Improved Human-Robot Team Performance Using *Chaski*, A Human-Inspired Plan Execution System

Julie Shah
MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar St. Room 32-D224
Cambridge, MA 02139
julie_a_shah@csail.mit.edu

James Wiken
MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar St. Room 32-D224
Cambridge, MA 02139
wyken@mit.edu

Brian Williams
MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar St. Room 32-D224
Cambridge, MA 02139
williams@mit.edu

Cynthia Breazeal
MIT Media Lab
20 Ames Street E15-468
Cambridge, MA 02139
cynthiab@media.mit.edu

ABSTRACT

We describe the design and evaluation of *Chaski*, a robot plan execution system that uses insights from human-human teaming to make human-robot teaming more natural and fluid. Chaski is a task-level executive that enables a robot to collaboratively execute a shared plan with a person. The system chooses and schedules the robot's actions, adapts to the human partner, and acts to minimize the human's idle time.

We evaluate Chaski in human subject experiments in which a person works with a mobile and dexterous robot to collaboratively assemble structures using building blocks. We measure team performance outcomes for robots controlled by Chaski compared to robots that are verbally commanded, step-by-step by the human teammate. We show that Chaski reduces the human's idle time by 85%, a statistically significant difference. This result supports the hypothesis that human-robot team performance is improved when a robot emulates the effective coordination behaviors observed in human teams.

Categories and Subject Descriptors: J.7 Computer Applications: Computers in other systems

General Terms: Algorithms, Experimentation, Performance.

Keywords: task allocation and coordination, experiments on HRI, HRI group dynamics, autonomy and trust.

1. INTRODUCTION

We envision a future in which collaboration between humans and robots will be indispensable to our work in many domains, ranging from surgery to space exploration. The success of these systems will depend in part on the ability of robots to integrate with existing human teams. Our goal is to develop robot partners that we can work with more easily and naturally, as inspired by the way we work with other people.

Today we treat robots primarily as tools that we explicitly command to perform tasks step-by-step. However, we know from studies of human teamwork that explicitly commanding is an inefficient means of coordinating the actions of multiple team members. Instead, the best human teammates anticipate what their partners will need and adapt to the actions of other team members [5, 17].

In this paper, we hypothesize and test whether human-robot team performance is improved when a robot teammate emulates the behaviors and teamwork strategies observed in human teams. We apply insights from human teamwork studies in order to design and evaluate Chaski, a robot plan execution system that makes human-robot teaming more natural and fluid. Chaski is a task-level executive that advances the state-of-the-art in dynamic plan execution. The system enables a robot to robustly anticipate and adapt to other team members, make decisions on-the-fly, and consider the consequences of its actions on others. A key strength of Chaski is that it generalizes naturally to dfferent styles of teamwork: *Equal Partners* and *Leader and Assistant*.

We report on human subject experiments in which a person works with a robot under the *Equal Partners* model of teamwork to collaboratively assemble structures using building blocks. We show that Chaski reduces the human's idle time by 85%, a statistically significant difference. This result supports the hypothesis that human-robot team performance is improved when a robot emulates the effective coordination behaviors observed in human teams.

HRI'11, March 6–9, 2011, Lausanne, Switzerland.

2. HUMAN TEAMING AS A GUIDE FOR HUMAN-ROBOT TEAMING

Our hypothesis is that the performance of human-robot teams is improved when a robot teammate emulates the effective coordination behaviors observed in human teams. There is a precedent for HHI informing the design of HRI (e.g. [11, 15, 22, 25, 8]). We draw from a body of human-human interaction (HHI) research that has not yet been applied to HRI: studies in human teamwork under stress induced by uncertainty, ambiguity, and time pressure. These include studies of military tactical teams [5, 4, 24], aviation crews [16], and medical teams [12]. Team dynamics have a significant impact on performance within these domains, producing a strong incentive for teams to understand and apply the communication and coordination strategies that improve performance.

Identifying the behaviors people use to coordinate actions within a team is the first step toward understanding how a robot may emulate an effective human teammate. In this section, we review the key results from these human-teamwork studies and, based on these insights, provide a set of design requirements for the Chaski system.

Teammates make decisions on-the-fly.

Effective teams tend to distribute work among team members on-the-fly. In other words, the best teams typically do not decide beforehand entirely who will do what and when. Instead team members show flexibility in making these decisions as circumstances unfold [5].

Teammates frequently communicate progress on the task.

Team members coordinate their actions through frequent updates on the status of the task. For example, teammates frequently update their partners on their progress by communicating when they start or finish parts of the task. Interestingly, studies show that the more team members communicate updates during the task, the better the team performs [4, 12]. This is the case even when team members coordinate to perform a task within a small shared workspace, such as a table surface [19].

Teammates consider the consequences of their actions on others.

Team members maintain shared mental models of the task and of each other's capabilities and use these models to consider the consequences of their actions on others [24]. Shared mental models provide team members with a common understanding of who is responsible for what task and what the information requirements are. In turn, this allows them to anticipate one another's needs so that team members can coordinate effectively. Evidence also suggests that people incorporate the capabilities of other team members into their own action planning [17], and that people act so as to minimize the idle time of other team members [19].

Design requirements for the Chaski Executive for Human-Robot Teaming

Based on results from these studies in human teamwork, we provide a set of design requirements for the Chaski Executive for Human-Robot Teaming.

(1) Chaski should take as input a shared plan that serves the same purpose as the shared mental model within a human team [24]. The shared plan should include the activities to be performed, plan deadlines, and information about the capabilities of each team member. Chaski should use the shared plan to choose and schedule the robot's activities. Chaski should make these decisions by considering the capabilities of each team member, so that the team can successfully complete the task within the plan deadlines.

(2) Chaski should enable a robot to choose just before execution which activities to perform and when. This should be based on knowledge of the plan execution so far. This ability to dynamically choose and schedule activities emulates the human ability to flexibly make decisions as circumstances unfold [5].

(3) Chaski should enable a robot to reason about the consequences of its actions on human teammates by favoring execution times that minimize the humans' idle time. This design requirement is based on the observation that human teammates consider the consequences of their actions on others [24], and that effective teams seem to act to minimize the team's idle time [19].

3. CHASKI EXECUTIVE

Chaski enables a human and a robot to execute a shared plan collaboratively under two different styles of teamwork: *Equal Partners* and *Leader and Assistant*. Chaski is presented in full detail in [18]. This paper focuses on the design and evaluation of the system for Equal Partners, one-on-one human-robot teaming.

Equal Partners teamwork is characterized by a flat, decentralized authority, meaning that each member of the team has equal authority to make decisions when executing the plan. Our model of Equal Partners assumes that team members fully know the capabilities of their teammates, in terms of which activities they may perform and bounds on how long the activities take. Also, the team does not negotiate beforehand who will do what and when, and instead makes decisions on-the-fly as circumstances unfold. Finally, the team members communicate to provide their team with timely information on the status of the task and rely on their team members to use this information when deciding what to do next.

We model Equal Partners teamwork along three dimensions: decision-making authority, decision-making strategy and communicative acts. Decision-making authority categorizes plan decisions as either within a team member's control or controlled exogenously by other teammates. Decision-making strategy refers to a team member's policy for deciding what activities to perform and when. Communicative acts describe the mechanism teammates use to coordinate their actions as they carry out the shared task.

Decision-making authority in the Equal Partners model of teamwork is characterized by three properties. (1) Each person has the authority to choose his or her own actions. (2) Each person has full control of the timing of their actions within specified bounds. (3) Each team member assumes that their teammates also have full authority to choose which actions to perform and have full control of the timing of their actions within specified bounds. As a result, in Equal Partner teamwork each member of the team has equal authority to make decisions when executing the plan.

Teammates use a dynamic **decision-making strategy** that delays task assignment and scheduling commitments until execution. In other words, rather than deciding who will do what and when ahead of time, the teammates make these decisions on-the-fly. This is consistent with results from human teamwork studies indicating that the most effective teams are able to redistribute tasks on-the-fly in response to changing circumstances. The model also assumes teammates employ a dynamic decision-making strategy that guarantees a successful plan execution. This means that teammates make decisions to ensure there is a way to complete the task that respects the temporal deadlines of the plan, and respects the model of the agents' capabilities.

Finally, teammates coordinate their actions through **communicative acts**. Specifically, a teammate communicates an "update" whenever he or she begins or finishes an activity in the plan. Studies of effective human teamwork indicate that the frequent offering of "updates" is correlated with improved team performance [19].

Next, we formally define the problem of Equal Partners plan execution.

3.1 Problem Statement: Equal Partners Plan Execution

3.1.1 Input

Chaski takes as input an Equal Partners plan that includes the activities to be performed, ordering constraints among the activities, and plan deadlines. The plan also includes information about the capabilities of the team members, including the activities that each agent may perform (e.g. Agent 1 may perform activity B), bounds on the amount of time each agent takes to perform each activity (e.g Agent 1 takes [5,7] minutes to perform B), as well as a description denoting which agents are human.

An Equal Partners plan encodes activities in terms of a set of variables $X_1, ..., X_n$, representing timepoints with real-valued domains. Each activity is composed of a *begin* timepoint and *end* timepoint.

Activity durations and other temporal constraints relating timepoints (e.x. "The entire plan must be completed within 250 seconds.") are formulated as binary constraints composed of simple intervals of the form:

$$(X_k - X_i) \in [a_{ik}, b_{ik}]. \tag{1}$$

An Equal Partners plan may also encode flexibility in which agent m performs each activity, and the corresponding choice in activity duration, by specifying an agent assignment to each interval in a disjunctive binary constraint as follows:

$$(X_k - X_i) \in P(\{agent_m : [a_{ik}, b_{ik}] | [a_{ik} \leq b_{ik}]\}), \tag{2}$$

Finally, the Equal Partners plan may include agent occupancy constraints, encoded as a set S of mutually exclusive intervals that cannot overlap in time.

3.1.2 Output

The output of Chaski is a *dynamic* and *least-commitment* policy, if one exists, for making task assignment and scheduling decisions. The policy ensures the team members work together to assign, schedule, and execute activities within the plan deadlines. The policy also includes a preference for task assignments and activity orderings that minimize a lowerbound on the humans' idle time.

A policy is *dynamic* if there exists an online strategy for making task assignments and scheduling decisions, given knowledge of all choices thus far, that will result in a full feasible schedule. A policy is *least-commitment* if each agent delays decisions until right before the commitment is made. In this case, agents delay deciding which activities they will perform and the timing of the activities.

The execution strategy generated by Chaski under the Equal Partners model is *correct* in that any complete task assignment and execution sequence generated by the executive also satisfies the constraints of the Equal Partners plan. Also, the execution strategy is *deadlock-free*, in that any partial execution generated by the executive can be extended to a complete execution that satisfies the constraints of the Equal Partners plan.

3.2 Technical Challenges

Development of an executive that adapts to a human on-the-fly is challenging because in high-tempo domains the robot must be able to choose and schedule its own activities very quickly in response to a human's actions. One approach to this problem is to make all the activity assignment and scheduling decisions ahead of time, before execution [14, 9]. The challenge with this approach is that any deviation from the initial activity assignment and schedule during execution requires re-planning. Assignment and scheduling for multi-agent temporal plans involving as few as three or four activities introduces time-consuming computations requiring tens of seconds [9], and as a result may significantly endanger the robot's ability to fulfill its role within the team.

Alternatively, many multi-agent systems employ an offline planning process to assign and order activities, but then enable the agents to schedule the precise timing of their activities online [1, 2, 10, 23]. Before execution, these systems perform *task assignment* to allocate activities among the agents, and then perform *synchronization* to introduce ordering constraints among activities so that concurrent execution remains logically valid. The process of task assignment and synchronization generates temporally flexible plans that the agents may use to schedule plan activities online, just before the activity is executed [13].

After task assignment and synchronization, agents are provided with a temporally flexible plan that enables them to make scheduling decisions and adapt to small disturbances online. For example, an agent may use this temporally flexible plan to decide to perform a task at 10:10am rather than 10am to adjust for schedule slip.

While this strategy allows the agent to adapt to some disturbances that occur prior to the activity, disturbances triggering task re-assignment or re-synchronization still require a deliberative capability to generate a new plan or perform plan repair. This re-planning process may require up to tens to hundreds of seconds [27], potentially endangering the robot's ability to fulfill its role within the team.

3.3 Approach

Chaski significantly improves the ability of robots to adapt on-the-fly, compared to prior work. The system's key innovation is a fast execution algorithm that operates on a compact encoding of the scheduling policies for all possible task assignments. Chaski first compiles the plan into to a compact encoding that can be efficiently executed. The compiled form of the plan makes explicit the consequences for each agent's activity choices and scheduling decisions. Agents then use this compiled plan to make task assignment and scheduling decisions online quickly.

By leveraging a compact encoding of multi-agent plans, Chaski enables agents to perform distributed dynamic execution while (1) reasoning on flexible scheduling policies for thousands of possible futures, and (2) often achieving execution latency within the bounds of human reaction time (250 ms) [20, 18]. Results show that Chaski enables agents to make task assignment and scheduling decisions one order of magnitude faster, on average, than prior work [26]. On moderately-sized benchmark plans composed of thousands of flexible scheduling policies, 89% of plans executed by Chaski exhibited an execution latency within human reaction time (250 ms), compared to only 24% executed using the Tsamardinos dispatcher [26].

The algorithms, and their empirical evaluation, for automatically compiling and executing Equal Partners plans are presented in full detail in [20, 18].

4. HUMAN-ROBOT TEAMING EXPERIMENTS

In this section, we report on experiments testing the hypothesis that human-robot team performance is improved when a robot teammate uses Chaski to emulate the behaviors and teamwork strategies observed in human teams. We measured team performance outcomes for robots controlled by Chaski (Implicit Teaming group) compared to robots that were verbally and explicitly commanded step-by-step by the human teammate (Explicit Teaming group).

4.1 Experiment Hypotheses

The experiments test the following two hypotheses about human-robot team performance.

Hypothesis 1: Chaski improves objective measures of team performance.

We hypothesize that human participants working with a robot controlled by Chaski will exhibit less idle time and take less time to complete the task than participants that verbally command the robot step-by-step.

This hypothesis is founded in human teamwork studies, ours and others', showing that improved performance is correlated with increased use of implicit coordination behaviors [24, 19]. In our human-robot teaming experiments, Chaski emulated effective human team coordination behaviors (e.g. adapting on-the-fly to other teammates, offering frequent updates on the status of the task, and acting to minimize the human's idle time).

Hypothesis 2: Chaski improves subjective measures of teaming quality.

We hypothesize that human participants working with a robot controlled by Chaski will agree more strongly that the team worked fluently together, the robot performed well, the team members shared common goals, and the robot was trustworthy, compared to participants that verbally command the robot step-by-step.

This hypothesis is informed by results reported in [8] that anticipatory action within a human-robot team positively impacted subjective measures of team performance and fluency.

4.2 Method

Participants

The participants consisted of 16 subjects (10 men and 6 women) recruited from the MIT and Greater Boston area. The average age was 29.4 years (SD = 16.1). The participants were randomly assigned to either the Implicit or Explicit teaming group.

Experiment Task

We developed an experimental task in which teams, each composed of one person and one robot, built pre-defined structures (presented in Figure 1) using a commercially available building block set.

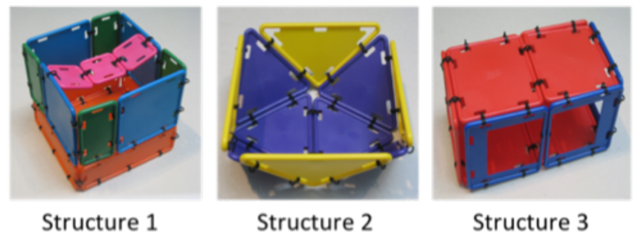

Figure 1: Three Structures for Teamwork Task

The base materials for each of the three structures were provided in hand to the human participant at the start of the task. The materials for the middles and tops of the structures were located in bags distributed on the floor within the experiment workspace. The human was pre-assigned the job of physically assembling the structures. However, either the human or robot was permitted to retrieve the bags with materials.

The team was tasked with collecting the building materials and assembling the three structures subject to the following four rules. The first two rules were developed to address the disparity in the humans' and robot's physical capabilities: (1) each team member may retrieve only one bag at a time, and (2) the human teammate is allowed to retrieve up to one bag between building each structure. The effect is that collaboration is required to complete the task; the robot must retrieve at least three bags.

The third rule (3) is that a teammate must follow through with an activity once he has communicated a commitment to perform the activity. This rule is required since, at this time, Chaski does not support the re-planning required when an agent changes its mind mid-activity. The fourth rule (4) is that the human teammate must finish gathering materials for and finish building Structures 1 and 2 before starting to build Structure 3. This rule imposes ordering constraints among the plan activities, and allows all bags to be placed within the limited dimensions of the experiment workspace.

Independent Variable

Sixteen human-robot teams performed the experimental task. Eight human teammates were randomly chosen to explicitly command the robot's actions step-by-step (Explicit Teaming Group). The other eight human teammates worked with a robot controlled by Chaski under the Equal Partners model of teamwork (Implicit Teaming Group). Chaski chose and scheduled the robot's activities with a preference to minimize the human's idle time.

Teams in the Implicit Teaming group coordinated their actions by communicating when they started and completed each activity. Each team member then relied on their partner to adapt based on these communications. This means that, within the Implicit Teaming group, the robot took the initiative to choose and schedule its own activities. The list of plan activities and the human's and robot's capabilities are listed in Table 1. The durations for each activity were chosen empirically based on team performance data from a pilot study.

Table 1: Team Capabilities

Activity	Agent	Duration(s)
Build the Base of Structure #1	Human	45-80
Build the Middle of Structure #1	Human	90-145
Build the Top of Structure #1	Human	15-50
Build the Base of Structure #2	Human	45-90
Build the Top of Structure #2	Human	45-90
Build the Base of Structure #3	Human	5-70
Build the Middle of Structure #3	Human	35-95
Build the Top of Structure #3	Human	25-70
Retrieve the Blue Squares	Human	15-30
	Robot	65-120
Retrieve the Green Rectangles	Human	15-30
	Robot	65-120
Retrieve the Pink Squares	Human	15-30
	Robot	65-120
Retrieve the Yellow Triangles	Human	15-30
	Robot	65-120
Retrieve the Blue Open Squares	Human	15-30
	Robot	65-120
Retrieve the Red Squares	Human	15-30
	Robot	65-120

Human team members in the Explicit Teaming group explicitly commanded the robot to perform the "retrieve materials" activities. The full list of commands is presented in Table 2. The robot began each activity immediately after receiving the command, and did not perform any activities outside of those commanded by the person. As with the Implicit Teaming group, both the human and robot also communicated when they started and completed each activity.

Table 2: Activity Commands

"Nexi, bring me the Blue Squares."
"Nexi, bring me the Green Rectangles."
"Nexi, bring me the Pink Squares."
"Nexi, bring me the Yellow Triangles."
"Nexi, bring me the Blue Open Squares."
"Nexi, bring me the Red Squares."

Dependent Measures Variable

Two team performance outcomes, time to complete the task and human idle time, were measured for each team. Both these measures were extracted from video recordings of the experiment trials. The human idle time was computed separately by two analysts. Agreement between the two analysts was found to be high, with a coefficient alpha of 0.98 [3]. Human idle time was defined as the cumulative amount of time a participant spent watching the actions of the robot, while not manipulating building materials. This is the same definition of idle time used in [19].

At the completion of the experiment, human participants were asked to rate their agreement with the statements presented in Table 3 on a 1-5 Likert scale, 1 for strongly disagree and 5 for strongly agree. The Likert questionnaire, similar to those used in [7, 8], addressed the robot's performance, the robot's contribution to the team, shared goals, team fluency, trust in the robot, and attribution of credit and blame. Participants were also asked to share their thoughts and comments in three open ended questions addressing the robot's performance, the robot's contribution to the team effort, and the fluency of the teamwork.

Table 3: Likert Questionnaire

1. Nexi's performance was an important contribution to the success of the team.
2. Nexi performed well as part of the team.
3. Nexi contributed equally to the team performance.
4. I felt like Nexi was committed to the success of the team.
5. Nexi perceives accurately what my goals are.
6. Nexi does not understand what I am trying to accomplish.
7. Nexi and I are working towards mutually agreed upon goals.
8. The team worked fluently together.
9. Nexi contributed to the fluency of the interaction.
10. I trusted Nexi to do the right thing at the right time.
11. Nexi was trustworthy.
12. Our success on the task was largely due to the things I said or did.
13. I am responsible for most of the things that we did well on this task.
14. Our success on this task was largely due to the things Nexi said or did.
15. Nexi should get credit for most of what we accomplished on this task.
16. I hold Nexi responsible for any errors that we made on this task.
17. Nexi is to blame for most of the problems we encountered in accomplishing this task.

4.3 Experiment Setup and Robot Platform

The experiment setup, pictured in Figure 2, consists of a work table where the person builds the experiment structures, and a floor area where the bags with building materials are initially placed.

The human participant works with Nexi, a Mobile-Dexterous-Social (MDS) robot. Nexi is a mobile robot platform capable of simple object manipulation and non-verbal social expression. The robot is approximately 48 inches tall, with a strength-to-mass ratio that allows it to interact safely with humans. Nexi has two manipulator arms, each with 6 degrees of freedom, and two hands, each with 2 degrees of freedom, to support pointing gestures and simple object manipulation. Nexi's 4-degree-of-freedom neck and 17-degree-of-freedom face supports a wide range of expressions and postures. Within the experiment, Nexi used the mobile base to

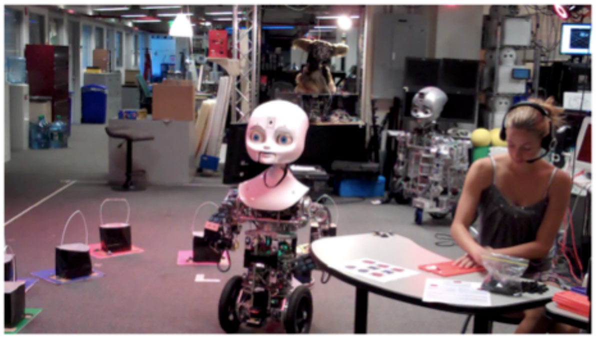

Figure 2: Experiment Workspace and Setup

drive to and from the bags and work table. The robot used its left manipulator arm and hand to pick up the bags. It used a previously developed saliency-based attention system to intermittently gaze towards nearby bags and the work table. The robot also nodded its head each time it received a verbal command from the person.

An off-board Vicon motion capture system was used for sensing of the experiment workspace. The Vicon system tracked the robot's position and orientation, and the locations of the bags and work table. The robot autonomously navigated around workspace using a map generated by Vicon data.

The robot used an open source speech recognition system, Sphinx-4, to recognize a simple grammar designed specifically for the experiment task. The grammar included predefined phrases for when the person began and finished each activity in Table 1, and also included the command phrases in Table 2. Participants wore a microphone headset and read phrases from a script to communicate with the robot. Software was developed to bypass the speech recognition system if necessary, so that experiment outcomes were not affected by erroneous speech recognition.

4.4 Procedure

The experiment was divided into a familiarization and test phase. Upon arrival, participants were seated at a table next to the robot. The table surface provided the workspace used to build the structures during both the familiarization and test phases. Prior to the familiarization phase, the participant was provided pictorial instructions for building the structures and a script with phrases for communicating with the robot.

During the familiarization phase, an independent experiment proctor read the participant instructions describing the experiment task, including the roles of the human and robot and the rules of the task. Participants were also informed that they would be video taped during the test phase of the experiment.

Participants were instructed that the experiment task involves building the three structures pictured in Figure 1 as fast as possible. Participants were also given the benchmark "best completion time to-date," calculated approximately 15% lower than the pilot study best completion time. Participants were told they must build each structure from the bottom up and that they start with the correct number of base materials for each structure already on the table. The building materials for the upper parts of the structures are located in black bags on the floor. Each bag contains a certain type of building material, indicated by the colored shape beneath each bag.

Participants were instructed that assembly of the structures is solely the human's responsibility. However either the human or the robot may retrieve the bags with building materials. Also, the human participant and robot must work together to build the structures subject to four rules described previously.

Next, participants were instructed to practice building the structures and communicating with the robot. Participants choose one structure and then practiced building it as fast as possible. Participants also practiced communicating updates to the robot while building the structure. Finally, participants in the Explicit Teaming group practiced commanding the robot to retrieve a bag.

In the test phase, each human-robot team performed the experiment task twice. At the end of each trial, participants were told their completion time and were reminded of the "best completion time to-date." Finally, at the completion of the experiment, participants were administered the Likert scale and open-ended questionnaires.

4.5 Results

In this section, we compare human idle time, time to complete the task, and subjective measures of teaming quality for the Implicit and Explicit Teaming groups. We interpret and discuss these results in the next section.

Idle Time

Human participants in the Implicit Teaming group spent 5 seconds (SD = 10s) idling in the first trial and 8 seconds (SD= 11s) idling in the second trial, on average. In comparison, human participants in the Explicit Teaming group spent on average 45 (SD= 34s) and 43 seconds (SD = 33s) idling in the first and second trials, respectively. Two-tailed, unpaired t-tests with unequal variance found the difference in idle time within each trial to be statistically significant (df=8, alpha=0.5, p=[0.01-0.02]). Within each group, no statistical difference was found between the means of the first and second trials.

Time to Complete Task

Implicit group teams on average performed the task in 13.6 minutes (SD = 1.9) and 11.2 minutes (SD=2.8) in the first and second trials, respectively. Teams in the Explicit group performed the task in 15.4 minutes (SD = 3.7) and 12.1 minutes (SD=2.9), respectively. Two-tailed, unpaired t-tests with unequal variance found the difference in completion time within each trial to not be statistically significant (df=8, alpha=0.5, p=[0.30-0.57]). Also, within each group, no statistical difference was found between the means of the first and second trials.

Subjective Measures

People in the Implicit Teaming group agreed with statement #11 in Table 3, "the robot is trustworthy," more strongly than people in the Explicit Teaming group. Two-tailed Wilcoxon-Mann-Whitney tests (df=8, alpha=0.5) found this difference to be statistically significant (U =11, p = 0.02). No statistically significant differences were found for responses to the other statements in Table 3.

Interestingly, there is a moderate correlation (r=+/-[0.4-0.5]) between a number of the Likert scores and the objective measures of team performance. A moderate, negative correlation was found between time to complete the task and Likert responses for statements #1,2 and 8 on robot performance and team fluency. A moderate, positive correlation was found between human idle time and Likert responses for statements #14 and 15 addressing attribution of credit to the robot.

Sample of Open-ended Responses

The open-ended responses for the two groups provide insight into the participants' experience of team fluency, robot performance, and common goals. The sample of open-ended responses provided in this section suggest that the experiences of participants in the two groups may have differed along these dimensions, even though the Likert questionnaire results do not report statistically significant differences for these measures.

Explicit Group:

> "It seems as though Nexi should be able to bring the materials I required without explicit orders based on which structure I was working on."

> "[Fluency of teamwork] largely depended on my foresight and ability to multi-task. If I asked for material out of order, it was my fault."

Implicit Group:

> "Nexi understood everything that I said and she knew what materials I needed, and in what order, to build all the structures. I think it was great (and helpful) that I didn't have to ask for specific materials."

> "Nexi understood what needed to be done and helped retrieve the materials necessary to build the structures. When I gave status updates and when I communicated if I had or hadn't all the materials, Nexi proved to know what needed to be done next. It was a big help having her work with me."

> "Nexi was helpful in making sure that I got all of the materials for the tasks and made sure that the building process was not delayed."

5. DISCUSSION

The results presented in the previous section provide the first evidence that human-robot teamwork is improved when a robot emulates the behaviors and teamwork strategies observed in human teams. Human participants in the Implicit Teaming group spent 85% less time idling, on average, than human participants in the Explicit Teaming group, a statistically significant difference ($p < 0.05$). Human idle time was reduced from 44 seconds to 6 seconds, on average. This result supports the hypothesis that human participants working with a robot controlled by Chaski exhibit less idle time than participants that verbally command the robot step-by-step. Of the reported results, this data most strongly supports the hypothesis that human-robot team performance is improved when a robot emulates the effective coordination behaviors observed in human teams.

Analysis also indicates that Implicit Teaming groups performed the task 7-12% faster, on average, than Explicit Teaming groups. This result is not statistically significant, and as a result we are unable to confirm the hypothesis that Implicit group teams take less time to complete the task than Explicit group teams. This is in part due to a large variance in time to complete the task and the low number of subjects. However, the trend is in the right direction and warrants further investigation.

Participants in the Implicit Teaming group agreed with the statement "the robot is trustworthy" more strongly than people in the Explicit Teaming group, a statistically significant difference ($p<0.05$). However, Implicit group participants did not agree more strongly than Explicit group participants that the team worked fluently together, the robot performed well, or that the team members shared common goals. These results are surprising considering previously reported results [8] that anticipatory action within a human-robot team positively impacted these subjective measures.

One possible explanation for these results is that a moderate correlation between Likert question scores and objective team performance measures dominated the Implicit versus Explicit group effect. Analysis shows a moderate, negative correlation between time to complete the task and Likert responses for robot performance and team fluency. This means that there is a correlation between finishing the task quickly and agreement that the robot performed well and the team worked fluently together. Analysis also shows a moderate, positive correlation between human idle time and Likert responses addressing attribution of credit to the robot. This means the participants' idle time was related to their agreement that the robot contributed to the success of the team.

6. FUTURE WORK

There are a number of aspects of real-world HRI that are idealized in these experiments and that warrant investigation in future work. Chaski, in its current form, does not interpret intent of the human, but rather adapts the plan based on certain and unambiguous information on the status of the task. Grounding and interpretation of ambiguous communications were also not a part of the experiments conducted. In future experiments, it would be interesting to investigate more free-form human-robot interactions by incorporating the use of implicit verbal and non-verbal communications, and strategies to disambiguate referents of these implicit communications (e.g. [25, 6]).

The lack of a re-planning capability also constrained the human's interaction with the robot. Chaski uses an incremental algorithm to compactly compile the robot's plan. A variant of this same algorithm has also been applied to incrementally repair plans online (see [21]). The natural next step is to unify these two capabilities. This would also enable more flexible interactions by providing either the human or robot teammates the ability to interrupt ongoing activities and add or remove activities in the plan on-the-fly.

7. CONCLUSIONS

In this paper, we described the design and evaluation of *Chaski*, a robot plan execution system that uses insights from human-human teaming to make human-robot teaming more natural and fluid. Chaski is a task-level executive that enables a robot to collaboratively execute a shared plan with a person. The system generalizes the state-of-the-art in dynamic plan execution by supporting just-in-time task assignment as well as scheduling. The key innovation of Chaski is a fast execution algorithm that operates on a compact encoding of the scheduling policies for all possible task assignments. By leveraging a compact encoding of multi-agent plans, Chaski enables a robot to robustly anticipate and adapt to other team members, make decisions on-the-fly, and consider the consequences of its actions on others.

We have evaluated Chaski in human subject experiments in which a person works with a robot to collaboratively assemble structures using building blocks. We show that Chaski reduces the human's idle time by 85%, a statistically significant difference. This result supports the hypothesis that human-robot team performance is improved when a robot emulates the effective coordination behaviors observed in human teams.

8. ACKNOWLEDGMENTS

We would like to thank Kenton Williams, Sigurdur Orn Adalgeirsson, and Akara Ambak for their help in running the human-robot teaming experiments.

This work was supported by ONR MURI Award No.: N000140710749, The Boeing Company grant MIT-BA-GTA-1, and NDSEG American Society for Engineering Education.

9. REFERENCES

[1] R. Alami, F. Ingrand, and S. Qutub. A scheme for coordinating multi-robot planning activities and plans execution. In *Proceedings of ECAI.* Brighton, UK, 1998.

[2] M. Brenner. Multiagent planning with partially ordered temporal plans. In *Proceedings of AIPS DC*, 2003.

[3] L. Cronbach. *Essentials of psychological testing (3rd ed.).* Harper and Row, New York, 1970.

[4] E. Entin and D. Serfaty. Adaptive team coordination. *Human Factors*, 41:312–325, 1999.

[5] E. Entin, D. Serfaty, and J. Deckert. *Report No TR-648-1.* ALPHATECH, Burlington, MA, 1994.

[6] R. R. Espinoza, S. Lemaignan, E. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken. Which one? grounding the referent based on efficient human-robot interaction. In *Proceedings of Ro-Man*, 2010.

[7] P. Hinds, R. Roberts, and H. Jones. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19:151–181, 2004.

[8] G. Hoffman and C. Breazeal. Cost-based anticipatory action-selection for human-robot fluency. *IEEE Transactions on Robotics and Automation*, 23(5):952–961, 2007.

[9] R. Huang and C. Ying. Ant colony system for job shop scheduling with time windows. *The International Journal of Advanced Manufacturing Technology*, 39(1):151–157, 2008.

[10] S. Lemai and F. Ingrand. Interleaving temporeal planning and execution in robotics domains. In *Proceedings of AAAI*, 2004.

[11] A. Lockerd and C. Breazeal. Tutelage and socially guided robot learning. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3475–3480, 2004.

[12] C. Mackenzie, Y. Xiao, and R. Horst. Video task analysis in high performance teams. *Cognition, Technology, and Work*, 6:139–147, 2004.

[13] N. Muscettola, P. Morris, and I. Tsamardinos. Reformulating temporal plans for efficient execution. In *Proceedings of KRR-98*, 1998.

[14] N. Muscettola, P. Nayak, B. Pell, and B. Williams. Remote agent: To boldly go where no ai system has gone before. *Artificial Intelligence*, 103(1):5–48, 1998.

[15] K. Sakita, K. Ogawara, S. Murakami, K. Kawamura, and K. Ikeuchi. Flexible cooperation between human and robot by interpreting human intention from gaze information. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 846–851, 2004.

[16] E. Salas, J. Fowlkes, R. Stout, D. Milanovich, and C. Prince. Does crm training improve teamwork skills in the cockpit?: Two evaluation studies. *Human Factors*, 41:326–343, 1999.

[17] N. Sebanz, H. Bekkering, and G. Knoblich. Joint action: Bodies and minds moving together. *Trends in Cognitive Science*, 10(2):70–76, 2006.

[18] J. Shah. *Fluid Coordination of Human-Robot Teams.* MIT PhD Thesis, Cambridge, Massachusetts, 2010.

[19] J. Shah and C. Breazeal. An empirical analysis of team coordination behaviors and action planning with application to human-robot teaming. *Human Factors*, 52, 2010.

[20] J. Shah, P. Conrad, and B. Williams. Fast distributed multi-agent plan execution with dynamic task assignment and scheduling. In *Proceedings of ICAPS-09*, 2009.

[21] J. Shah, J. Stedl, B. Williams, and P. Robertson. A fast incremental algorithm for maintaining dispatchability of partially controllable plans. In *Proceedings of ICAPS-07*, 2007.

[22] C. Sidner, C. Lee, C. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1):140–164, 2005.

[23] S. Smith, A. Gallagher, T. Zimmerman, L. Barbulescu, and Z. Rubinstein. Multi-agent management of joint schedules. In *AAAI Spring Symposium on Distributed Plan and Schedule Management*, 2006.

[24] R. Stout, J. Cannon-Bowers, E. Salas, and D. Milanovich. Planning, shared mental models, and coordinated performance: an empirical link established. *Human Factors*, 41:61–71, 1999.

[25] J. Trafton, N. Cassimatis, M. Bugajska, D. Brock, F. Mintz, and A. Schultz. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics*, 35(4):460–470, 2005.

[26] I. Tsamardinos and M. Pollack. Flexible dispatch of disjunctive plans. In *Proceedings of ECP*, 2001.

[27] I. Tsamardinos and M. Pollack. Efficient solution techniques for disjunctive temporal reasoning problems. *Artificial Intelligence*, 151(1):43–90, 2003.