Computational Theories of Collaboration

Ph.D. Comprehensive Exam

Mohammad Shayganfar - mshayganfar@wpi.edu

May, 26 2015

1 Introduction to Collaboration Theories

The construction of computer systems that are intelligent, collaborative problem-solving partners is an important goal for both the science of AI and its application. There is no question that there is a need to make computer systems better at helping us do what we use them to do. collaboration must be designed into systems from the start; it cannot be patched on. To build collaborative sys- tems, we need to identify the capabilities that must be added to individual agents so that they can work with other agents [15].

Most collaborative situations involve agents who have different beliefs and capabilities. Partial knowledge is the rule, not the exception. As a result, collaborative planning requires an ability to ascribe beliefs and intentions.

Another feature of collaborative activity is collaborative plans are not simply the sum of individual plans.

To be collaborative, partners, e.g., a robot and a human, need to meet the specifications stipulated by some theories that we review in this document. As we discuss in Section 5.2, collborators need to commit to the group activity and to their role in it; they need to divide the task load according to their capabilities so they can carry out the individual plans that constitute the group activity; and they need to commit to the success of others. Collaborators also need to be able to communicate with others effectively, and to interpret others' actions and utterances in the collaboration context. Furthermore, collaborators need to be willing to help others in doing their own tasks, and to reconcile between commitments to existing collabortion and their other activities [16].

Collaboration is a special type of coordinated activity in which the participants work jointly, together performing a task or carrying out the activities needed to satisfy a shared goal [20].

Existing collaboration theories (including SharedPlans) consider the nature of a collaboration to be more than a set of individual acts. These theories argue for an essential distinction between a collaboration and a simple interaction or even a coordination in terms of commitments [14, 37].

2 Shared Activity

Joint action can best be described as doing something as a team where the participants share the same goal and a common plan of execution. Grosz and Sidner has pointed out that collaborative plans do not reduce to the sum of the individual plans [22] [15].

For example, if we were to move a table jointly through a doorway, your picking up one side of the table and starting to walk through the door does not make sense outside our joint activity. Even the sum of both our picking-up and moving actions would not amount to the shared activity without the existence of a collaborative plan that both of us are sharing (namely to move the table out the door). It seems that we both hold a joint intention, as well as individual intentions related to this joint intention.

3 Communication

Cohens Joint Intention Theory predicts that an efficient and robust collaboration scheme in a changing environment requires an open channel of communication. Sharing information through communication acts is critical given that each teammate often has only partial knowledge relevant to solving the problem, different capabilities, and possibly diverging beliefs about the state of the task.

4 Teamwork & Collaboration

Bratman's view is that collective intention can be described by referring to individual intention in combination with other mental attitudes. Searle's opposing view is that collective intention cannot be so reduced.

According to Joint Intentions theory, the notion of teamwork is characterized by joint commitment (also known as joint persistent goal or JPG). The definition of JPG states that the agents mutually believe they have the appropriate goal, and that they mutually believe a persistent weak achievement goal (which represents the one-way commitment of one agent directed towards another) to achieve it persists until the agents mutually believe that the goal has either been achieved, impossible, or irrelevant.

5 Computational Theories of Collaboration

There are prominent collaboration theories that are mostly based on plans and often analysis of the discourse between collaborators revolving around these plans [22, 36]. In these theories the discourse analysis is based on search over these tree plans [50].

The two theories Joint Intentions and SharedPlans have been extensively used to examine and describe teamwork.

Two important theories for modeling teamwork collaboration were derived from the BDI paradigm.

5.1 Theory of Joint Intentions/Teamwork

1. Mutual belief in the joint intention

2. Joint execution until MB in goal termination

3. Termination:

- Goal achieved
- Goal unachievable
- Goal irrelevant

[Use my own example] Supporting Bratmans guidelines, Cohen and Levesque propose a formal approach to building artificial collaborative agents. Their notion of joint intention is viewed not only as a persistent commitment of the team to a shared goal, but also implies a commitment on part of all its members to a mutual belief about the state of the goal. Teammates are committed to inform the team when they reach the conclusion that a goal is achievable, impossible, or irrelevant. In our table-moving example, if one team member reaches the conclusion that the table will not fit through the doorway, it is an essential part of the implicit collaborative contract, to have an intention to make this knowledge common. In a collaboration, agents can count on the commitment of other members, first to the goal and thenif necessaryto the mutual belief of the status of the goal.

A joint intention of a team Θ is based on its joint commitment, which is defined as a *Joint Persistent Goal* (JPG). A JPG to achieve a team action p, denoted JPG(Θ, p) requires all team members to mutually believe that p is currently false and want p to be eventually true. A JPG guarantees that team members cannot decommit until p is mutually known to be *achieved, unachievable* or *irrelevant.* Basically, JPG(Θ, p) requires team members to each hold p as a *Weak Achievement Goal* (WAG). WAG(μ, p, Θ), where μ is a team member in Θ, requires p to achieve p if it is false. However, if p privately believes that p is either achieved, unachievable or irrelevant, JPG(Θ,

p) is dissolved, but p is left with a commitment to have this belief become Θ's mutual belief. Such a commitment is required to establish mutual belief in Θ that an agent must typically communicate with its teammates.

The Joint Intentions Framework [35] [9] [10] is a theoretical framework founded on BDI logics. The framework focuses on a team's joint mental state, called a joint intention. A team jointly intends a team action if all team members are jointly committed to perform an action while in a specified mental state.

According to this theory, joint action is conceptualized as doing something together as a team where the teammates share the same goal and the same plan of execution. Sharing information through communication acts is critical given that each teammate often has only partial knowledge relevant to solving the problem, different capabilities, and possibly diverging beliefs about the state of the task. Communication plays an important role in coordinating their roles and actions to accomplish the task. It also serves to establish and maintain a set of mutual beliefs (also called common ground) among the team members. For instance, all teammates need to establish and maintain a set of mutual beliefs regarding the current state of the task, the respective roles and capabilities of each member, the responsibilities of each teammate, etc. What happens when things go wrong? Teammates must share a commitment to achieving the shared goal. They cannot abandon their efforts, but must instead continue to coordinate their efforts to try a different, mutually agreed upon plan. [–¿ Leonardo]

Joint Intentions Theory describes how a team of agents can jointly act together by sharing mental states about their actions. An intention is viewed as a commitment to perform an action while in a mental state. And a joint intention is a shared commitment to perform an action while in a group mental state [9]. Communication is required to establish and maintain mutual beliefs and joint intentions. A team of agents jointly intend to perform an action if and only if the members have a joint persistent goal [8].

In order to enter a joint commitment, the team members have to establish appropriate mutual beliefs and individual commitments. Although the Joint Intentions Theory does not mandate communication and several techniques are available to establish mutual beliefs about actions from observations (see [26]), currently communication seems to be the only feasible way to attain joint commitments. A key aspect of the Joint Intention Theory is the commitment to attain mutual belief about the termination of a team action. This helps to ensure that the team stays updated about the status of the team actions. This behaviour is achieved by enforcing that agents committing to a joint intention also commit to inform their team about any relevant failures or premature terminations. Joint intentions and joint commitments provide a basic framework to reason about coordination required for teamwork as well as guidance for monitoring and maintaining team activities. However, a single joint intention for a high-level team goal is not sufficient to model team behaviour in detail and to ensure coherent teamwork.

There are also some other theories with similarities and contrasts conveying collaboration concepts including Cohen and Levesque's work describing the concept of *joint intentions* in [8, 35].

In [8] Cohen and Levesque establish that joint intention cannot be defined simply as individual intention with the team regarded as an individual. This is because after the initial formation of an intention, team members may diverge in their beliefs and hence in their attitudes towards the intention. Instead, Cohen and Levesque generalise their own definition of intention. First they present a definition of individual persistent goal and, in terms of this, individual intention. Both definitions use the notion of individual belief. Next, they define precise analogues of these concepts – joint persistent goal and joint intention – by invoking mutual belief in place of individual belief. The definition of joint persistent goal additionally requires each team member to commit to informing other members – to the extent of the team's mutual belief – if it comes to believe that the common goal has been achieved, becomes impossible or is no longer relevant. The result is that, while a team is not an individual, nevertheless joint intention is similar to individual intention. In Cohen and Levesque's theory, then, a team with a joint intention is a group that shares a common objective and a certain shared mental state. In particular, joint intentions are held by the team as a whole [29].

5.2 SharedPlans Theory

5.2.1 Full Vs. Partial Shared Plan

A shared plan is either a *Full Shared Plan* (FSP) or a *Partial Shared Plan* (PSP). An FSP to do α represents a situation where every aspect of a joint activity α is fully determined. This includes mutual belief and agreement in the complete recipe to do α. Recipe is a specication of a set of actions A_i, which constitutes performance of α when executed under speci ed constraints. *FSP(*$\boldsymbol{P}$*,* $\Theta, \alpha, T_p, T_\alpha, \boldsymbol{R}_\alpha$*)* denotes a group Θ's plan $\boldsymbol{P}$ at time T_p to do action α at time T_α using recipe $\boldsymbol{R}_\alpha$. In short, *FSP* holds if and only if the following conditions are satisfied:

1. All members of group Θ mutually believe that they intend to do α.

2. All members of group Θ mutually believe that $\boldsymbol{R}_\alpha$ is the recipe for α.

3. For each step A_i in recipe $\boldsymbol{R}_\alpha$:

   - A subgroup Θ_j has an *FSP* for A_i, using recipe $\boldsymbol{R}_{A_i}$.
   - Other members of group Θ believe that there exists a recipe such that subgroup Θ_j can bring about A_i and have an FSP for A_i.
   - Other members of group Θ intend that subgroup Θ_j can bring about A_i using some recipe.

Most of the times a team and its members do not possess an *FSP* to achieve their shared goal. In this case, the concept of *FSP* put limits on the SharedPlans theory. However, the SharedPlans theory provides the concept of *Partial Shared Plan* (PSP) which since a team might only have the plan partially. The PSP is a snapshot of the team's mental state in a particular situation in their teamwork, and further communication and planning is often used to fulfill the conditions of an FSP. Notice that for the same reason recipes can be partial.

5.2.2 Communicating Intentions

Using discourse plans can help to encode the knowledge about conversation.

The SharedPlans theory recognises three interrelated levels of discourse structure.

In [22], Grosz and Sidner argue that the components of the discourse structure are a trichotomy of linguistic structure, intentions structure and the attention state. In their work, the linguistic structure of a discourse is a sequence of utterances aggregating into discourse segments just as the words in a single sentence form constituent phrases. They also discuss the idea of the discourse purpose as the intention that underlies engagement in the particular discourse. They believe this intention is the reason behind performing a discourse rather than some other actions, and also the reason behind conveying a particular content of the discourse rather than some other contents. They describe mechanisms for plan analysis looking at Discourse Segment Purposes (DSPs). In fact, the DSPs specify how the discourse segments contribute to achieving the overall discourse purpose. Finally, the third component in their theory, the attentional state, provides an abstraction of the agent's focus of attention as the discourse unfolds.

The focusing structure contains DSPs and the stacking of focus spaces reflects the relative salience of the entities in each space during the discourse. In short, the focusing structure is the central repository for the contextual content required for processing utterances during the discourse [22].

5.2.3 Collaboration Vs. Sum of Coordinate Actions

[15]

5.2.4 SharedPlans

Grosz and Sidner (1996) propose that collaboration must have three elements:

1. the participants must have commitment to the shared activity,
2. there must be a process for reaching an agreement on a recipe for the group action, and
3. there must be commitment to the constituent actions.

The SharedPlans formalization distinguishes partial plans and complete plans. A Full SharedPlan (FSP) is a complete plan in which agents have fully determined how they will perform an action. The Partial SharedPlan (PSP) definition provides a specification of the minimal mental-state requirements for collaboration to exist and gives criteria governing the process of completing the plan. Consequently, a full recipe for doing an action A is a set of actions and constraints such that the doing of those actions under those constraints constitutes the doing of action A. A partial recipe is a set of actions and constraints that may be extended to a full recipe [19].

Grosz and Sidner [22] Grosz and Kraus [20]

SharedPlans is a general theory of collaborative planning that requires no notion of joint intentions, accommodates multi-level action decomposition hierarchies and allows the process of expanding and elaborating partial plans into full plans.

SharedPlans is rooted in the observation that collaborative plans are not simply a collection of individual plans, but rather a tight interleaving of mutual beliefs and intentions of different team members.

The SharedPlans model of collaborative action [19] [20] [22] aims to provide the theoretical foundations needed for building collaborative robots/agents [15]. It specifies four key characteristics for participants in a group activity to be collaborative partners, and thus for their joint activity to be collaborative. The SharedPlans definition states that for a group activity to be collaborative, the collaborators must have:

Look at this carefully!!!

a) individual intentions that the group perform the group activity;

b) mutual belief of a (partial or complete) recipe;

c) individual or group plans for the constituent subactions of the recipe;

d) intentions that their collaborators succeed in doing the constituent subactions.

In other words, to successfully complete a plan the collaborators must mutually believe that they have a common goal and have agreed on a sequence of actions for achieving that goal. They should believe that they are both capable of performing their own actions and intend to perform those actions while they are committed to the success of their plans.

The intentions that this definition specifies constitute different kinds of commitments required of the collaborators.

The idea behind partial shared plans is enabling the agents to modify the shared plan over the course of planning without impairing the achievement of the shared goals.

5.2.5 Intention-to and Intention-that

In Grosz and Sidner's SharedPlans theory [22], two intentional attitudes are employed: "intending to" (do an action) and "intending that" (a proposition will hold). The notion of *intention to*, as an individual-oriented intention, models the intention of an agent to do any single-agent action while the agent not only believes that it is able to execute that action, but it also committs to doing so. In short, it is an intention to perform an action, similar to Bratmans conception. In contrast with *intention to*, an *intention that*, as the notion of an intention directed toward group activity, does not directly imply an action. In fact, an individual agent's *intention that* is directed towards its collaborator's action or towards a group's joint action. *Intention that* guides an agent to take actions (including the communication), that enable or facilitate other collaborators to perform assigned tasks. This leads an agent to behave in a manner consistent with a collaborative effort. Therefore, agents will adopt intentions to communicate about the plan [20]. There is a significant difference between "Intention that" and "Intention to". "Intention to" commits an agent to means-end reasoning and acting [3]. In contrast, "Intention that" does not necessarily entail this commitment.

5.2.6 Recipes

The SharedPlans definition of mutual beliefs states that when agents have a shared plan for doing some act, they must hold mutual beliefs about the way in which to perform that act. Following Pollack [45], the term recipe refers to what collaborators know when they know a way of doing something. Recipes are specified at a particular level of detail. Hence, the agents need to have mutual beliefs about acts specified at the particular level of detail of the recipe, and they do not to have mutual beliefs about all levels of acts that each agent will perform. Mutual belief of the recipe essentially means that all the collaborators hold the same beliefs about the way in which the activity will be accomplished. Therefore, the collaborators must agree on how to do the activity. Grosaz and Sidner in their earlier work [22] have considered only simple recipes in which each recipe consisted of only a single act-type relation [38]. Recipes are aggregations of act-types and relations among them. Act-types, rather than actions, are the main elements in recipes. Recipes can be partial, meaning thay can expand and be modified over time.

5.2.7 Plans

Figure 1 shows what we need to add to individual plans in order to have plans for group action, and lists the major components of group plans; to provide a base for comparison, the top of the figure lists the main components for individual plans. First, just as an individual agent must know how to do an action, agents in a group must have knowledge of how theyre going to do an action. In the case of a group plan to do a joint activity, there must be mutual belief of the recipe; agents must agree on how they are going to do the action. Then, just as individual agents must have the ability to perform the constituent actions in an individual plan and must have intentions to perform them, the participants in a group activity must have individual or group plans for each of the constituent actions in their agreed-on recipe [22] [15].

Plans for group actions include two major constituents that do not have correlates in the individual plan. First, the agents must have a commitment to the group activity; they must all intend-that the group will do the action. For instance, a robot and an astronaut need to have intentions that they install solar panels. Among other things, these intentions will keep them both working on the panels until the panels are installed. Second, the participants must have some commitment to the other agents being able

PLANS FOR COLLABORATIVE ACTION

- To have an individual plan for an act, need
  - knowledge of a recipe
  - ability to perform subacts in the recipe
  - intentions to do the subacts
- To have a group plan for an act, need
  - mutual belief of a recipe
  - individual or group plans for the subacts
  - intentions that group perform act
  - intentions that collaborators succeed

Figure 1: Plans for collaborative action [15].

to do their actions. For example, the robot must have an intention that the astronaut be able to measure the quality of installation. This intention will prevent the robot from interrupting astronaut's measurement action or prevent the robot from using astronaut's measurement tool [22] [15].

5.2.8 Commitment

Teamwork is:
Mutual commitment to joint activity:
- Agreement on the joint activity
- Cannot abomdom activity without involving teammates
Mutual suppport:
- Must be active in helping teammate activity
Mutual Responsiveness:
- Take over tasks from teammates if necessary
Joint Intentions theory is expressed in a modal language.
Individual Commitment:(Joint Intentions)
PGOAL (Persistent Goal) to achieve p relative to q
1. she believes that p is currently false.
2. she wants p to be true eventually.
3. 2. will continue to hold until she comes to believe that p is true, or will never be true, or that q is false (q is irrelevance clause for the case that

p becomes unnecessary).

Joint Commitment: (Joint Intentions)

JPG (Joint Persistent Goal) to achieve p relative to q

1. they mutually believe that p is currently false.

2. they mutually believe that they all want p to be true eventually.

3. until they come to mutually believe that p is true, or will never be true, or that q is false, they will continue to mutually believe that they each have p as a AGOAL (Achievement Goal)

Joint Intention:

A Join Intention of team t to achieve p relative to q ts members have a JPG to achieve p relative to q + ts members are believing throughout that they are doing it.

(a joint intention to perform a particular action is a joint commitment to enter a future state wherein the agents mutually believe the collaborative action is imminent just before they perform it)

Theorem:

If a team jointly intends to do a complex action consisting of team members concurrently doing individual actions, then the individuals will privately intend to do their share relative to the joint intention.

The notion of an agents commitment to achieving some state in the world is expressed as a persistent goal or PGOAL in Joint Intentions Theory [9]. An agent x has a persistent goal (PGOAL x p q) if x wants p to become true and cannot give up the goal until it believes that p is accomplished, or is impossible, or is irrelevant. An intention to do an action is defined as a persistent goal in which the agent is committed to performing the action believing throughout that it is doing that action.

Bratman in [3] has argued that intentions noramlly pose the problems for the agent and agents needs to determine an action to achieve those intentions. He also argued that intentions constrain an agents choice of what else it can intend, and they provide the context for an agents replanning when something goes wrong. The commitment to joint activity leads to a need to communicate.

Finally, the definition of the overall plan in terms of constituent plans of individuals or groups is recursive, with the recursion ending at the level of basic, individual actions [16].

Grosz, Sidner and Lochbaum in [22] and [38] present a model of plans to account for how agents with partial knowledge collaborate in the construction of a domain plan. Agents have a library of partially speci ed plan schemas (recipes). These recipes might be underspeci ed as to how an action is executed or how an action contributes to a goal. Agents then collaborate in constructing a shared plan by uttering statements about their beliefs and intentions about the plan. This collaboration will terminate with each agent mutually believing that each act in the plan can be executed by one of the agents, that that agent intends to perform the act, and that each act in the plan contributes to the goal.

Grosz, Sidner and Lochbaum in [22] and [38] are interested in the type of plans that underlie discourse in which the agents are collaborating in order to achieve a shared goal. They propose that agents are building a shared plan in which participants have a collection of beliefs and intentions about the actions in the plan.

Grosz, Sidner and Lochbaum in [22] and [38] model how several agents with partial knowledge collaborate on constructing a shared domain plan . Each agent communicates their beliefs and intentions by making utterances about what actions they can contribute to the shared plan. Collaboration is again modelled by the agents establishing a mutual belief that each action in the shared plan contributes to the goal of the plan, and that each action can and will be performed by one of the agents.

Shared plan is another essential concept in the collaboration context. The definition of the shared plan is derived from the definition of plans Pollack introduced in [44, 45] since it rests on a detailed treatment of the relations among actions and it distinguishes the intentions and beliefs of an agent about those actions. However, since Pollack's plan model is just a simple plan of a single agent, Grosz and Sidner extended that to plans of two or more collaborative agents. The concept of the shared plan provides a framework in which to further evaluate and explore the roles that particular beliefs and intentions play in collaborative activity [38]. However, this formulation of shared plans (a) could only deal with activities that directly decomposed into single-agent actions, (b) did not address the requirement for the commitment of the agents to their joint activities, and (c) did not adequately deal with agents having partial recipes [20]. Grosz and Kraus in [20], reformulate Pollack's definition of the individual plans [45], and also revise and expand the SharedPlans to address these shortcomings.

5.2.9 Coordinated Cultivation of SharedPlans

Grosz and Hunsberger [17] claim to reconcile the two approaches. They provide the "Coordinated Cultivation of SharedPlans" (CCSP) model, which, while relying solely on individual intention, captures the essential properties argued for in accounts that require group-oriented intention. CCSP also provides a general architecture for collaboration-capable agents.

5.3 Hybrid Collaboration Approaches

5.3.1 Hybrid Collaboration Approaches

Tambe in [59] argues that teamwork in complex, dynamic, multi-agent domains requires the agent to obtains flexibility and reusability by using integrated capabilities. Tambe created STEAM based on this idea. STEAM's operationalization in complex, real-world domains is the key in its development to address important teamwork issues some of which are discussed in Section 8. STEAM is founded on the Joint Intentions theory and it uses joint intentions as the basic building block of teamwork. According to Tambe's claim, several advantages accrue due to the this use of Joint Intentions theory, such as achieving a principled framework for reasoning about coordination and communication in a team which the joint intention can provide. Or, the guidance for monitoring and maintenance of a team activity which again the joint commitment in joint intention provides. Ans lastly, Tambe believes the joint intention in a team can facilitate reasoning about team activity and team member's contribution to that activity. However, he believes for a high level team goal, one single joint intention is not sufficient to achieve all these advantages. Thus, STEAM borrows some of the concepts of SharedPlans theory. First, STEAM uses the concept of "intention that" (see Section 5.2) towards an activity as well as the fact that SharedPlans theory mandates team members' mutual belief in a common recipe and shared plans for individual steps in the common recipe. Thus, in this case, SharedPlans helps STEAM to achieve coherency of within the teamwork. However, STEAM uses joint intentions as a building block to ensure the teamwork coherency to build mental attitudes of team members. In other words, in STEAM as the recipe evolves, STEAM requires all team members to agree on execution of a step and form joint intentions to execute it while other joint intentions are formed, leading to a hierarchy. Second,

However, Tambe added more practical concepts into the STEAM's architecture. For instance, STEAM has team synchronization protocol to establish joint intention (see JPG in Section 5.1), or it has constructs for monitoring joint intentions which helps the agent to be able to monitor team performance. STEAM facilitates this monitoring by exploiting its explicit representation of team goals and plans. In particular, STEAM allows an explicit specification of monitoring conditions to determine achievement, unachievability or irrelevancy conditions of team operators. Finally, in STEAM, communication is driven by commitments embodied in the joint intentions theory, i.e., team members may communicate to obtain mutual belief while building and disbanding joint intentions. Thus, joint intentions provide STEAM a principled framework for reasoning about communication. Also, STEAM addresses some practical issues, not addressed in other teamwork theories. One of these issues is STEAM's detailed attention to communication overheads and risks, which can be significant [60]. Furthermore, operationalization of STEAM is based on enhancements to the Soar architecture [34], plus a set of about 300 domain-independent Soar rules.

Building on the well developed theory of joint intentions [8] and shared plans [22] [20], the STEAM teamwork model [59] was operationalized as a set of domain independent rules that describe how teams should work together.

Tambe's work on *STEAM teamwork model* [59].

STEAM enables explicit representation of team goals and plans, and teams joint commitments.

STEAM (Shell for Teamwork) builds on both Joint Intention Theory and Shared Plan Theory and tries to overcome their shortcomings. Based on joint intentions, STEAM builds up hierarchical structures that parallel the Shared Plan Theory as described in the previous chapter. Hence, STEAM formalises commitments by building an maintaining Joint Intentions, and uses SharedPlans to formulate the team's attitudes in complex tasks.

In [59] Tambe presents STEAM, an implemented model of teamwork based primarily on Cohen et al.'s theory of Joint Intentions, but informed by key concepts from SharedPlans. Following Cohen et al., a team initially adopts a joint intention for a high-level team goal that includes commitments to maintain the goal until it is deemed already achieved, unachievable or irrelevant. The agents then construct a hierarchy of individual and joint intentions analogous to partial SharedPlans. Tambe notes that as the hierarchy evolves, if a step involves only a subteam then that subteam must form a joint intention to perform that step, and the remaining team members need only track the subteam's joint intention, requiring that they be able to infer whether or not the subteam intends to, or is able to, execute that step [27].

5.3.2 Joint Responsibility

Jennings work in [30]

5.3.3 Planned Team Activity

6 Relation to Psychology and Sociology

[from Hoffman & Breazeal] In Bratmans detailed analysis of Shared Cooperative Activity (SCA), he defines certain prerequisites for an activity to be considered shared and cooperative: he stresses the importance of mutual responsiveness, commitment to the joint activity and commitment to mutual support. His work also introduces the idea of meshing singular sub-plans into a joint activity. In our implementation, we generalize this concept to the idea of dynamically meshing sub-plans.

Referring expressions [23]

7 Similarities and Differences

- SharedPlans theory:

+ Teammates agree on SharedPlan

+ Plan it together, execute it together

+ Specifies conditions for assistance, monitoring

- Joint Intentions theory

+ Teammates agree on intentions

+ Teammates agree on selecting/deselecting goals

+ i.e., goal unachievable, achieved, irrelevant

Similar to SharedPlans theory by Grosz and Sidner, Joint Intentions theory specifies what it means for agents to execute actions as a team. [58].

Cohen and Levesque Joint Intentions theory also states that a joint action could not be seen as a collection of individual ones but that agents working together need to share belief.

In contrast with Joint Intentions theory, the concept of SharedPlans theory is not based on a joint mental attitude , i.e., joint intention. Instead, SharedPlans relies on a novel intentional attitude, intending that (see Section 5.2), which is similar to an agent's normal intention to do an action. However, an individual agent's intention that is directed towards its collaborator's actions or towards a group's joint action.

1. None of SharedPlans' four components (see Section 5.2) has the notion of a joint intention. This is a significant difference between SharedPlans and Joint Intentions theories, since the notion of joint intention is an integral part of Cohen and Levesque's theory. In particular, SharedPLans theories emphesizes on the agents individually intending

that the joint action be done successfully as well as the agents individually intending the success of their collaborators' actions which is introduced in [20] by Grosz and Kraus as the notion of *intention-that*.

2. Communication requirements are derived from any intention that's, as opposed to being "hard-wired" in Joint Intentions.

3. In contrast to Joint Intentions, the SharedPlans Theory employs hierarchical structures over intentions, thus overcoming the shortcoming of a single Joint Intention for complex team tasks. The Shared Plans Theory is not based on a joint mental attitude but on an intentional attitude called intending that, which is very similar to an agents normal intention to perform an action.

4. Another main difference between the Joint Intentions Theory and the SharedPlans Theory is that the Shared Plans Theory describes the way to achieve a common goal through the hierarchy of plans, whereas the Joint Intentions Theory describes only this common goal [56].

5. Joint Intentions theory assumes that knowledge about the team-mates is always available.

8 Application in Human-Computer Collaboration

COLLAGEN [49, 50] is the first implemented system based on the SharedPlans theory. It incorporates certain algorithms for discourse generation and interpretation, and is able to maintain a segmented interaction history, which facilitates the discourse between human user and the intelligent agent. The model includes two main parts: (1) a representation of discourse state and (2) a discourse interpretation algorithm utterances of the user and agent [51].

In [23] Heeman presents a computational model of how a conversational participant collaborates in order to make a referring action successful. The model is based on the view of language as goal-directed behaviour, and in his work, he refers to SharedPlans as part of the planning and conversation literature.

In [38], Lochbaum and Sidner modify and expand the SharedPlan model of collaborative behavior [22]. They present an algorithm for updating an agents beliefs about a partial SharedPlan and describe an initial implementation of this algorithm in the domain of network management.

The system GRATE* by Jennings [30] is based on the Joint Intention Theory. GRATE* provides a rule-based modelling approach to cooperation using the notion of Joint Responsibilities, which in turn is based on Join Intentions. GRATE* is geared towards industrial settings in which both agents and the communication between them can be considered to be reliable.

CAST (Collaborative Agents for Simulating Teamwork) [61] [62] is a teamwork framework based on the SharedPlans Theory. CAST focuses on flexibility in dynamic environments and on proactive information exchange enabled by anticipating what information team members will need. Petri Nets are used to represent both the team structure and the teamwork process, i.e., the plans to be executed.

Researchers in [25] discuss developing an ontology of microsocial concepts for use in an instructional system for teaching cross-cultural communication. They believe being acquainted with one another is not a strong enough relationship to create a society from. Hence, there is a need for commitment and shared plans (as the basis of social life) to achieve a shared goal. In this work, Gorsz and Sidner's SharedPlans theory [22] is used to explain the concept of shared plans within the interpersonal relationships of societies in an industrial environment.

In [28] Hunsberger and Grosz discuss the idea of whether the rational, utility-maximizing agents should determine commiting to a group activity when there is an opportunity to collaborate. They call this problem as the "initial-commitment decision problem" (ICDP) and provide a mechanism that agents can use to dolve the ICDP. They use the representation of action, act-types and recepies in the SharedPlans theory.

In [63] an integrated agent-based model for Group Decision Support Systems is proposed and discussed. The decisional model that authors outline in this paper is based on the SharedPlans theory.

Rauenbusch and Grosz in [47] formally define a search problem with search operators that correspond to the team planning decisions. They provide an algorithm for making the three types of interrelated decisions by recasting the problem as a search problem. Their model respects the constraints on mental states specified by the SharedPlans theory of collaboration.

In [39] authors provide their in RoboCup (robotics soccer testbed) in which their focus is on teamwork and learning challenges. Their research investigationin RobotCup is based on ISI Synthetic, a team of synthetic soccer-players. They also investigate the use of STEAM as their model of teamwork which is influenced by the Joint Intentions and SharedPlans theories.

Babaian et. al. ini [2] describe Writers Aid, a system that deploys AI planning techniques to enable it to serve as an authors collaborative assistant. While an author writes a document, Writers Aid helps in identifying and inserting citation keys and by autonomously finding and caching potentially relevant papers and their associated bibliographic information from various on-line sources. They believe the underlying concepts of SharedPlans is relevent since in collaborative interfaces like Writers Aid, the users establish shared goals with the system and user and the system both take initiative in satisfying them.

In [41] researchers address high-level robot planning issues for an interactive cognitive robot that acts in presence or in collaboration with a human partner. They describe a Human Aware Task Planner (HATP) which is designed to provide socially acceptable plans to achieve collaborative tasks. They use notions of plans based on SharedPlans theory.

In [55] Sidner and Dzikovska argue that robots, in order to participate in conversations with humans, need to make use of conventions of conversation and the means to be connected to their human counterparts. They provide an initial research on engagement in human-human interaction and applications to stationary robots in hosting activities. They believe hosting activities are collaborative because neither party completely determines the goals to be undertaken nor the means of reaching the goal. To build a robot host, they rely on an agent built using Collagen which is implemented based on the SharedPlans theory.

In [32] authors introduce a language for representing joint plans for teams of agents. They describe how agents can organize the formation of a suitably skilled team to achieve a joint goal, and they explain how such a team can execute these plans to generate complex, synchronized team activity. In this paper, authors adopt the underlying concepts of the Joint Intentions theory as the structure of their collaborative agents.

Breazeal et. al. in [4] present an overview of their work towards building socially intelligent, cooperative humanoid robots, Leonardo, that can collaborate and learn in partnership with humans. They employ the Joint Intentions theory of collaboration to implement the collaborative behaviors while performing a task in collaboration with humans.

In [58] researchers' goal is to develop an architecture (based on the concepts of Joint Intentions theory) that can guide an agent during collaborative teamwork. They describe how a joint intention interpreter that is integrated with a reasoner over beliefs and communicative acts can form the core of a dialogue engine. Ultimately, the system engages in dialogue through the planning and execution of communicative acts necessary to attain the collaborative task at hand.

Mutlu et. al. in [42] discuss key mechanisms for effective coordination toward informing the design of communication and coordination mechanisms for robots. They present two illustrative studies that explore how robot behavior might be designed to employ these mechanisms (particularly joint attention and action observation) to improve measure of task performance in human-robot collaboration. Their work uses Joint Intentions theory to develop shared task representations and strategies for task decomposition.

In [31] researchers propose a behavioral architecture C^2BDI that allows to enhance the knowledge sharing using natural language communication between team members. They define collaborative conversation protocols that provide proactive behavior to agents for the coordination between team members. Their agent architecture provides deliberative and conversational behaviors for collaboration, and it is based on both of the SharedPlans and Joint Intentions theories.

This domain independent teamwork model, STEAM, has been successfully applied to a variety of domains. From combat air missions [24] to robot soccer [33] to teams supporting human organizations [46] to rescue response [52], applying the same set of STEAM rules has resulted in successful coordination between heterogeneous agents. The successful use of the same teamwork model in a wide variety of diverse domains provides compelling evidence that it is the principles of team- work, rather than exploitation of specic domain phenomena, that underlies the success of teamwork based approaches.

There are many research focusing on different aspects of collaboration each of which are different than my own work. In my thesis, I focus on emotion functions and how they impact collaboration's structure and processes, and how the dynamics of the collaboration structure influences emotion-regulated processes. Some of the other works focus on the concepts of robot assistants [7], or teamwork and its challenges in cognitive and behavioral levels [8, 43, 53, 59]. Some researchers have an overall look at a collaboration concept at the architectural level. In [12] authors present a collaborative architecture, COCHI, to support the concept of emotional awareness. In [11] authors present the integration of emotional competence into a cognitive architecture which runs on a robot, MEXI. In [57] authors discuss the challenges of integrating natural language, gesture understanding and spatial reasoning of a collaborative humanoid robot situated in the space. The importance of communication during collaboration has been considered by some researchers from human-computer interaction and human-robot collaboration [6, 40, 50] to theories describing collaborative negotiation, and discourse planning and structures [1, 21, 54]. There are other concepts such as joint actions and commitments [18], dynamics of intentions during collaboration [35], and task-based planning providing more depth in the context of collaboration [5, 48].

The concept of collaboration has also received attention in the industry and in research in robotic laboratories [13].

9 Conclusion

References

[1] Jerry Andriessen, Koenraad de Smedt, and Michael Zock. Discourse planning: Empirical research and computer models. In Anton Dijkstra and Koenraad de Smedt, editors, *Computational psycholinguistics: AI and connectionist models of human language processing*, pages 247–278. Taylor & Francis, 1996.

[2] Tamara Babaian, Barbara J. Grosz, and Stuart M. Shieber. A writer's collaborative assistant. In *In Proceedings of the International Conference on Intelligent User Interfaces (IUI2000)*, pages 7–14. ACM Press, 2002.

[3] Michael E. Bratman. *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press, 1987.

[4] Cynthia Breazeal, Andrew Brooks, Jesse Gray, Guy Hoffman, Cory Kidd, Hans Lee, Jeff Lieberman, Andrea Lockerd, and David Mulanda. Humanoid robots as cooperative partners for people. *Journal of Humanoid Robots*, 1(2):1–34, 2004.

[5] Catherina Burghart, RalfMikut, Rainer Stiefelhagen, Tamim Asfour, Hartwig Holzapfel, Peter Steinhaus, and Ruediger Dillmann. A cognitive architecture for a humanoid robot: A first approach. In *5th IEEE-RAS International Conference on Humanoid Robots*, pages 357–362, 2005.

[6] Aaron B. St. Clair and Maja J. Matarić. Modeling action and intention for the production of coordinating communication in human-robot task collaborations. In *21st IEEE International Symposium on Robot and*

Human Interactive Communication: Workshop on Robot Feedback in HRI, Paris, France, 2012.

[7] William J. Clancey. Roles for agent assistants in field science: Understanding personal projects and collaboration. *IEEE Transactions on Systems, Man and Cybernetics, special issue on Human-Robot Interaction*, 34(2):125–137, 2004.

[8] Philip Cohen and Hector J. Levesque. *Teamwork*. SRI International, 1991.

[9] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–261, 1990.

[10] Philip R. Cohen and Hector J. Levesque. Persistence, intention, and commitment. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, pages 33–69. MIT Press, Cambridge, MA, 1990.

[11] Natascha Esau, Lisa Kleinjohann, and Bernd Kleinjohann. Integrating emotional competence into man-machine collaboration. In *Biologically-Inspired Collaborative Computing, September 8-9, Milano, Italy*, pages 187–198, 2008.

[12] Octavio García, Jessús Favela, Guillermo Licea, and Roberto Machorro. Extending a collaborative architecture to support emotional awareness. In *Emotion Based Agent Architectures (ebaa99*, pages 46–52, 1999.

[13] Scott A. Green, Mark Billinghurst, XiaoQi Chen, and J. Geoffrey Chase. Human-robot collaboration: A literature review and augmented reality approach in design. *International Journal of Advanced Robotic Systems*, 5(1):1–18, 2008.

[14] Barbara Grosz and Sarit Kraus. The evolution of shared plans. In *Foundations and Theories of Rational Agency*, pages 227–262, 1998.

[15] Barbara J. Grosz. AAAI-94 presidential address: Collaborative systems. *AI Magazine*, 17(2):67–85, 1996.

[16] Barbara J. Grosz. Beyond mice and menus. *Proceedings of the American Philosophical Society*, 149(4):523–543, 2005.

[17] Barbara J. Grosz and Luke Hunsberger. The dynamics of intention in collaborative activity. *Cognitive Systems Research*, 7(2-3):259–272, 2006.

[18] Barbara J. Grosz and Luke Hunsberger. The dynamics of intention in collaborative activity. *Cognitive Systems Research*, 7(2-3):259–272, 2007.

[19] Barbara J. Grosz, Luke Hunsberger, and Sarit Kraus. Planning and acting together. *AI Magazine*, 20(4):23–34, 1999.

[20] Barbara J. Grosz and Sarit Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.

[21] Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, July 1986.

[22] Barbara J. Grosz and Candace L. Sidner. Plans for discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 417–444. MIT Press, Cambridge, MA, 1990.

[23] Peter Anthony Heeman. *A Computational Model of Collaboration on Referring Expressions*. PhD thesis, University of Toronto, 1991.

[24] Randall W. Hill, Jr., Johnny Chen, Jonathan Gratch, Paul Rosenbloom, and Milind Tambe. Intelligent agents for the synthetic battlefield: A company of rotary wing aircraft. In *Innovative Applications of Artificial Inttigence (IAAI-97)*, pages 227–262, 1997.

[25] Jerry R. Hobbs, Alicia Sagae, and Suzanne Wertheim. Toward a commonsense theory of microsociology: Interpersonal relationships. In *Formal Ontology in Information Systems - Proceedings of the Seventh International Conference*, pages 249–262, 2012.

[26] Marcus Huber and Edmund H. Durfee. On acting together: Without communication. In *In Working Notes of the AAAI Spring Symposium on Representing Mental States and Mechanisms*, pages 60–71, 1995.

[27] Luke Hunsberger. Making Shared Plans More Concise and Easier to Reason About. In *Proceedings of the 5th International Workshop on Intelligent Agents V : Agent Theories, Architectures, and Languages (ATAL-98)*, volume 1555, pages 81–98. Springer-Verlag: Heidelberg, Germany, 1999.

[28] Luke Hunsberger and Barbara J. Grosz. A combinatorial auction for collaborative planning. In *In Proceedings of ICMAS*, 2000.

[29] Bevan Jarvis, Dennis Jarvis, and Lakhmi Jain. Teams in multi-agent systems. In *Intelligent Information Processing III*, volume 228. Springer US, 2007.

[30] Nicholas R. Jennings. Controlling cooperative problem solving in industrial multi-agent systems using joint intentions. *Artificial Intelligence*, 75(2):195–240, 1995.

[31] Alexandre Kabil, Camille De Keukelaere, and Pierre Chavaillier. Coordination mechanisms in human-robot collaboration. In *Proceeding of the 7th International Conference on Advances in Computer-Human Interactions*, pages 389–394, 2014.

[32] David Kinny, Magnus Ljungberg, Anand Rao, Gil Tidhar, Eric Werner, and Elizabeth Sonenberg. Planned team activity. In *Lecture notes in artificial intelligence*. Springer-Verlag, 1992.

[33] Hiroaki Kitano, Minoru Asada, Yasuo Kuniyoshi, Itsuki Noda, Eiichi Osawai, and Hitoshi Matsubara. Robocup: A challenge problem for AI. *AI Magazine*, 18(1):73–85, 1997.

[34] John Laird. *The Soar Cognitive Architecture*. MIT Press, 2012.

[35] Hector J. Levesque, Philip R. Cohen, and Jos H. T. Nunes. On acting together. In *AAAI*, pages 94–99. AAAI Press / The MIT Press, 1990.

[36] D. J. Litman and J. F. Allen. Discourse processing and commonsense plans. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 365–388. MIT Press, Cambridge, MA, 1990.

[37] Karen E Lochbaum. A collaborative planning model of intentional structure. *Computational Linguistics*, 24(4):525–572, 1998.

[38] Karen E. Lochbaum, Barbara J. Grosz, and Candace L. Sidner. Models of plans to support communication: An initial report. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 485–490. AAAI Press, 1990.

[39] Stacy Marsella, Jafar Adibi, Yaser Al-Onaizan, Ali Erdem, Randall Hill, Gal A. Kaminka, Zhun Qiu, and Milind Tambe. Using an explicit teamwork model and learning in robocup: An extended abstract. In *RoboCup-98: Robot Soccer World Cup II*, volume 1604, pages 237–245. Springer Berlin Heidelberg, 1999.

[40] Laetitia Matignon, Abir Beatrice Karami, and Abdel-Illah Mouaddib. A model for verbal and non-verbal human-robot collaboration. In *AAAI Fall Symposium Series*, pages 62–67, 2010.

[41] Vincent Montreuil, Aurélie Clodic, Maxime Ransan, and Rachid Alami. Planning human centered robot activities. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 2618–2623, 2007.

[42] Bilge Mutlu, Alison Terrell, and Chien-Ming Huang. Coordination mechanisms in human-robot collaboration. In *Proceedings of the HRI 2013 Workshop on Collaborative Manipulation*, 2013.

[43] Stefanos Nikolaidis, Przemyslaw A. Lasota, Gregory F. Rossano, Carlos Martinez, Thomas A. Fuhlbrigge, and Julie A. Shah. Human-robot collaboration in manufacturing: Quantitative evaluation of predictable, convergent joint action. In *ISR*, pages 1–6, 2013.

[44] Martha E. Pollack. A model of plan inference that distinguishes between the beliefs of actors and observers. In *Proceedings of the 24th Annual Meeting on Association for Computational Linguistics*, pages 207–214. Association for Computational Linguistics, 1986.

[45] Martha E. Pollack. Plans as complex mental attitudes. In *Intentions in Communication*, pages 77–103. MIT Press, 1990.

[46] David V. Pynadath and Milind Tambe. An automated teamwork infrastructure for heterogeneous software agents and humans. *Journal of Autonomous Agents and Multi-Agent Systems, Special Issue on Infrastructure and Requirements for Building Research Grade Multi-Agent Systems,*, 7(1-2):71–100, 2003.

[47] Timothy W. Rauenbusch and Barbara J. Grosz. A decision making procedure for collaborative planning. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, 2003.

[48] Charles Rich. Building task-based user interfaces with ANSI/CEA-2018. *IEEE Computer*, 42(8):20–27, July 2009.

[49] Charles Rich and Candace L. Sidner. COLLAGEN: A collaboration manager for software interface agents. *User Modeling User-Adapted Interaction*, 8(3-4):315–350, 1998.

[50] Charles Rich, Candace L. Sidner, and Neal Lesh. COLLAGEN: Applying collaborative discourse theory to human-computer interaction. *AI Magazine*, 22(4):15–26, 2001.

[51] Jeff Rickel, Neal Lesh, Charles Rich, Candace L. Sidner, and Abigail Gertner. Collaborative discourse theory as a foundation for tutorial dialogue. In *In Proceedings Sixth International Conference on Intelligent Tutoring Systems*, 2002.

[52] Paul Scerri, David Pynadath, Lewis Johnson, Paul Rosenbloom, Mei Si, Nathan Schurr, and Milind Tambe. A prototype infrastructure for distributed robot-agent-person teams. In *The Second International Joint Conference on Autonomous Agents and Multiagent Systems*, 2003.

[53] Paul Scerri, David Pynadath, Lewis Johnson, Paul Rosenbloom, Mei Si, Nathan Schurr, and Milind Tambe. A prototype infrastructure for distributed robot-agent-person teams. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '03, pages 433–440, New York, NY, USA, 2003. ACM.

[54] Candace Sidner. An artificial discourse language for collaborative negotiation. In *In Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 814–819. MIT Press, 1994.

[55] Candace L. Sidner and Myroslava Dzikovska. A first experiment in engagement for human-robot interaction in hosting activities. In *Advances in Natural Multimodal Dialogue Systems*, volume 30 of *Cognitive Technologies*, pages 55–76. Springer Netherlands, 2005.

[56] Hendrik Skubch. *Modelling and Controlling of Behaviour for Autonomous Mobile Robots*. Springer Science Business Media, 2012.

[57] Donald Sofge, Magdalena D. Bugajska, J. Gregory Trafton, Dennis Perzanowski, Scott Thomas, Marjorie Skubic, Samuel Blisard, Nicholas Cassimatis, Derek P. Brock, William Adams, and Alan C. Schultz. Collaborating with humanoid robots in space. *International Journal of Humanoid Robotics*, 2(2):181–201, 2005.

[58] Rajah Annamalai Subramanian, Sanjeev Kumar, and Philip Cohen. Integrating joint intention theory, belief reasoning, and communicative action for generating team-oriented dialogue. In *AAAI*, pages 1501–1507. AAAI Press, 2006.

[59] Milind Tambe. Towards flexible teamwork. *Journal of Artificial Intelligence Research*, 7:83–124, 1997.

[60] Milland Tambe. Agent architecture for flexible, practical teamwork. In *Proceedings of the National Conference on Artificial Intelligence*, pages 22–28, 1997.

[61] John Yen, Jianwen Yin, Thomas R. Ioerger, Michael S. Miller, Dianxiang Xu, and Richard A. Volz. Cast: Collaborative agents for simulating teamwork. In *In Proceedings of IJCAI2001*, pages 1135–1142, 2001.

[62] Jianwen Yin, Michael S. Miller, Thomas R. Ioerger, John Yen, and Richard A. Volz. A knowledge-based approach for designing intelligent team training systems. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 427–434. ACM, 2000.

[63] Zamfirescu and Candea. On integrating agents into gdss. In *In Preprints of the 9th IFAC / IFORS / IMACS / IFIP/ Symposium on Large Scale Systems: Theory and Applications*, 2001.