1

Teamwork in Multi-Agent Environments

The Master doesn't talk, he acts.
When his work is done,
the people say, 'Amazing:
we did it, all by ourselves!'

Tao Te Ching (Lao-Tzu, Verse 17)

1.1 Autonomous Agents

What is an autonomous agent? Many different definitions have been making the rounds, and the understanding of agency has changed over the years. Finally, the following definition from Jennings *et al*. (1998) has become commonly accepted:

> An agent is a computer system, *situated* in some environment, that is capable of *flexible autonomous* action in order to meet its design objectives.

The environment in which agents operate and interact is usually dynamic and unpredictable.

Multi-agent systems (MASs) 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, these complex social interactions depend on the circumstances and may vary from altruistic cooperation through to open conflict. Therefore, in multi-agent systems one of the central issues is the study of how groups work, and how the technology enhancing complex interactions can be implemented. A paradigmatic example of joint activity is *teamwork*, in which a group of autonomous agents choose to work together, both in advancement of their own individual goals as well as for the good of the system as a whole. In the first phase of designing multi-agent systems in the 1980s and 1990s, the emphasis was put on

Teamwork in Multi-Agent Systems: A Formal Approach Barbara Dunin-Kęplicz and Rineke Verbrugge

cooperating teams of software agents. Nowadays there is a growing need for teams consisting of computational agents working hand in hand with humans in *multi-agent environments*. Rescue teams are a good example of combined teams consisting of robots, software agents and people (Sycara and Lewis, 2004).

1.2 Multi-Agent Environments as a Pinnacle of Interdisciplinarity

Variety is the core of multi-agent systems. This simple statement expresses the many dimensions immanent in agency. Apparently, the driving force underlying multi-agent systems is to relax the constraints of the previous generation of complex (distributed) intelligent systems in the field of knowledge-based engineering, which started from expert systems, through various types of knowledge-based systems, up to blackboard systems (Engelmore and Morgan, 1988; Gonzalez and Dankel, 1993; Stefik, 1995). Flexibility is essential for ensuring goal-directed behavior in a dynamic and unpredictable environment. Complex and adaptive patterns of interaction in multi-agent systems, together with agents' autonomy and the social structure of cooperative groups, determine the novelty and strength of the agent-based approach.

Variety is the core of multi-agent systems also because of important links with other disciplines, as witnessed by the following quote from Luck *et al*. (2003):

> A number of areas of philosophy have been influential in agent theory and design. The philosophy of beliefs and intentions, for example, led directly to the BDI model of rational agency, used to represent the internal states of an autonomous agent. Speech act theory, a branch of the philosophy of language, has been used to give semantics to the agent communication language of FIPA. Similarly, argumentation theory – the philosophy of argument and debate, which dates from the work of Aristotle – is now being used by the designers of agent interaction protocols for the design of richer languages, able to support argument and non-deductive reasoning. Issues of trust and obligations in multiagent systems have drawn on philosophical theories of delegation and norms.
>
> Social sciences: Although perhaps less developed than for economics, various links between agent technologies and the social sciences have emerged. Because multiagent systems are comprised of interacting, autonomous entities, issues of organisational design and political theory become important in their design and evaluation. Because prediction of other agents' actions may be important to an agent, sociological and legal theories of norms and group behavior are relevant, along with psychological theories of trust and persuasion. Moreover for agents acting on behalf of others (whether human or not), preference elicitation is an important issue, and so there are emerging links with marketing theory where this subject has been studied for several decades.

1.3 Why Teams of Agents?

Why cooperation?

> Cooperation matters. Many everyday tasks cannot be done at all by a single agent, and many others are done more effectively by multiple agents. Moving a very heavy object is an example of the first sort, and moving a very long (but not heavy) object can be of the second (Grant *et al*., 2005a).

Teams of agents are defined as follows (Gilbert, 2005):

> The term 'team' tends to evoke, for me, the idea of a social group dedicated to the pursuit of a particular, persisting goal: the sports team to winning, perhaps with some proviso as to how this comes about, the terrorist cell to carrying out terrorist acts, the workgroup to achieving a particular target.

Teamwork may be organized in many different ways. Bratman characterizes shared cooperative activity by the criteria of mutual responsiveness, commitment to joint activity, commitment to mutual support and formation of subplans that mesh with one another (Bratman, 1992). Along with his characteristics, the following essential aspects underlie our approach to teamwork:

- working together to achieve a common goal;
- constantly monitoring the progress of the team effort as a whole;
- helping one another when needed;
- coordinating individual actions so that they do not interfere with one another;
- communicating (partial) successes and failures if necessary for the team to succeed;
- no competition among team members with respect to achieving the common goal.

Teamwork is a highly complex matter, that can be characterized along different lines. One distinction is that teamwork can be primarily defined:

1. In terms of achieving a certain outcome, where the *roles* of agents are of prime importance.
2. In terms of the motivations of agents, where agents' *commitments* are first-class citizens.

In this book, the second point of view is taken.

1.4 The Many Flavors of Cooperation

It is useful to ask initially: what makes teamwork tick? A fair part of this book will be devoted to answering this question.

Coordinated group activity can be investigated from many different perspectives:

- the software engineering perspective (El Fallah-Seghrouchni, 1997; Jennings and Wooldridge, 2000);
- the mathematical perspective (Procaccia and Rosenschein, 2006; Shehory, 2004; Shehory and Kraus, 1998);
- the information theory perspective (Harbers *et al.*, 2008; Sierra and Debenham, 2007);
- the social psychology perspective (Castelfranchi, 1995, 2002; Castelfranchi and Falcone, 1998; Sichman and Conte, 2002);
- the strictly logical perspective (Ågotnes *et al.*, 2008; Goranko and Jamroga, 2004);
- in the context of electronic institutions (Arcos *et al.*, 2005; Dignum, 2006).

We take the *practical reasoning* perspective.

1.5 Agents with Beliefs, Goals and Intentions

Some multi-agent systems are intentional systems implementing *practical reasoning* – the everyday process of deciding, step by step, which action to perform next (Anscombe, 1957; Velleman, 2000). The intentional model of agency originates from Michael Bratman's theory of human rational choice and action (Bratman, 1987). He posits a complex interplay of informational and motivational aspects, constituting together a belief-desire-intention (BDI) model of rational agency.

Intuitively, an agent's *beliefs* correspond to information it has about the environment, including other agents. An agent's *desires* represent states of affairs (options) that it would choose. We usually use the term *goal* for this concept, but for historical reasons we use the abbreviation BDI. In human practical reasoning, *intentions* are first class citizens, as they are not reducible to beliefs and desires (Bratman, 1987). They form a rather special consistent subset of an agent's goals, that it chooses to focus on for the time being. In this way they create a screen of admissibility for the agent's further, possibly long-term, decision process called *deliberation*.

During deliberation, agents decide what state of affairs they want to achieve, based on the interaction of their beliefs, goals and intentions. The next substantial part of practical reasoning is means-ends analysis (or planning), an investigation of actions or complex plans that may best realize agents' intentions. This phase culminates in the construction of the agent's *commitment*, leading directly to action.

In this book, we view software agents from the *intentional stance* introduced by Dennett (1987) as the third level of abstraction (the first two being the physical stance and the design stance, respectively). This means that agents' behavior is explained and predicted by means of mental states such as beliefs, desires, goals, intentions and commitments. The intentional stance, although possibly less accurate in its predictions than the two more concrete stances, allows us to look closer on essential aspects of multi-agent systems. According to Dennett, it does not necessarily presuppose that the agents actually have explicit representations of mental states. In contrast, taking the computer science perspective, we will make agents' mental state representations explicit in our logical framework.

1.6 From Individuals to Groups

A logical model of an agent as an *individual, autonomous* entity has been successfully created, starting from the early 1990s (Cohen and Levesque, 1990; Rao and Georgeff, 1991; Wooldridge, 2000). These systems have been proved to be successful in real-life situations, such as Rao and Georgeff's system OASIS for air traffic control and Jennings and Bussmann's contribution to making Daimler–Chrysler production lines more efficient (Jennings and Bussmann, 2003; Rao and Georgeff, 1995a).

More recently the question how to organize agents' cooperation to allow them to achieve their common goal while striving to preserve their individual autonomy, has been extensively debated. Bacharach notes the following about individual motivations in a team setting (Gold, 2005):

> First, there are questions about motivations. Even if the very concept of a team involves a common goal, in real teams individual members often have private interests as well. Some individuals may be better motivated than others to 'play for the team' rather than for

themselves. So questions arise for members about whether other members can be trusted to try to do what is best for the team. Here team theory meets *trust* theory, and the currently hot topic of when and why it is rational to trust. Organizational psychology studies how motivations in teams are determined in part by aspects of personality, such as leadership qualities, and by phenomena belonging to the affective dimension, such as mood and 'emotional contagion'.

The intentional stance towards agents has been best reflected in the BDI model of agency. However, even though the BDI model naturally comprises agents' individual beliefs, goals and intentions, these do not suffice for teamwork. When a team is supposed to work together in a planned and coherent way, it needs to present a collective attitude over and above individual ones. Without this, sensible cooperation is impossible, as agents are not properly motivated and organized to act together as a team. Therefore, the existence of collective (or joint) motivational attitudes is a necessary condition for a loosely coupled group of agents to become a strictly cooperative team. As in this book, we focus on cooperation within strictly cooperative teams, cases of competition are explicitly excluded. Strangely enough, many attempts to define coordinated team action and associated group attitudes have neglected the aspect of ruling out competition.

1.7 Group Attitudes

The formalization of informational attitudes derives from a long tradition in philosophy and theoretical computer science. As a result of inspiring discussions in philosophical logic, different axiom systems were introduced to express various properties of the notions of knowledge and belief. The corresponding semantics naturally reflected these properties (Fagin *et al*., 1995; Hintikka, 1962; Lenzen, 1978). Informational attitudes of groups have been formalized in terms of epistemic logic (Fagin *et al*., 1995; Meyer and van der Hoek, 1995; Parikh, 2002). Along this line such advanced concepts as general, common and distributed knowledge and belief were thoroughly discussed and precisely defined in terms of agents' individual knowledge or, respectively, belief.

The situation is much more complex in case of motivational attitudes. Creating a conceptually coherent theory is challenging, since bilateral and collective notions cannot be viewed as a straightforward extension or a sort of sum total of individual ones. In order to characterize their collective flavor, additional subtle and diverse aspects of teamwork need to be isolated and then appropriately defined. While this process is far from being trivial, the research presented in this book brings new results in this respect. The complex interplay between environmental and social aspects resulting from the increasing complexity of multi-agent systems significantly contributes to this material. For example, in an attempt to answer what it means for a group of agents to be *collectively committed* to do something, both the circumstances in which the group is acting and properties of the organization it is part of, have to be taken into account. This implies the importance of differentiating the scope and strength of team-related notions. The resulting characteristics may differ significantly, and even become logically incomparable.

1.8 A Logical View on Teamwork: TeamLog

Research on a methodology of teamwork for BDI systems led us first to a static, descriptive theory of collective motivational attitudes, called TeamLog. It builds on individual goals, beliefs and intentions of cooperating agents, addressing the question what it means for a group of agents to have a *collective intention*, and then a *collective commitment* to achieve a common goal.

While investigating this issue we realized the fundamental role of collective intention in consolidating a group to a strictly cooperating team. In fact, a team is glued together by collective intention, and exists as long as this attitude holds, after which the team may disintegrate. Plan-based collective commitment leads to team action. This plan can be constructed from first principles, or, on the other extreme of a spectrum of possibilities, it may be chosen from a depository of pre-constructed plans. Both notions of collective intentions and collective commitments allow us to express the potential of strictly cooperative teams.

When building a logical model of teamwork, agents' *awareness* about the situation is essential. This notion is understood here as the state of an agent's beliefs about itself, about other agents and about the environment. When constructing collective concepts, we would like to take into account all the circumstances the agents are involved in. Various versions of group notions, based on different levels of awareness, fit different situations, depending on organizational structure, communicative and observational abilities, and so on.

Various epistemic logics and various notions of group information (from distributed belief to common knowledge) are adequate to formalize agents' awareness (Dunin-Kęplicz and Verbrugge, 2004, 2006; Fagin *et al.*, 1995; Parikh, 2002). The (rather strong) notion of *common belief* reflects ideal circumstances, where the communication media operate without failure and delay. Often, though, the environment is less than ideal, allowing only the establishment of weaker notions of group information.

1.9 Teamwork in Times of Change

Multi-agent environments by their very nature are constantly changing:

> As the computing landscape moves from a focus on the individual standalone computer system to a situation in which the real power of computers is realised through distributed, open and dynamic systems, we are faced with new technological challenges and new opportunities. The characteristics of dynamic and open environments in which, for example, heterogeneous systems must interact, span organisational boundaries, and operate effectively within rapidly changing circumstances and with dramatically increasing quantities of available information, suggest that improvements on the traditional computing models and paradigms are required. In particular, the need for some degree of autonomy, to enable components to respond dynamically to changing circumstances while trying to achieve over-arching objectives, is seen by many as fundamental (Luck *et al.*, 2003).

Regardless of the complexity of teamwork, its ultimate goal is always *team action*. Team attitudes underpin this activity, as without them proper cooperation and coordination wouldn't be possible. In TeamLog, intentions are viewed as an inspiration for goal-directed activity, reflected in the strongest motivational attitudes, that is in social (or bilateral) and collective commitments. While social commitments are related to individual actions, collective commitments pertain to plan-based team actions.

Basically, team action is nothing more than a coordinated execution of actions from the social plan by agents that have socially committed to do them. The kind of actions is not prescribed: they may vary from basic individual actions like picking up a violin, to more compound ones like carrying a piano, requiring strict coordination of the agents performing them together. In order to start team action, the underlying collective commitment should first be properly constructed in the course of teamwork. Indeed, different individual, social and collective attitudes that constitute the essential components of collective commitment have to be built carefully in a proper sequence. Our approach is based on the four-stage model of Wooldridge and Jennings (1999).

First, during *potential recognition*, an initiator recognizes potential teams that could actually realize the main goal. Then, the proper group is to be selected by him/her and constituted by establishing a collective intention between team members. This takes place during *team formation*. Finally, in the course of *plan formation*, a social plan realizing the goal is devised or chosen, and all agents agree to their shares in it, leading ultimately to collective commitment. At this point the group is ready to start *team action*. When defining these stages we abstract from particular methods and algorithms meant to realize them. Instead, the resulting team attitudes are given.

The explicit model of teamwork provided by TeamLog helps the team to monitor its performance and especially to re-plan based on the present situation. The dynamic and unpredictable environment poses the problem that team members may fail to realize their actions or that new favorable opportunities may appear. This leads to the *reconfiguration problem*: how to re-plan properly and efficiently when the situation changes during plan execution? A generic solution of this problem in BDI systems is provided by us in the *reconfiguration algorithm*, showing the phases of construction, maintenance and realization of collective commitment. In fact, the algorithm, formulated in terms of the four stages of teamwork and their complex interplay, is devised to efficiently handle the necessary re-planning, reflected in an *evolution* of collective commitment. Next to the algorithm, the dynamic logic component of TeamLog$^{\text{dyn}}$ addresses issues pertaining to adjustments in collective commitment during reconfiguration.

The static definitions from TeamLog and dynamic properties given in TeamLog$^{\text{dyn}}$ express solely vital aspects of teamwork, leaving room for case-specific extensions. Under this restriction both parts can be viewed as a set of *teamwork axioms* within a BDI framework. Thus, TeamLog formulates postulates to be fulfilled while designing the system. However, one has to realize that *any* multi-agent system has to be tailored to the application in question.

1.10 Our Agents are Planners

"Variety is the core of multi-agent systems." This saying holds also for agents' planning. In early research on multi-agent systems, successful systems such as DMARS, TouringMachines, PRS and InteRRaP were based on agents with access to plan depositories, from which they only needed to select a plan fitting the current circumstances (d'Inverno *et al.*, 1998; Ferguson, 1992; Georgeff and Lansky, 1987; Müller, 1997). The idea behind this approach was that all possible situations had to be foreseen, and procedures to tackle each of them had to be prepared in advance. These solutions appear to be quite effective in some practical situations. However, over the last few years the time has become ripe for more refined and more flexible solutions.

Taking reconfiguration seriously, agents should be equipped with planning abilities. Therefore our book focuses on the next generation of software agents, who are capable to plan from first principles. They may use contemporary planning techniques such as continual distributed planning (desJardins *et al*., 1999; Durfee, 2008). Planning capabilities are vital when dealing with real-life complex situations, such as evacuation after ecological disasters. Usually core procedures are pre-defined to handle many similar situations as a matter of routine. However, the environment may change in unpredictable ways that call for time-critical planning as addition to these pre-defined procedures. In such dynamic circumstances, a serious methodological approach to (re-)planning from first principles is necessary. Even so, ubiquitous access to complex planning techniques is still a 'song of the future'.

In this book, we aim to provide the vital methodological underpinnings for teamwork in dynamic environments.

1.11 Temporal or Dynamic?

TeamLog has been built incrementally starting from individual intentions, which we view as primitive notions, through social (bilateral) commitments, leading ultimately to collective motivational attitudes. These notions play a crucial role in practical reasoning. As they are formalized in multi-modal logics, their semantics is clear and well defined; this enables us to express many subtle aspects of teamwork like various interactions between agents and their attitudes. The static theory TeamLog has been proved sound and complete with respect to its semantics (see Chapter 3 for the proof).

Multi-agent systems only come into their own when viewed in the context of a dynamic environment. Thus, the static logic TeamLog is embedded in a richer context reflecting these dynamics. When formally modeling dynamics in logic, the choice is between dynamic logic and temporal logic. Shortly stated, in dynamic logic *actions* (or programs) are first-class citizens, while in temporal logic the *flow of time* is the basic notion (Barringer *et al*., 1986; Benthem, 1995; Benthem *et al*., 2006; Doherty and Kvarnström, 2008; Fischer and Ladner, 1979; Fisher, 1994; Harel *et al*., 2000; Mirkowska and Salwicki, 1987; Salwicki, 1970; Szałas, 1995). Both approaches have their own advantages and disadvantages, as well as proponents and detractors. Lately, the two approaches are starting to be combined and their interrelations are extensively studied, including translations from dynamic presentations into temporal ones (Benthem and Pacuit, 2006). However, the action-related flavor so typical for dynamic logic is hidden in the complex formulas resulting from the translation. Even though the solution is technically satisfying, for modeling applicable multi-agent systems it is appropriate to choose a more recognizable and explicit representation.

We choose agents, actions and plans as the prime movers of our theory, especially in the context of reconfiguration in a dynamic environment. Dynamic logic is eminently suited to represent agents, actions and plans. Thus, we choose dynamic logic on the grounds of clarity and coherence of presentation. Some aspects, such as an agent's commitment strategies, specifying in which circumstances the agent drops its commitments, can be much more naturally formalized in a temporal framework than in a dynamic one. As commitment strategies have been extensively discussed elsewhere (see, for example Dunin-Kęplicz and Verbrugge (1996); Rao and Georgeff (1991)), we shall only informally discuss them in Chapter 4. In addition, the interested reader will find a temporal framework in which our teamwork theory could be embedded in the appendix.

We are agnostic as to which of the two approaches, dynamic or temporal, is better. As Rao and Georgeff did in their version of BDI logic, one can view the semantics of the whole system as based on discrete temporal trees, branching towards the future, where the step to a consecutive node on a branch corresponds to the (successful or failing) execution of an atomic action (Rao and Georgeff, 1991, 1995b). In this view, the states are worlds at a point on a time-branch within a time-tree, so in particular, accessibility relations for individual beliefs, goals and intentions point from such a state to worlds at a (corresponding) point in time.

1.12 From Real-World Data to Teamwork

Formal approaches to multi-agent systems are concerned with equipping software agents with functionalities for reasoning and acting. The starting point of most of the existing approaches is the layer of beliefs, in the case of BDI systems extended by goals and intentions. These attitudes are usually represented in a symbolic, qualitative way. However, one should view this as an idealization. After all, agent attitudes originate from real-world data, gathered by a variety of sources at the *object level* of the system. Mostly, the data is derived from sensors responsible for perception, but also from hardware, different software platforms and last, but not least, from people observing their environment. The point is that this information is inherently quantitative. Therefore one deals with a meta-level duality: sensors provide quantitative characteristics, while reasoning tasks performed at the *meta-level* require the use of symbolic representations and inference mechanisms.

Research in this book is structured along the lines depicted in Figure 1.1. The object-level information is assumed to be summarized in queries returning Boolean values. In this way we will be able to abstract from a variety of formalisms and techniques applicable in the course of reasoning about real-world data. This abstraction is essential, since the

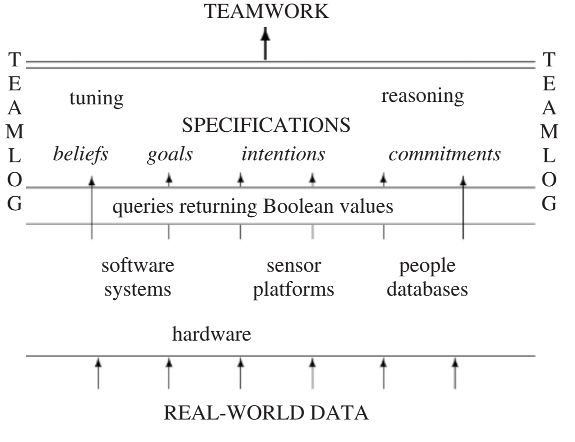

Figure 1.1 The object- and meta-level views on teamwork.

focus of this book is on the *meta-level*, including formal specification and reasoning about teamwork, as exemplified by the static and dynamic parts of TeamLog.

1.13 How Complex are Models of Teamwork?

Having a complete static logic TeamLog at hand, a natural next step is to investigate the complexity of the satisfiability problem of TeamLog, with the focus on individual and collective attitudes up to collective intention. (The addition of collective commitment does not add to the complexity of the satisfiability problem.) Our logics for teamwork are squarely multi-modal, in the sense that different operators are combined and may interfere. One might expect that such a combination is much more complex than the basic multi-agent logic with one operator, but in fact we show that this is not the case. The individual part of TeamLog is PSPACE-complete, just like the single modality case. The full system, modeling a subtle interplay between individual and group attitudes, turns out to be EXPTIME-complete, and remains so even when propositional dynamic logic is added to it.

Additionally we make a first step towards restricting the language of TeamLog in order to reduce its computational complexity. We study formulas with bounded modal depth and show that in case of the individual part of our logics, we obtain a reduction of the complexity to NPTIME-completeness. We also show that for group attitudes in TeamLog the satisfiability problem remains EXPTIME-hard, even when modal depth is bounded by 2. We also study the combination of reducing modal depth and the number of propositional atoms. We show that in both cases this allows for checking the satisfiability of the formulas in linear time.