Chapter 1

Introduction

1.1 Motivation

The more capable, autonomous, and interactive that computer systems become, the more important it is for them to have some ability to act collaboratively, whether interacting with other computer systems or within heterogeneous groups of computers and people. For example, teams of mobile rovers gathering soil samples and taking atmospheric measurements on a distant planet (Estlin et al., 2000) need to be able to coordinate their activity to avoid duplicating each other's work, to avoid interfering with one another, and to offer assistance to one another when things go wrong. Similarly, large-scale military-training simulation systems in which computer agents simulate cooperating (and competing) battlefield participants (Tambe, 1997) demand a certain level of collaborative sophistication in their simulated combatants: teams of helicopter agents must carry out a bombing mission together; when one of them gets shot down, the others may want to think about picking up any survivors. Furthermore, consumer demand for more realistic video games provides a non-trivial economic incentive to construct computer agents able to act collaboratively (Laird, 2000).

1.2 Challenges of Group Decision Making in Collaborative Planning

It has been argued that "collaboration must be designed into systems from the start; it cannot be patched on" (Grosz, 1996). Attempts to design collaborative multi-agent systems without an underlying theory of collaboration have met a range of difficulties. For example, Jennings (1995) and Tambe (1997) report such undesirable behavior as (1) agents continuing to pursue an activity even after some team members discover it is destined to fail, (2) agents abandoning a collaborative activity without bothering to inform the rest of the team, and (3) agents needlessly waiting for the results of some activity that has been abandoned or significantly delayed.

Several researchers in Artificial Intelligence (AI) have proposed formal specifications of what it means for a group of agents to collaborate on a group activity (Levesque, Cohen, and Nunes, 1990; Kinny et al., 1994; Grosz and Kraus, 1996; Grosz and Kraus, 1999). In each case, the specifications stipulate an assortment of intentions, beliefs and mutual beliefs relating to their activity that collaborating agents must hold. However, none of these specifications generally explain the underlying source of such attitudes or the ways they change over time. They provide static definitions, but do not address the dynamics of group planning. A partial exception is the SharedPlans formalization of collaborative activity (Grosz and Sidner, 1990; Grosz and Kraus, 1996; Grosz and Kraus, 1999), which specifies that collaborating agents must be committed not only to *doing* their activity together, but also to *planning* their activity together. In particular, the SharedPlans formalization stipulates that collaborating agents must be committed to group planning processes aimed at selecting a method of doing their activity, assigning tasks to members of the group, and finding suitable values for action parameters.

However, the SharedPlans formalization does not specify any such processes nor make explicit how the *results* of such processes cause the intentions, beliefs and mutual beliefs of the group members to change. In effect, the SharedPlans formalization specifies discrete snapshots of collaborative activity, but does not explicate the transitions from one such snapshot to another. As a result, it provides an incomplete specification of the dynamic function of group decision making in collaborative activity.

This thesis addresses the need for a more complete specification of the dynamic function of group decision making in collaborative activity. It describes how group decisions authorize and oblige collaborating agents to coordinate the updating of their intentions as they expand their partial plans for a collaborative activity. It also (1) provides a framework for rigorously specifying group decision-making mechanisms that agents can use to generate group decisions, (2) specifies sample mechanisms according to that framework, and (3) provides a suite of algorithms that agents can use to effectively participate in such mechanisms.

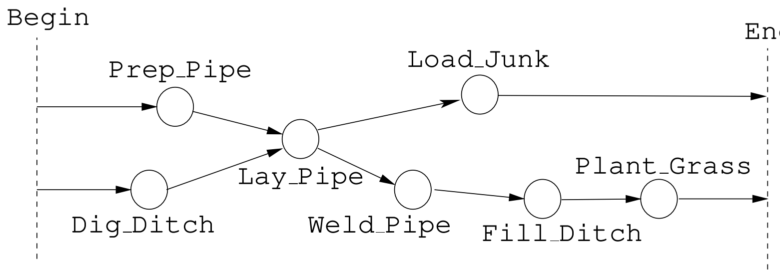
An important part of group decision making in collaborative activity centers on the problem of coordinating the times at which various tasks can be done. The second part of this thesis focuses on a range of temporal reasoning problems that agents need to be able to solve when planning their group activities. Algorithms for solving these temporal reasoning problems are derived, and their properties formally analyzed, using the Simple Temporal Networks framework (Dechter, Meiri, and Pearl, 1991).

1.3 Background and General Framework for Group Decision Making

1.3.1 Actions, Act-types and Recipes

The representation of actions, act-types and recipes used in this thesis follows that of Grosz and Kraus (1996). An action is an instance of an act-type. Act-types are either basic or complex. A basic action is an action that may be executed at will by an individual agent under appropriate conditions; a complex action is executed indirectly using a recipe. A *recipe* for a complex act-type is a set of subacts and constraints such that the doing of those subacts

Recipe: `R39 (LAY_PIPELINE)`



Subacts with Precedence Constraints

Additional Constraints

$$\begin{aligned} Agent(\texttt{Prep\_Pipe}) &= Agent(\texttt{Weld\_Pipe}) \\ Agent(\texttt{Lay\_Pipe}) &= Agent(\texttt{Load\_Junk}) \\ Agent(\texttt{Dig\_Ditch}) &= Agent(\texttt{Fill\_Ditch}) \end{aligned}$$

Figure 1.1: A sample recipe

under those constraints constitutes the doing of an action of that type. Typically, recipe constraints include precedence constraints on the execution times of the various subacts, as well as constraints that certain subacts be executed by the same agent or subgroup. As is standard in work on planning, act-type definitions specify preconditions, application constraints and necessary effects for an action of that type, as well as any parameters required in doing the action.

Figure 1.1 shows a sample recipe, `R39`, that specifies one way of doing an action of the `LAY_PIPELINE` type.[1] Precedence constraints are indicated by arrows in the figure; for example, the `Weld_Pipe` subact must be done before the `Fill_Ditch` subact. Although not shown in the figure, precedence constraints may include *offsets.* Thus, the `Load_Junk` subact might be constrained to begin no sooner than twenty minutes after the `Lay_Pipe` subact ends. Recipes may contain complex subacts; recursively choosing recipes for these subacts gives rise to a multi-level recipe hierarchy (Hunsberger, 1999). However, this thesis assumes that all recipes are fully expanded and, hence, that agents need only consider basic subacts.

[1] This thesis uses a teletype font for the names of act-types and recipes.

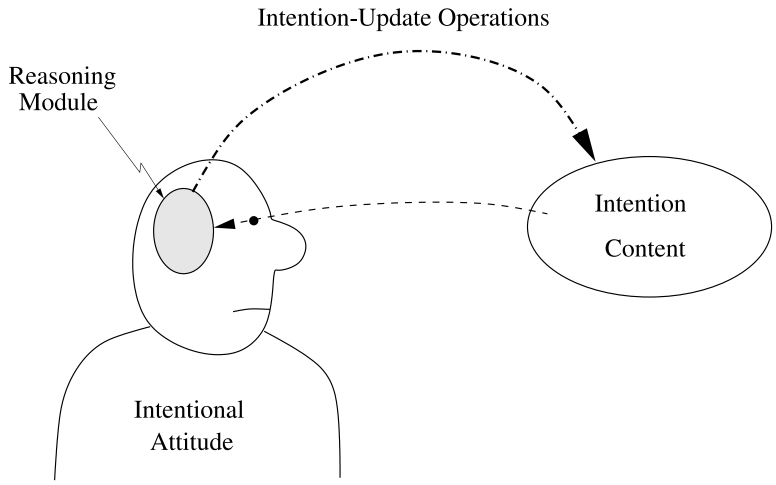


Figure 1.2: Intention cultivation in the context of single-agent activity

1.3.2 Intentions and Intention Cultivation

When an agent commits to doing some task, it typically forms a future-directed intention (Bratman, 1987). Intentions serve many purposes. They not only motivate, but also focus an agent's subsequent planning processes (Bratman, 1987). For example, Bob's intention to see a movie sometime next week might serve to focus his attention on deciding which movie he will see and whether he will drive or walk to the theater.

This thesis focuses on the class of planning decisions that agents use either to adopt new intentions or to update existing intentions. For example, given his intention to see a movie sometime next week, Bob might decide to see the movie *Notorious*, in which case he would update the content of his intention to reflect that choice; he would now intend to see *Notorious* sometime next week. Should he subsequently decide to drive to the theater, he would need to adopt a subsidiary intention (to drive to the theater). To simplify the presentation, the term *intention-update operation* will be taken to include operations for adopting new intentions as well as updating existing intentions.

In this thesis, the characteristic planning activity associated with an intention is called intention *cultivation*.[2] The process of intention cultivation is iterative: the content of an intention motivates an agent to make a planning decision, which causes it to update that intention, which motivates it to make further planning decisions, and so on. Importantly, in the context of single-agent activity, the iterative, intention-cultivation process is under the complete control of one agent, as illustrated in Figure 1.2. However, in the context of collaborative, multi-agent activity, the process of intention cultivation is more complex in several ways.

[2]Grosz and Kraus (1999) used this term to refer to the refinement of a particular class of intentions. This thesis uses the term in a more general sense.

1.3.3 The CCGI Model of the Coordinated Cultivation of Group-related Intentions

For collaborative, multi-agent activities, there are interactions among agents' individual intentions and, as a result, constraints on their intention-cultivation activity. These interactions and constraints pose several challenges for the design of collaboration-capable agents.

First, to maintain the compatibility of their group-related intentions over time, collaborating agents may not update their group-related intentions without some consensus from the group. For example, if a group of musicians are collaborating on a gig, none of them may make unilateral decisions about the type of van to rent or how much to charge for the gig; such decisions must be agreed to, or *authorized* by, the group. In the model presented in this thesis, the *coordinated-cultivation requirement* (CCR) explicitly represents this general prohibition against unilateral planning decisions in the context of collaborative activity.

Second, a group planning decision (e.g., to rent a luxurious van for transporting their musical equipment) not only authorizes the agents to update their group-related intentions, but also compels them to do so. In making an agreement, agents incur an obligation to update their intentions accordingly. This thesis explicitly represents the obligations that ensue from group decisions.

Third, agents require a means of generating group decisions. Although people working in small groups can be very adept at using casual methods of generating group decisions, larger groups benefit from structured decision-making mechanisms. For example, the people of the United States engage in a highly structured group decision-making process once every four years to decide which of them shall be the next president. This thesis provides a framework for rigorously specifying group decision-making mechanisms that computer agents can use to generate group decisions. The group decisions generated by any such mechanism must be mutually believed by the group; such mechanisms must also establish the obligations that result from the decisions they generate.

Fourth, since the decisions generated by a group decision-making mechanism typically subject agents to various obligations, agents must be able to reason about the obligations that might ensue from their participation in such a mechanism. For example, an agent bidding to do a set of tasks must be able to determine whether its bid might conflict with any of its pre-existing commitments. This thesis provides a set of algorithms that agents can use to solve the sorts of temporal reasoning problems that arise from the use of group decision-making mechanisms. Each algorithm solves a temporal reasoning problem that is formally defined and analyzed using the Simple Temporal Networks framework (Dechter, Meiri, and Pearl, 1991).

Figure 1.3 illustrates the model of the coordinated cultivation of group-related intentions in the context of collaborative activity (the CCGI model) proposed by this thesis. This model includes a specification of the coordinated-cultivation requirement that prohibits the unilateral updating of group-related intentions, depicted in Figure 1.3 as a clamp on the ability of agents to update their group-related intentions. The CCGI model also includes mechanisms for generating group decisions that both authorize and oblige agents to update their group-related intentions. The authorization and obligation resulting from such group

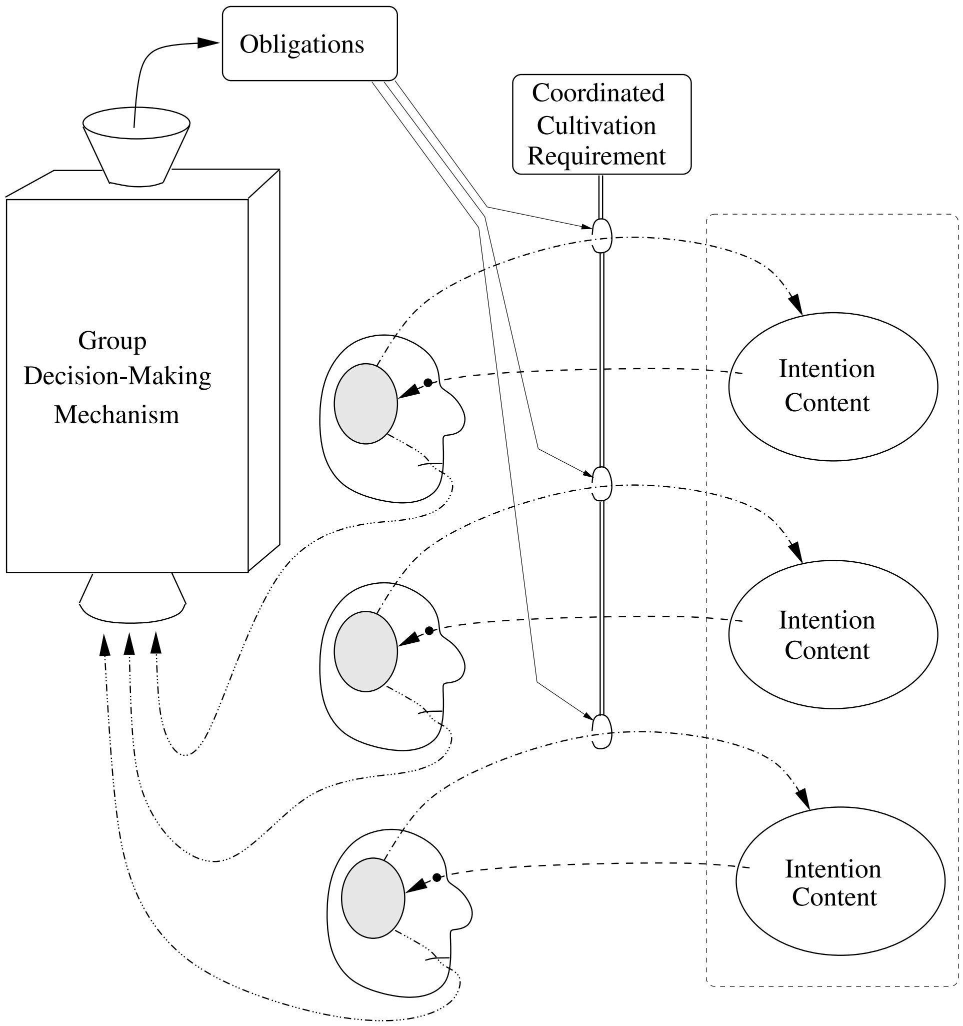


Figure 1.3: The CCGI model of the coordinated cultivation of group-related intentions in collaborative, multi-agent activity

decisions are depicted in Figure 1.3 by arrows emanating from the "obligations" box. Finally, as in the individual case, the content of an agent's group-related intention motivates it to update its intention; however, given the coordinated-cultivation requirement, the agent must seek the group's authorization for any intention-update operations it might want to perform. Thus, an agent's intention motivates it to participate in group decision-making mechanisms, as depicted by arrows from agents to the GDMM box in Figure 1.3.

1.3.4 Related Prior Work in Philosophy and AI

The CCGI model of the coordinated cultivation of group-related intentions described above succinctly captures many of the requirements of collaborative activity highlighted by a variety of researchers in philosophy and AI. In addition, the CCGI model provides a unifying framework for the different commitments to group planning processes specified by the SharedPlans formalization (Grosz and Kraus, 1996; Grosz and Kraus, 1999), as well as a more complete specification of the function of group decision making in the context of collaborative activity than any prior formalization.

Comparison to Approaches in Philosophy

Bratman (1999) argues that the intentions existing within the minds of agents are of one fundamental kind, whether in the context of single-agent or collaborative, multi-agent activity; that it is solely the *content* of agent intentions that distinguishes the two cases, as illustrated by the difference between my intending *to play my guitar by myself* versus my intending *that we paint the house together*. However, in his analysis of "We intend to J", Bratman requires not only that each of us "intend that we J", but also that each of us "intend that we J in accordance with and because of [our intentions that we J] and meshing subplans [arising from our intentions that we J]."

In contrast to Bratman, Searle (1990) argues that intentions in single-agent and collaborative, multi-agent activity cannot be distinguished solely by their differing content. Instead, he claims that a fundamentally different kind of intention, called a *we-intention*, is at work in collaborative activity. According to Searle, a we-intention has the form "*We* intend that we perform act A." Although we-intentions exist in the minds of individual agents, Searle claims that "the reference to the collective lies outside the bracket that specifies the propositional content of the intentional state." Part of Searle's argument that a web of I-intentions is insufficient to generate "collective intentionality" is that "the notion of a we-intention, of collective intention, implies the notion of *cooperation*."

Tuomela (1995) argues that collaborative groups require what he calls a *socially-existing authority system* (SEAS) and that members of a group "submit their will to a group will" with respect to a range of issues related to their group activity. He also observes that "for an SEAS to do the work it is supposed to do, the members of [a group] must have some motivation to use it."

Gilbert (2000) argues that shared intentions are distinguished by a group's joint commitment to "intending as a body."

In the CCGI model, the coordinated-cultivation requirement forces me to seek your approval for planning decisions concerning our doing of J; the model thus provides a mechanism that accounts for how I can intend that we J in accordance with *your* intention that we J. The CCGI model also explains how a web of I-intentions, suitably constrained by the coordinated-cultivation requirement, motivates agents to participate in group decision-making mechanisms and, hence, to cooperate and to plan their activity together. In that sense, it accommodates the notion of "group will".

As will be seen in Chapter 2, the CCGI model includes a framework for rigorously specifying group decision-making mechanisms. Furthermore, it explains the motivation of collaborating agents to *use* such mechanisms, as well as the ways agents can intend "as a body" (i.e., coordinate the cultivation of their group-related intentions).

Comparison to Other Approaches in AI

The SharedPlans formalization of collaborative activity (Grosz and Kraus, 1996; Grosz and Kraus, 1999) stipulates that collaborating agents must be committed to selecting a recipe for their activity, to assigning tasks to agents or subgroups, and to finding suitable values for action parameters. The CCGI model provides a unifying framework for this network of commitments to planning processes by identifying and highlighting their common roots in the need for the coordinated cultivation of group-related intentions. In addition, the CCGI model distinguishes the motivation for participating in group decision-making mechanisms from the specification of the various decision problems that collaborating agents need to address. The motivation for participating in group decision-making mechanisms arises both from agent intentions and the coordinated-cultivation requirement; the various decision problems that agents must address arise from the content of their intentions. Finally, by explicitly representing group decisions and the resulting obligations, the CCGI model clarifies how group decisions cause agent intentions to change over time as a group expands its partial plan.

Levesque, Cohen and Nunes (1990), building on earlier work by Cohen and Levesque (1990), focus on the persistence of agent intentions in collaborative activity. However, as observed by Singh (1992), their model does not adequately provide for agents being motivated to eventually do something to achieve their goals.

In the context of an alternative account of joint intentions, Kinny et al. (1994) provide algorithms agents can use to agree upon a completely specified plan, including a complete allocation of tasks to agents, at the time a team is formed for a proposed group activity. The members of the team establish the corresponding network of intentions, beliefs and mutual beliefs in a single, synchronized act. Like SharedPlans, the CCGI model generalizes the approach of Kinny et al. along several dimensions; for instance, recipes for group activities can be partial and agents can incrementally expand their partial plans. In addition, the CCGI model provides a more general explication of the process whereby group decision-making mechanisms lead to the establishment and subsequent modification of agent intentions over time.

1.4 Road Map for the Remainder of the Thesis

Chapter 2 presents a framework for rigorously specifying group decision-making mechanisms that agents can use to generate group decisions to coordinate the cultivation of their group-related intentions, as specified in the CCGI model. The properties of a sample, proposal-based mechanism are formally analyzed to show that, under a specified set of assumptions about agent beliefs, group decision-making mechanisms properly associate group decisions with the intention updates that they authorize. The coordinated-cultivation requirement is also formally defined in this chapter.

The remainder of the thesis focuses on the problem of how agents, when faced with a proposed group activity, can determine, both individually and collectively, whether to commit to that activity, a problem this thesis refers to as the Initial-Commitment Decision Problem (ICDP). Chapter 3 specifies a group decision-making mechanism based on a combinatorial auction in which agents bid on sets of tasks in a proposed group activity. To enable agents to protect the feasibility of their private schedules of pre-existing commitments, agents are allowed to condition their bids on temporal constraints. Thus, to participate in an ICDP auction, agents must be able to solve a variety of temporal reasoning problems.

Chapter 5 formally treats an important family of temporal reasoning problems that are directly related to the auction-based mechanism. Each problem is formally defined and analyzed using the Simple Temporal Networks framework (Dechter, Meiri, and Pearl, 1991). The existing theory of Simple Temporal Networks is summarized in Chapter 4. In addition, that chapter provides several useful extensions to the theory. Chapter 6 contains concluding remarks.