G(\varphi)$ is meant to be true if everyone in G intends φ, everyone in G intends that everyone in G intends φ, etc. (as in $\text{M-INT}_G(\varphi)$); moreover, everyone intends that there be common belief in the group of this infinite conjunction ($\text{E-INT}_G(\text{C-BEL}_G(\text{M-INT}_G(\varphi)))$):

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

The notion of $\text{M-INT}'_G$ is appropriate for unstable situations in which communication is hard or impossible and in which a team needs to be formed. From this perspective, $\text{M-INT}'_G$ may be called a "pre-collective intention", from which the team members will in a later stage hopefully establish a common belief. This leads to an alternative definition of collective intention $\text{C-INT}'_G$ based on $\text{M-INT}'_G$. This notion is stronger than its standard counterpart C-INT_G: now, both intended and factual establishment of a common belief about the mutual intention are present. It is defined by the following axiom:

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

5.5 Discussion and conclusions

We have shown that for coordination in rescue or emergency situations, a group's informational and motivational attitudes are vital. In time-stressed or even time-critical dynamic and unpredictable environments, the communication necessary for proper coordination is crucial, but hard to achieve. Thus one needs to develop methods of creating such complex notions like common belief or collective intention on the basis of minimal communication. A procedure for establishing common belief that is applicable in rescue situation when the communication may be hampered is the main contribution in this paper.

From the time when the notion of common knowledge and common belief were first studied, there has been a puzzle about their establishment and assessment (*Mutual Knowledge Paradox* described in [21]). How can it be that to check whether one makes a felicitous reference when saying "Have you seen the movie showing at the Roxy tonight", one has to check an infinitude of facts about reciprocal knowledge, but people do this in a finite, indeed short, time? Can common knowledge (belief) be established in finite time? Even if common knowledge in general cannot be created by communication in distributed system, our somewhat devious procedure shows that common belief can.

In [22], another procedure is given for establishing shared beliefs (of the form $\text{E-BEL}_G(\text{E-BEL}_G(\varphi))$ between two agents, and it is argued that, by reasoning, these shared beliefs lead to a common belief between the agents. The authors also make use of the fact that (common) beliefs need not be true, but their protocol is much more complex than ours. We believe that their protocol is correct, but there is unfortunately a gap in their proof that $\text{E-BEL}_G\text{E-BEL}_G(\varphi)$ implies $\text{C-BEL}_G(\varphi)$, which they use to prove correctness.

The presented analysis of emergency situations may be viewed as a starting point for a more refined classification of rescue situations along the need for teamwork. On this basis, various attitudes necessary for proper organization of a rescue team activity, like collective intentions, bilateral and collective commitments, and different types of group beliefs, could be tuned to the strength of teamwork needed. As such, a resulting formal model could be viewed as a natural extension of our theory of motivational attitudes (see [17, 19, 6, 20]) realized in multimodal logics. Finally, there is room for various implementations of prototypical systems based on a the previous modelling.

Acknowledgments

We would like to thank Rohit Parikh and Marcin Dziubinski for fruitful discussion about this work. This work is partially supported by the Polish KBN Grant supporting the EU funded ALFEBIITE++ project.

References

1. Jennins, N.R., Sycara, K., Wooldridge, M.: A roadmap of agent research and development. Autonomous Agents and Multi-agent Systems **1** (1998) 7–38

2. Levesque, H., Cohen, P., Nunes, J.: On acting together. In: Proceedings Eighth National Conference on AI, AAAI-Press and MIT Press (1990) 94–99
3. Bratman, M.: Intention, Plans, and Practical Reason. Harvard University Press, Cambridge (MA) (1987)
4. Fagin, R., Halpern, J., Moses, Y., Vardi, M.: Reasoning about Knowledge. MIT Press, Cambridge, MA (1995)
5. Meyer, J.J.C., van der Hoek, W.: Epistemic Logic for AI and Theoretical Computer Science. Cambridge University Press (1995)
6. Dunin-Kęplicz, B., Verbrugge, R.: Calibrating collective commitments. In: Proceedings of the 3rd International Central and Eastern European Conference on Multi-Agent Systems. Volume 2691 of LNAI., Springer Verlag (2003) 73–83
7. Dunin-Kęplicz, B., Verbrugge, R.: A tuning machine for cooperative problem solving. Fundamenta Informaticae **to appear** (2004)
8. Steup, M.: The analysis of knowledge. In Zalta, E.N., ed.: The Stanford Encyclopedia of Philosophy. (Spring 2001)
9. van der Hoek, W., Verbrugge, R.: Epistemic logic: A survey. In Petrosjan, L., Mazalov, V., eds.: Game Theory and Applications. Nova Science Publishers, vol. 8, New York (2002) 53–94
10. Parikh, R., Krasucki, P.: Levels of knowledge in distributed computing. Sadhana: Proceedings of the Indian Academy of Sciences **17** (1992) 167–191
11. Halpern, J.Y., Moses, Y.: Knowledge and common knowledge in a distributed environment. In: Symposium on Principles of Distributed Computing. (1984) 50–61
12. van Ditmarsch, H., Kooi, B.: Unsuccessful updates. In: Proceedings of the 12th International Congress of Logic, Methodology, and Philosophy of Science (LMPS), Oviedo University Press (2003) 139–140
13. Halpern, J., Zuck, L.: A little knowledge goes a long way: Simple knowledge-based derivations and correctness proofs for a family of protocols. In: 6th ACM Symposium on Principles of Distributed Computing. (1987) 268–280
14. Stulp, F., Verbrugge, R.: A knowledge-based algorithm for the internet protocol TCP. Bulletin of Economic Research **54** (2002) 69–94
15. Castelfranchi, C., Tan, Y.H., eds.: Trust and Deception in Virtual Societies. Kluwer, Dordrecht (2001)
16. Dunin-Kęplicz, B., Verbrugge, R.: Dialogue in teamwork. In: Proceedings of The 10th ISPE International Conference on Concurrent Engineering: Research and Applications, Rotterdam, A.A. Balkema Publishers (2003) 121–128
17. Dunin-Kęplicz, B., Verbrugge, R.: Collective intentions. Fundamenta Informaticae **51(3)** (2002) 271–295
18. Dignum, F., Dunin-Kęplicz, B., Verbrugge, R.: Creating collective intention through dialogue. Logic Journal of the IGPL **9** (2001) 145–158
19. Dunin-Kęplicz, B., Verbrugge, R.: A reconfiguration algorithm for distributed problem solving. Engineering Simulation **18** (2001) 227 – 246
20. Dunin-Kęplicz, B., Verbrugge, R.: Evolution of collective commitments during teamwork. Fundamenta Informaticae **56** (2003) 329–371
21. Clark, H.H., Marshall, C.: Definite reference and mutual knowledge. In Joshi, A., Webber, B., Sag, I., eds.: Elements of Discourse Understanding, Cambridge University Press (1981) 10–63
22. Paurobally, S., Cunningham, J., Jennings, N.R.: Ensuring consistency in the joint beliefs of interacting agents. In: Proceedings of the second international joint conference on Autonomous agents and multiagent systems, ACM Press (2003) 662–669