Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams

Noname manuscript No.
(will be inserted by the editor)

Decision-Making Authority, Team Efficiency and Human Worker Satisfaction in Mixed Human-Robot Teams

Matthew C. Gombolay · Reymundo A. Gutierrez · Shanelle N. Clarke · Giancarlo F. Sturla · Julie A. Shah

Received: date / Accepted: date

Abstract In manufacturing, advanced robotic technology has opened up the possibility of integrating highly autonomous mobile robots into human teams. However, with this capability comes the issue of how to maximize both team efficiency and the desire of human team members to work with these robotic counterparts. To address this concern, we conducted a set of experiments studying the effects of shared decision-making authority in human-robot and human-only teams. We found that an autonomous robot can outperform a human worker in the execution of part or all of the process of task allocation ($p < 0.001$ for both), and that people preferred to cede their control authority to the robot ($p < 0.001$). We also established that people value human teammates more than robotic teammates; however, providing robots authority over team coordination more strongly improved the perceived value of these agents than giving similar authority to another human teammate ($p < 0.001$). In *post-hoc* analysis, we found that people were more likely to assign a disproportionate amount of the work to themselves when working with a robot ($p < 0.01$) rather than human teammates only. Based upon our findings, we provide design guidance for roboticists and industry practitioners to design robotic assistants for better integration into the human workplace.

Keywords Human-Robot Teaming · Planning and Scheduling · Team Performance · Human-Robot Interaction

Matthew C Gombolay
77 Massachusetts Avenue, Cambridge, MA 02139
E-mail: gombolay@csail.mit.edu

Reymundo A. Gutierrez
77 Massachusetts Avenue, Cambridge, MA 02139
E-mail: ragtz@mit.edu

Shanelle N. Clarke
77 Massachusetts Avenue, Cambridge, MA 02139
E-mail: shang418@mit.edu

Giancarlo F. Sturla
77 Massachusetts Avenue, Cambridge, MA 02139
E-mail: sturlag@mit.edu

Julie A. Shah
77 Massachusetts Avenue, Cambridge, MA 02139
E-mail: julie_a_shah@csail.mit.edu

1 Introduction

There is a growing desire within the manufacturing field to leverage the unique strengths of humans and robots to form highly effective human-robot teams [11,35,36]. Robots are often not capable of performing the same tasks as their human counterparts; and, upon the introduction of a robot worker into their environment, human workers often shift their focus toward the performance of a smaller set of tasks that are better suited for human dexterity and intelligence. The proper functioning of a human-robot manufacturing team requires strict coordination between human and robotic work that satisfies hard temporal and spatial constraints. Academic researchers and industry practitioners alike have developed systems for the planning or scheduling of human and robot work, where the humans are either included in the decision-making process [11,13,45] or the work is scheduled autonomously [2,7]. In this work, we experimentally investigate whether decision-making authority over how best to allocate work should be shared between robots and human workers in order to maximize both human-robot team fluency and human worker satisfaction.

Human workers often develop a sense of identity and security from their roles or jobs, and many are used to some degree of decision-making autonomy. As a result, a human worker may feel devalued when tasked by an automated scheduling algorithm. Even if the algorithm increases process efficiency at first, taking control away from human workers may alienate them and, in turn, ultimately damage overall productivity. On the other hand, workers may find the process of scheduling to be burdensome, and prefer to be part of an efficient team rather than have a role in the scheduling process, if maintaining such a role decreases their efficiency. While autonomous scheduling algorithms can provide near-optimal schedules within seconds, we also want to determine how much decision-making authority humans should have in the task allocation process, so that they continue to feel appreciated while still maintaining a high level of team efficiency.

We conducted a set of experiments studying the effects of shared decision-making authority within human-robot and human-only teams. In our first experiment, we studied the effects of allowing human and robotic teammates more or less decision-making authority over *task allocation*, or the assignment of which worker will perform which task. We hypothesized that human workers would be less capable of coordinating a human-robot team than an autonomous scheduling algorithm. Secondly, we posited that a person's desire to work on a team would be greatest when that person had some control over their role on the team. Rather than finding a desired "middle ground" between fully autonomous and manual scheduling scenarios, we observed statistically significant evidence that allowing human subjects greater authority over task allocation negatively influenced both team fluency ($p < .02$) and the desire of the subject to work with the robot again ($p < 0.001$). In *post-hoc* analysis, we also found evidence of a desire for subjects to monopolize the robot's time, as well as a tendency for the subjects to take a disproportionate amount of the work onto themselves.

In our second experiment, we tested the effects of shared decision-making authority over *task allocation* between human teammates. By collecting data for subjects working within a human-only team, we could study how human behavior might differ when working with another human compared with a robotic assistant. We hypothesized that people would value their human assistant more, even if the human and robotic teammates served the same role on the team, but that a robotic teammate that demonstrated novel scheduling capability would have a stronger positive impact on subjects' perceptions of the value of the robot than a human. Both of these hypotheses were supported by the resulting data.

From our findings across both experiments, we are able to provide guidelines for how to design effective robotic teammates to work alongside human counterparts (Section 6).

2 Background

The development of effective human-machine systems has been the focus of research for many within the fields of human factors, robotics, manufacturing and aerospace, among others. A key goal of this work has been to leverage the unique strengths of both the human and robot. Researchers have defined an effective robot teammate as one that permits humans to choose their own actions and the timing of those actions on the fly, dynamically anticipates and adapts to these decisions, and supports fluid interaction that feels natural to the human [8,14,31,39,42].

However, the human-robot interface has long been identified as a major bottleneck for the utilization of these robotic systems to their full potential [9]. Significant research efforts have been aimed at enabling easier use of these systems in the field, including the careful design and validation of supervisory and control interfaces [4,12,18,24,26]. For example, Barnes *et al.* investigated mixed-autonomy for robot soldiers, with a focus on various issues of interface design. They found that operators were not efficient at the concurrent use of more than one asset. Furthermore, operators performed worse when controlling semi-autonomous ground vehicles than when directly teleoperating those assets [4].

Goodrich *et al.* studied human-robot interaction in the context of wilderness search and rescue. They presented several lessons learned from field trials using camera-equipped UAVs, and observed that the ability to effectively search for persons lost in remote areas was hampered by an inability to clearly see targets in the video obtained from the UAVs, as well as inadequate representation of the quality and progress of the search by the visualization tools.

Jones *et al.* investigated robotic agents assisting in law-enforcement – specifically, Special Weapons and Tactics (SWAT) operations. In these operations, a SWAT team either has a practically unlimited time to plan before responding to a situation, or must react in real-time to environmental uncertainties and actions taken by adversaries. For this study, researchers identified the need for robots to be able to effectively plan in both situations, but found that SWAT members were reluctant to introduce uncertainty into the environment through the use of the robot. Thus, roboticists must develop robots and human-robot interfaces that allow for a robot to be a valuable, dependable member of a team in order to achieve effective utilization [26].

Recently, the DARPA Robotics Challenge provided an opportunity for the robotics community to develop and test real-world, human-machine systems [15, 21, 25, 32, 41]. The teams participating in the challenge deployed robots to perform a series of tasks in a mock disaster-response scenario. Murphy [32] reported that teams seemed to struggle with whether deploying robots with task-level autonomy was practical or even feasible. Where possible, the teams would incorporate autonomy into the system and, when that autonomy would fail, the system would revert back to manual control [32]. The dominant strategy employed was *execution approval*, where only a small set of actions was reviewed and approved by the supervisor at each step in the process [32]. Murphy identified the lack of effort and expertise in user-interface design and the number of operators required to control the robot as two key areas of concern [32].

Related research efforts have focused on the inclusion of a human in the decision-making loop to improve the quality of task plans and schedules for robots or semi-autonomous systems [3, 11, 12, 19, 20, 30, 45]. This is particularly important in the event that the human operators have knowledge of factors not explicitly captured by the system model, or if scheduling decisions may have serious consequences. Studies of these cases have supported preserving the human in the decision-making loop by demonstrating that veteran operators were sometimes able to use heuristics to quickly generate an efficient plan that outperformed optimization algorithms [38].

These works aim to leverage the strengths of both humans and machines in scheduling by soliciting user input in the form of quantitative, qualitative, hard or soft constraints over various scheduling options. For example, Ardissono *et al.* presented a mixed-initiative scheduler that supports users to organize and revise their calendars. Their system relies on common temporal reasoning techniques to provide possible schedules, and even suggest ways to improve the schedule by changing the timing of tasks [3].

Similarly, Zhang *et. al* developed a collaborative planning system called Mobi. This system is a form of "crowdware", which relies on feedback from crowd participants in order to improve schedule quality. Mobi presents the current schedule to participants, who then provide feedback on which activities are to be performed, along with any temporal constraints that relate to those activities. Mobi is able to iteratively generate schedules as more constraints are added or revised, and can provide guidance for the user when certain constraints cannot be satisfied [45].

Berry *et al.* went a step further, and introduced a personalized meeting scheduling assistant that attempts to learn users' preferences for scheduling meetings within an office environment. This system allows users to input activity requests and their preferences for when those activities occur using a restricted natural language and direct manipulation of the schedule. Berry *et al.* employed general-purpose machine learning algorithms to predict users' preferences in order to improve interaction with the scheduling assistant [6].

Supervisory systems have also been developed to assist human operators with coordination of the activities of four-robot or eight-robot teams [10]. Experiment results demonstrated that operators were less able to detect key surveillance targets when controlling a larger number of robots. Similarly, other studies have investigated the perceived workload and performance of subjects operating multiple mobile robots [1]. Results indicated that the presence of more than two robots greatly increased perceived workload and decreased the performance of human subjects.

Some industry practitioners, however, have taken a different approach to scheduling within human-robot teams. For example, when fulfilling online orders in warehouses, workers must navigate the warehouse to find and collect the correct items, then return to the packaging area to complete the order. Kiva Systems has developed robots that are able to fetch items for the worker and ensure that each worker is never idle while waiting for the next item to package. One might initially think that narrowing the role of workers in a factory setting may cause them to feel less important; however, CNN has reported that "robots make for a more pleasant work environment" because they "eliminate much of the mundane physical labor employees once did to retrieve products off shelves." [2].

In this work, we are motivated by the application of robotics in the manufacturing domain, where human workers will perform physical tasks in coordination with robotic partners. In some cases, the human workers may also be responsible for tasking the team and tracking progress. We seek to understand how much control human workers should have over the assignment of roles and schedules when working in teams that include robots. We specifically investigate the role of a robotic teammate as opposed to a virtual scheduling assistant. The embodiment of artificial intelligence in a phyiscal platform has been shown to have an important effect on the performance of the human-robot team and the desire of a person to work with the system [28, 29, 37, 43].

The following sections will describe our experiment to lend insight into the relationship between team efficiency and worker satisfaction, as a function of the control authority possessed by human workers over team scheduling, and the effects of sharing that control with human versus robotic teammates.

3 Aim of the Experiment

We sought to understand the contributions of efficiency, worker decision-making authority and human idle time to objective and subjective measures of team performance and worker satisfaction. Understanding the relationship between these measures will provide researchers and industry practitioners with better insight into how to design successful human-robot teams.

3.1 Experiment 1: Human-Robot Team

In our first experiment, we controlled the level of decision-making authority over task allocation that the subject has during work scheduling for their team. This independent variable can have one of three values:

1. *Manual control* - The subject decides who will perform which tasks
2. *Semi-Autonomous control* - The subject decides which tasks he will perform, while the robot allocates the remaining tasks to itself and a human assistant
3. *Autonomous control* - The robot allocates all tasks.

The robot performed task sequencing in all three conditions. We explored decision-making authority over task allocation alone, rather than over both task allocation and sequencing, in order to isolate the effects of task allocation and mitigate experimental confound. We leave the investigation of sequencing and joint task allocation and sequencing to our future experimentation.

3.2 Experiment 2: Human-Only Team

While it is critical to understand how the level of decision-making authority over task allocation affects team fluency and worker satisfaction within a human-robot team, we must also determine whether these effects are common to teams consisting solely of humans, or if there is an intrinsic difference when working with a robotic teammate. To compare the effects of team composition, we conducted the same experiment as depicted above, but with a human-only team.

3.3 Hypotheses

H1 *Team productivity degrades when the subject has more control over the rescheduling process. As a metric of productivity, we measured both the time it takes to reschedule and the time it takes to finish all tasks.*
Determining the optimal schedule while under hard upper- and lowerbound temporal constraints is NP-Hard [7]. Even for problems of a modest size, optimal scheduling becomes intractable under these circumstances. While we have seen a great deal of work in the development of supervisory control interfaces and human in-the-loop systems to leverage the strengths of human insight and the computational power of autonomous scheduling algorithms ([4, 11, 13, 18, 26]), we expected a near-optimal scheduling algorithm to generate better schedules than those generated by human subjects.

H2 *Subjects prefer having partial control over the rescheduling process to complete control, and prefer having complete control to no control. We utilized a series of subjective Likert-scale questions to determine which level of control the subjects preferred.*
We posited that allowing subjects decide which tasks they would perform and having the robot complete the remainder of the rescheduling would be most satisfying for the subject. In this scenario, a subject can select their preferred tasks according to perceived physical and mental demands, and has a more substantial role in the success of their team. On the other hand, we expected that giving subjects the responsibility of quickly and optimally rescheduling all work would be overwhelming and least desirable. Furthermore, allowing the robot complete control would improve team fluency, but at the cost of possibly devaluing the role of the subject. Interestingly, as we discuss in Section 5, the data did not support this hypothesis. Instead, a *post-hoc* analysis with a Bonferroni correction showed that subjects preferred the robot having complete control over task allocation decisions.

H3 *Subjects are more satisfied with their experience working on the team when they are less idle. To test this hypothesis, we utilized timing information about task execution during the assembly process and the same set of subjective Likert-scale questions used to test Hypothesis 2.*
Many studies have used idle time as a proxy for team fluency [23, 34, 40], and we posited that subjects' satisfaction would be negatively correlated with idle time.

H4 *Subjects are more satisfied and perceive the team as more fluent when working with a human-only team rather than a human-robot team.*
In our experimental design, we afforded the human and robot co-leaders identical functions and capabilities, by

restricting the human co-leaders' capabilities such that they were the same as those of the robot. We nonetheless hypothesized that subjects would value working with a human co-leader more highly than working with a robot co-leader.

H5 *Providing robots authority over team coordination more strongly improves the perceived value of these agents than giving similar authority to another human teammate.*

Humans are intelligent agents capable of advanced reasoning, dexterous manipulation, and multi-modal communication. On the other hand, researchers are still learning how to endow robots with these capabilities. We anticipate that giving a robotic teammate authority over task allocation provides a significant, novel advancement with respect to people's expectations of a robotic agent. We hypothesize that subjects' views on the perceived value of the robot co-leader will vary more substantially with the robot's increased authority over task allocation, as compared to the perceived value of a human teammate.

4 Experimental Methods

We designed an environment analogous to that of a manufacturing setting. In our experiment, the subject is identified as a member of a manufacturing team responsible for completing a set of tasks that includes the fetching and building of part kits. For each trial, the team must schedule and complete this set of tasks. The goal was to assemble various components of a Lego kit, as shown in Figure 1. A video describing the experiment is available at `http://tiny.cc/k4hzgx`.

4.1 Materials and Setup

We used a Willow Garage PR2 platform, depicted in Figure 2, as the robotic co-leader for our human-robot team. Relevant to the experiment, the PR2 has a holonomic base with optical encoders for each wheel and a 270 deg Hokuyo laser at the base. We mapped the laboratory using this laser, as well as the standard *Gmapping* package in the Robot Operating System (ROS). For navigation, we used the Adaptive Monte Carlo Localization (AMCL) [16] probabilistic localization package and a hybrid-dynamical *proportional-derivative* (PD) controller. The locations of the pick-up and drop-off locations for each part kit were hard-coded into the robot controller. The inspection component of the fetching task was simulated.

Fig. 1: The Assembled Lego Model.

4.2 Human-Robot Team Composition

In the first experiment, our human-robot manufacturing team consisted of the human subject (the leader), a robotic co-leader and a human assistant. The human-only manufacturing team for the second experiment consisted of the human subject (the leader), a human co-leader and a human assistant. The subject was capable of both fetching and building, while the co-leader was only capable of fetching. The role of the human assistant, who was capable of both fetching and building, was played by one of the experimenters in both scenarios. This human assistant was included to more realistically represent the composition of a multi-member team. The subjects either performed task allocation alone or shared decision-making authority over task allocation with the co-leader, depending on the experimental condition. The co-leader was always responsible for sequencing the team's work. The human assistant did not aid in the task allocation or sequencing process.

4.3 Experiment Task

In our scenario, the fetching of a kit required walking to one of two inspection stations where the kits were located, inspecting the part kit and carrying it to the build area (depicted in Figure 3). The architecture of our fetching task is analogous to what is required in many manufacturing domains: In order to adhere to strict quality assurance standards, fetching a part kit requires verification from one to two people that all correct parts are in the kit, along with certification from another person that this verification has been performed. Building a part kit required putting together the parts in the kit (depicted in Figure 4.

There were a number of constraints imposed on the analog assembly process in order to model relevant constraints encountered during assembly manufacturing: First, a part kit must have been fetched before it could

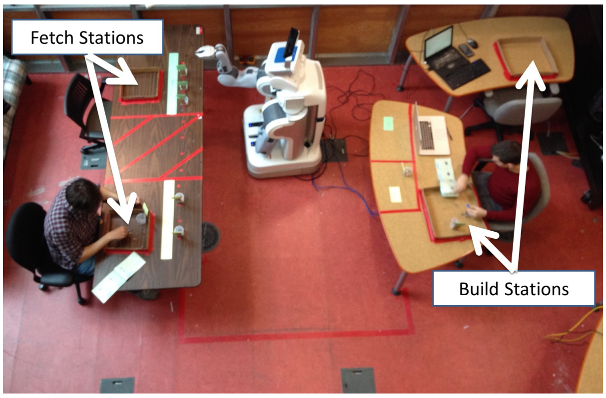


Fig. 2: This figure depicts the laboratory room where the experiment took place. There are two locations where the human and robot workers can inspect part kits during a fetching task, and two locations where the human workers can assemble the part kits.

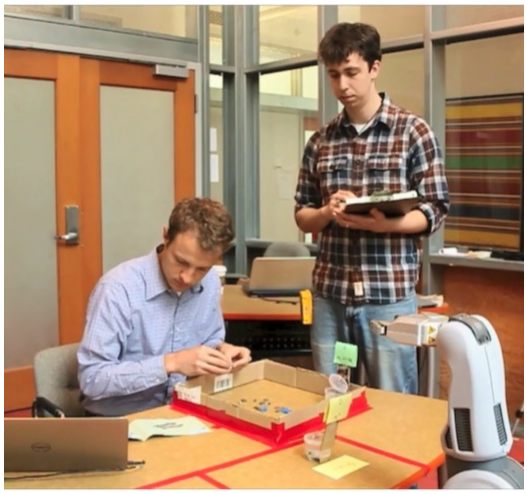


Fig. 4: This figure shows the participant putting together the pieces of a kit as part of the build process. An experimenter is shown here instructing the subject on how to build the assembled lego model from the kit before the experimental trials begin.

be built. Also, no two agents were able to occupy the same fetching or build station simultaneously. There were two fetching stations and two build stations, as shown in Figure 2, with four part kits located at each fetching station. When fetching a part kit, inspection of that kit must have been performed at the station where it was initially located.

Fig. 3: This figure shows the participant verifying that one of the kits has all of the necessary parts and putting the parts one-by-one back into the part kit. This inspection is part of the fetching process. After verifying that the correct parts are located within the kit, the participant will bring the kit to the build stations. An experimenter is shown here instructing the subject on how to fetch a part kit before the experimental trials begin.

Because there were an equal number of building stations and agents able to build, there were no additional constraints imposed exclusively on build tasks. However, because there were three agents able to fetch but only two fetching stations, the agents were required to take turns using these stations. Allowing workers to sort through parts from multiple kits at the same location risked the mixing of parts from different kits. We imposed a 10-minute deadline from the time that the fetching of a part kit began until that part kit had been built, for similar reasons. In manufacturing, if a part or part kit is missing from an expected location for too long, work in that area of the factory will temporarily halt until the missing pieces are found.

4.4 Formulation of the Human-Robot Scheduling Problem

Assembly of the Lego model involved eight tasks $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_8\}$, each of which was composed of a *fetch* and *build* subtask $\tau_i = \{\tau_i^{fetch}, \tau_i^{build}\}$. The time each subject took to complete each subtask $C_i^{subject-fetch}$ and $C_i^{subject-build}$ was measured during an experiment training round. The times necessary for the human assistant $C_i^{assist-fetch}$ and $C_i^{assist-build}$ tasks were measured prior to the experiments. To provide a fair comparison, the human co-leader in the human-only team worked at the same constant pace as the robot co-leader for each subtask $C_i^{co\text{-}leader-fetch} = C_j^{co\text{-}leader-fetch} =$ 120 seconds, $\forall i, j$. Because the robot required significantly more time to complete the fetching task than the average person, we justified the longer duration to the subjects by explaining that the experimenter was asking both the human and robotic co-leaders to perform an extra auditing process on the part kits and would therefore take longer than normal.

Constraints on lowerbound completion time of tasks are presented in Equations 1-5

$$f_i^{build} - s_i^{build} \geq C_i^{subject-build} - M(1 - A^{subject}_{\tau_i^{build}}) \quad (1)$$

$$f_i^{fetch} - s_i^{fetch} \geq C_i^{subject-fetch} - M(1 - A^{subject}_{\tau_i^{fetch}}) \quad (2)$$

$$f_i^{fetch} - s_i^{fetch} \geq C_i^{co\text{-}leader-fetch} - M(1 - A^{co\text{-}leader}_{\tau_i^{fetch}}) \quad (3)$$

$$f_i^{build} - s_i^{build} \geq C_i^{assist-build} - M(1 - A^{assist}_{\tau_i^{build}}) \quad (4)$$

$$f_i^{fetch} - s_i^{fetch} \geq C_i^{assist-fetch} - M(1 - A^{assist}_{\tau_i^{fetch}}) \quad (5)$$

where $A^{agent}_{\tau_i^{subtask}}$ is a binary decision variable for the assignment of $agent \in \{subject, assist, co\text{-}leader\}$ to each $subtask \in \{fetch, build\}$ of $\tau_i \in \boldsymbol{\tau}$. Variables s_i^{build}, s_i^{fetch}, f_i^{build}, and f_i^{fetch} are the start and end times of the build and fetch subtasks, respectively. M is a large constant that allows for constraints to be selectively enforced[5].

Constraints 6 and 7 ensured that each agent performed only one subtask at a time.

$$s_x^y - f_i^j \geq -M\left(1 - x_{\langle \tau_i^j, \tau_x^y \rangle}\right) - M\left(2 - A^{agent}_{\tau_i^j} - A^{agent}_{\tau_x^y}\right), \forall \tau_i^j, \tau_x^y \in \boldsymbol{\tau} \quad (6)$$

$$s_i^j - f_x^y \geq -M x_{\langle \tau_i^j, \tau_x^y \rangle} - M\left(2 - A^{agent}_{\tau_i^j} - A^{agent}_{\tau_x^y}\right), \forall \tau_i^j, \tau_x^y \in \boldsymbol{\tau}, \quad (7)$$

where $x_{\langle \tau_i^j, \tau_x^y \rangle} \in \{0, 1\}$ is a binary decision variable specifying whether τ_i^j comes before or after τ_x^y.

Temporal constraints 8 and 9 ensured that the parts necessary for each task were fetched before the building process began, and that assembly was completed within $D = 10$ minutes of fetching the parts.

$$\infty \geq s_i^{build} - f_i^{fetch} \geq 0, \forall \tau_i \in \boldsymbol{\tau} \quad (8)$$

$$D \geq f_i^{build} - s_i^{fetch} \geq 0, \forall \tau_i \in \boldsymbol{\tau} \quad (9)$$

The spatial constraint in Equation 10 ensured that no two agents occupied the same fetching station at the same time.

$$s_j^{fetch} - f_i^{fetch} \geq 0 \ \vee \ s_i^{fetch} - f_j^{fetch} \geq 0, \quad \forall \tau_i, \tau_j \in \boldsymbol{\tau} \text{ s.t. } R_i^{fetch} = R_j^{fetch} \quad (10)$$

where R_k^{fetch} denotes the physical floor area reserved for the fetching subtask τ_k^{fetch}. Fetching subtasks $\{\tau_i^{fetch} | i \in \{1, 2, 3, 4\}\}$ incorporated the first inspection station R_1^{fetch}, while $\{\tau_{i'}^{fetch} | i' \in \{5, 6, 7, 8\}\}$ used the second inspection station R_2^{fetch}.

Finally, one pair of tasks, $\boldsymbol{\tau_{two-step}} = \langle \tau_3, \tau_4 \rangle$, was related through a precedence constraint. Specifically, participants were instructed that the first task in the pair, τ_3, be completed before beginning to fetch parts for the second task, τ_4. This constraint is presented in Equation 11.

$$s_j^{fetch} - f_i^{build} \geq 0, \forall \langle \tau_i, \tau_j \rangle \in \boldsymbol{\tau_{two-step}} \quad (11)$$

The objective of the problem (Equation 12) was to minimize the maximum amount of work assigned to any one agent while satisfying the constraints in Equations 1 - 11.

$$obj = \arg\min \left(\max_{agent} \left(\sum_{\tau_i} \left(\sum_{subtask} C^{agent}_{\tau_i^{subtask}} A^{agent}_{\tau_i^{subtask}} \right) \right) \right) \quad (12)$$

Figure 6a shows the set of tasks and associated temporal constraints as a Simple Temporal Network [33]. Nodes represent events (e.g., s_2^{fecth} is the start of subtask τ_2^{fetch}), and edges represent interval temporal constraints (e.g., an edge $[0, d]$ between s_2^{fetch} and f_2^{build} means that subtask τ_2^{build} must finish between 0 and d units of time after the start of τ_2^{fetch}). To schedule the task set, one must assign agents to subtasks and order the nodes to minimize the makespan such that each $agent \in \{subject, assist, co-leader\}$ performs no more than one subtask at a time, and resources R_1^{fetch} and R_2^{fetch} are used by at most one agent at a time. This network shows subtask durations as a variable C_i^j (e.g., C_1^{build}). When an agent is assigned to a subtask, the duration of the subtask is set to the duration required by that agent to complete that subtask ($e.g., C_1^{build}$ assigned to *subject* would become $C_1^{subject-build}$). Lastly, node s is simply a reference point to be scheduled at $t = 0$.

4.5 Human-Robot Coordination

Subjects were provided the expected time necessary for each agent to complete each of the 16 subtasks under conditions where the subject performed the task allocation. For the manual condition, subjects specified the assignment of $agent \in \{subject, assist, co\text{-}leader\}$ to each $subtask \in \{fetch, build\}$ of $\tau_i \in \boldsymbol{\tau}$ by writing the assignment list on a blank paper. In the semi-autonomous condition, subjects selected only the subtasks that they would complete themselves.

Under the manual condition, the experimenter provided the subjects' selections to the human or robot co-leader, and the co-leader would then sequence all of the subtasks. In the semi-autonomous condition, the experimenter provided the subjects' selections to the human or robot co-leader, who would then divide the remaining subtasks between the co-leader and the human assistant and sequence all of the subtasks. Under the autonomous condition, the co-leader allocated and sequenced all of the subtasks.

4.6 Scheduling Mechanism

To enable the co-leader to schedule with varying degrees of decision-making input from the subject, we adapted Tercio, a fast, near-optimal scheduling algorithm that divides the scheduling process into task allocation and sequencing subroutines [17].

As shown in Figure 5, the algorithm takes as input a temporal constraint problem, a list of agent capabilities (i.e., the lowerbound, upperbound and expected duration for each agent performing each task) and the physical location of each task. Tercio first solves for an optimal task allocation by ensuring that the minimum amount of work assigned to any agent is as large as possible, as depicted in Equation 13. In this equation, $\boldsymbol{Agents}$ is the set of agents, $A^a_{\tau_i^j}$ is a task allocation variable that equals 1 when agent a is assigned to subtask τ_i^j and 0 otherwise, $\boldsymbol{A}$ is the set of task allocation variables, $\boldsymbol{A^*}$ is the optimal task allocation and $C^a_{\tau_i^j}$ is the expected time it will take agent a to complete subtask τ_i^j.

$$\boldsymbol{A^*} = \max_{\{\boldsymbol{A}\}} \min_{\boldsymbol{Agents}} \sum_j A^a_{\tau_i} \times C^a_{\tau_i}, \forall a \in \boldsymbol{Agents} \tag{13}$$

After determining the optimal task allocation, $\boldsymbol{A^*}$, Tercio uses a fast sequencing subroutine to complete the schedule. The sequencer orders the tasks through simulation over time. Before each commitment is made, the sequencer conducts an analytical schedulability test to determine whether task τ_i can be scheduled at time t given prior scheduling commitments. If the schedulability test returns that the commitment can be made, the sequencer then orders τ_i and continues. If the schedulability test cannot guarantee commitment, the sequencer evaluates the next available task.

If the schedule, consisting of a task allocation and a sequence of tasks, does not satisfy a specified makespan, a second iteration is performed by finding the second-most optimal task allocation and the corresponding sequence. The process terminates when the user is satisfied with the schedule quality, or when no better schedule can be found. In this experiment, we specified that Tercio run for 25 iterations and return the best schedule.

In the first experiment, where the subject worked with a robot co-leader, the robot used Tercio to perform task allocation and sequencing. In the scenario where the subject performed task allocation, the robot used Tercio to sequence tasks and return a flexible, dispatchable schedule [33]. When the subject decided which tasks he or she would perform, the robot used Tercio to find an efficient schedule by iterating over different allocations to the robot and the human assistant. Tercio receives the upperbound, lowerbound, and expected duration of each task, and uses the expected durations to compute near-optimal schedules. The upperbound and lowerbound times are used for computing a flexible, dispatchable schedule to allow subjects to work faster or slower than expected, if necessary. We set the lowerbound duration of subtasks assigned to the subject to be 25% faster than the speed observed during training to mitigate subject idle time due to learning effects.

Figure 6b shows an example solution generated by Tercio. Tercio takes as in put the Simple Temporal Network from Figure 6a. Tercio then generates an assignment of which agents will perform each task and the sequencing of those tasks such that each agent only performs one subtask at a time and each resource is used by only one agent at a time. Rather than generating a set of fixed timepoints for the schedule, Tercio adds ordering constraints to the original Simple Temporal Network to enforce the schedule while preserving flexibility where possible [44].

In the second experiment, where the subject worked with a human co-leader, the human co-leader simulated

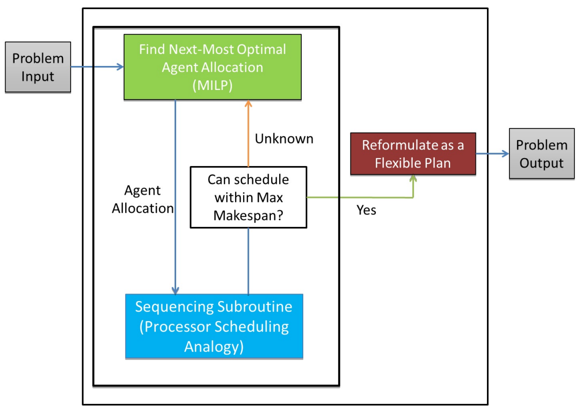
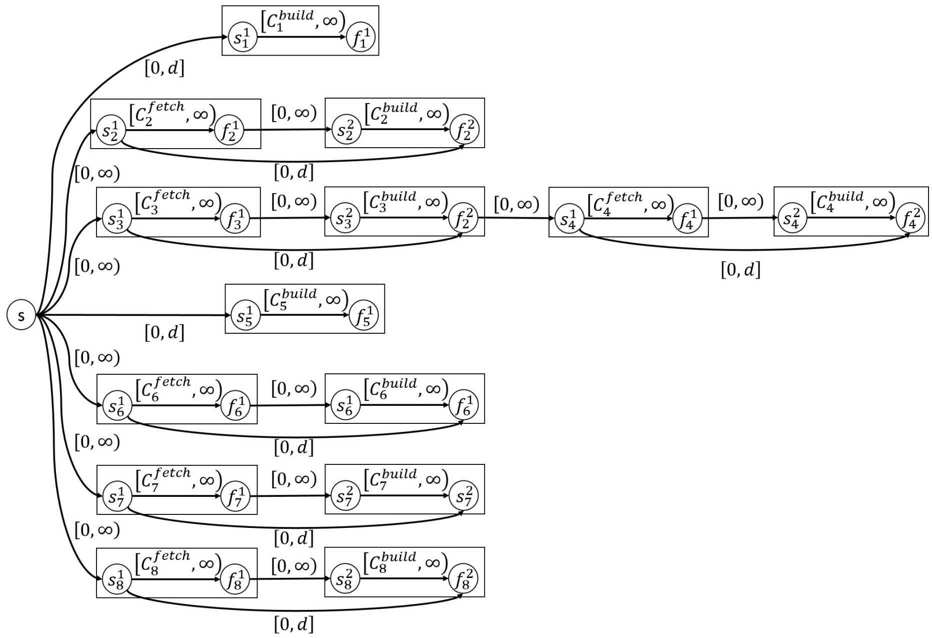
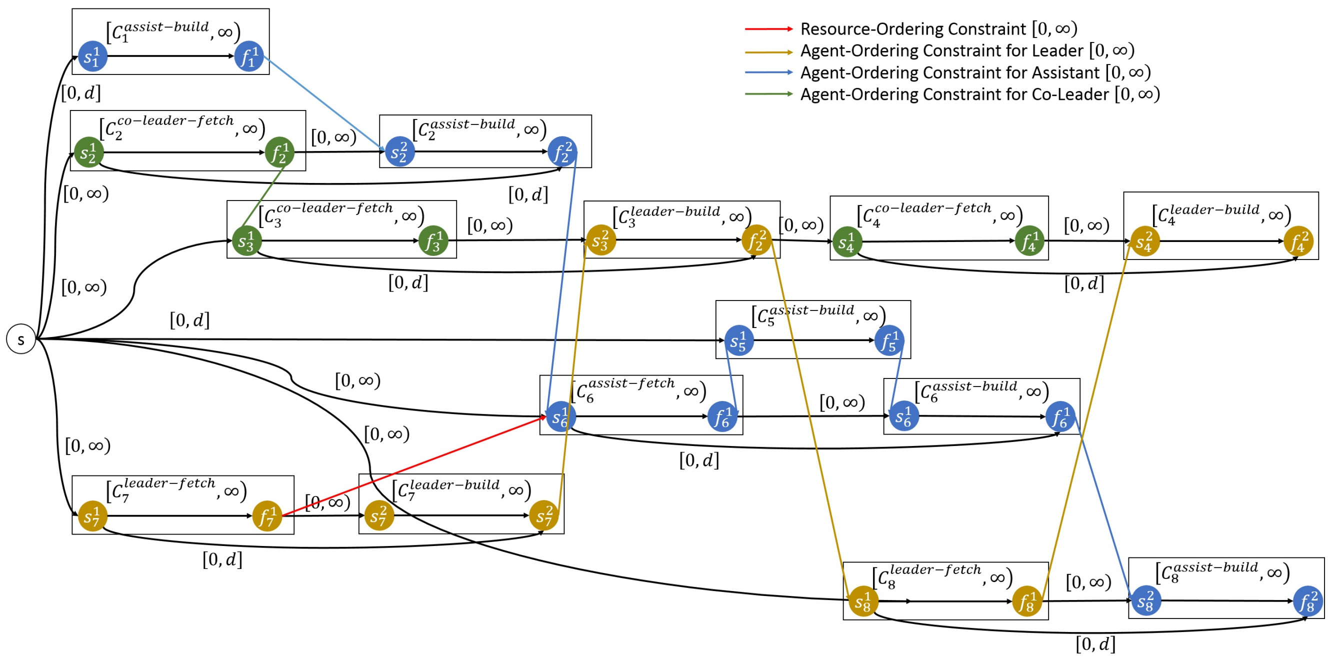


Fig. 5: Tercio takes as input a temporal constraint problem and finds a satisficing, flexible schedule by utilizing an analytical schedulability test to ensure a feasible solution.



(a) This figure shows the unscheduled task set as a Simple Temporal Network.



(b) This figure shows an example of a scheduled task set.

Fig. 6: The robot uses Tercio to schedule the task set shown in Figure 6a. Tercio returns the solution in the form of a flexible, dispatchable schedule shown in Figure 6b.

allocating and sequencing tasks with pen and paper. The human co-leader provided a simulated schedule to an experimenter, who would run Tercio directly instead of using the human co-leader's schedule. This simulation is intended to suggest to the subject that the human co-leader was actually responsible for scheduling, rather than an autonomous scheduling algorithm.

In both experiments, Tercio served as a dispatcher, communicating to the subject, human assistant and co-leader when to initiate their next subtasks. Tercio would tell each agent when they were able to start or finish each subtask, and each agent would send a message acknowledging the initiation or completion of a subtask. The team members communicated this information by sending simple, text-based messages over a TCP/IP GUI[1].

4.7 Procedure

We first introduced the subject to the manufacturing scenario. Each subject was told that they were a member of a manufacturing team. We explained the various temporal and spatial constraints of our analog manufacturing task, as well as the capabilities and roles of each team member.

We then conducted a training round where the subject fetched and built each of the eight part kits. We timed how long it took the participant to complete each task, and provided this information to the co-leader and the subject for use when scheduling the work. Participants were instructed to work as quickly as possible without making mistakes. Next, the experimenter explained the constraints imposed on the assembly process, and the subject was trained on how to communicate with the Tercio dispatcher via the TCP/IP GUI.

We then performed three trials in which the subject was exposed to each of the three conditions (i.e., manual, semi-autonomous or autonomous control), varying the order of the conditions across subjects. Each trial consisted of scheduling the work and completing all tasks according to that schedule. In scenarios where the subject participated in the task allocation process, we provided the subject with information about how long it took each agent to perform each task. The subject was instructed to quickly construct an efficient task allocation with the goal of minimizing the amount of time spent rescheduling and completing the tasks. Subjects took approximately 5 minutes when asked to allocate all of the tasks, and approximately 2 minutes when deciding which tasks to complete themselves.

Either autonomously or according to the task allocation information provided by the subject, the co-leader completed the rescheduling and the assembly process began. Each trial took approximately 15 minutes. After each trial, subjects were asked to answer a post-test questionnaire, which consisted of 21 Likert-scale questions assessing their experience. The experiment concluded with a final post-test questionnaire consisting of three Likert-scale questions and two free-response questions, as shown in Tables 1 and 2.

After reviewing the results of the first experiment, where subjects worked with a robot co-leader, we wanted to ask subjects an additional set of questions to better understand their task allocation strategies, preferences and impressions of the team dynamic. As such, we included an additional post-test questionnaire for subjects in our second experiment, where subjects worked with a human co-leader. The contents of this second post-test questionnaire are depicted in Table 3.

4.8 Experimental Design

The goal of our experiment was twofold. First, we sought to understand the relationship between efficiency and worker satisfaction, as a function of how much control the worker has over his or her own role as part of a human-robot manufacturing team. Second, we wanted to know whether the effects of varying subjects' authority over their own role differed when the subjects worked within a human-robot versus a human-only team.

To test our experimental hypotheses, we employed a 2x3 mixed-factorial design. Our first factor, team composition, was of a between-subjects design; the second factor, how decision-making authority was shared amongst the team members, was within-subjects.

In prior experience, we have observed human subjects assemble Lego models at vastly differing rates. A within-subjects design for varying decision-making authority can help to mitigate the effects of inter-subject variability. However, because of this design, we needed to account for possible learning effects over the different trials, as the speed at which subjects build generally increases with practice. To account for this factor, we balanced the assignment subjects into groups for each of the $k!$ orderings of our $k = 3$ conditions.

To address the potential variability in the characteristics of the human assistant on each team, a laboratory researcher played the role of the human assistant in all trials. This assistant performed tasks at a nearly constant speed, and did not aid the subject in rescheduling the work.

[1] SocketTest v3.0.0 ©2003-2008 Akshathnkumar Shetty (http://sockettest.sourceforge.net/)

4.9 Objective Evaluation

Objective measures of team fluency consist of *assembly time*, *rescheduling time* and *idle time*. “Assembly time” is defined as the difference between the time the last task was completed and the time the first task was initiated. “Rescheduling time” is defined as the sum of the time taken to allocate and sequence tasks. (The experimenter was required to input the task allocation of the subject into the robot’s scheduling algorithm, but we did not include this as part of the rescheduling time.) Lastly, we defined “idle time” as the sum of all periods during which the subject was not working.

4.10 Subjective Evaluation

Subjects received questionnaires after each trial, consisting of 21 Likert-scale questions, as shown in Table 1. Subjects who worked with a human co-leader received the same questionnaire as those who did not, with the word “robot” and the phrase “human worker” each replaced by “human co-leader”. Hoffman proposed a set of composite measures for the evaluation of human-robot fluency [22]. Questions 1-3 corresponded to Hoffman’s measure of *Robot Teammate Traits*, and Questions 4-13 represented Hoffman’s adaptation of the “Working Alliance Index” for human-robot teams, measuring the quality of the alliance amongst the teammates. We added questions 14-21 based on our own insight. Subjects were not informed of their rescheduling and build times during the experiment.

Subjects also received a post-test questionnaire after completing the three trials, as shown in Table 2. This questionnaire collected demographic information, and also included three additional Likert-scale questions summarizing the experience of the subject, along with two open-ended questions. Subjects who worked with a human co-leader received the same questionnaire as those who worked with a robot, with the word “PR2” replaced by “human co-leader,” and the phrase “human teammate” replaced with “human assistant”.

For our second experiment, where subjects worked with a human co-leader, we included a second post-test questionnaire, as depicted in Table 3. The goal of this questionnaire was to better understand how subjects performed task allocation, their views on different team roles and their preferences for performing different types of tasks.

Table 1: Subjective measures - post-trial questionnaire

Robot Teammate Traits
1. The robot was intelligent.
2. The robot was trustworthy.
3. The robot was committed to the task.
Working Alliance for Human-Robot Teams
4. I feel uncomfortable with the robot. (reverse scale)
5. The robot and I understand each other.
6. I believe the robot likes me.
7. The robot and I respect each other.
8. I feel that the robot worker appreciates me.
9. The robot worker and I trust each other.
10. The robot worker perceives accurately what my goals are.
11. The robot worker does not understand what I am trying to accomplish. (reverse scale)
12. The robot worker and I are working towards mutually agreed upon goals.
13. I find what I am doing with the robot worker confusing. (reverse scale)
Additional Measures of Team Fluency
14. I was satisfied by the teams performance.
15. I would work with the robot the next time the tasks were to be completed.
16. The robot increased the productivity of the team.
17. The team collaborated well together.
18. The team performed the tasks in the least time possible.
19. The robot worker was necessary to the successful completion of the tasks.
20. The human worker was necessary to the successful completion of the tasks.
21. I was necessary to the successful completion of the tasks.

5 Results

In this section, we report the demographics of the participants, as well as statistically significant and insightful findings from our experiment. We define statistical significance at the $\alpha = .05$ level.

5.1 Effects of Sharing Decision-Making Authority with a Robot Co-Leader

5.1.1 Participants

Twenty-four participants were included in the experiment. Each subject worked on the human-robot manufacturing team under each level of decision-making authority, in accordance with a within-subjects design. To control for learning effects, participants were balanced between one of six groups, including one group for each of the six possible sequences of the three conditions, and four subjects for each sequence. The participants (14 men and 10 women) had a mean age of 27 ± 7 years (range 20-42) and were recruited via email and fliers distributed around a university campus.

Table 2: Subjective measures - post-test questionnaire # 1

Overall Preference
22. If it was the PR2s job to reschedule the work, I would want to work with the robot again.
23. If it was my job to reschedule my work and the PR2 reschedule the work for the PR2 and my human teammate, I would want to work with the robot again.
24. If it was my job to reschedule the work for myself, my human teammate, and the PR2, I would want to work with the robot again.
Open Response Questions
25. Which of the three scenarios did you prefer and why?
26. If you were going to add a robotic assistant to a manufacturing team, to whom would you give the job of rescheduling the work and why?

Table 3: Subjective measures - post-test questionnaire # 2

Overall Preference
27. I used a different strategy when scheduling when I could only pick which tasks I would perform versus when I could pick which tasks each team member would perform.
28. When I could only pick which tasks I would perform, I tried to isolate my work (or work separately) from the rest of the team.
29. When I could pick which tasks each team member would perform, I tried to isolate my work (or work separately) from the rest of the team.
30. I thought that it would be better if I performed more of the work.
31. The teammate that performs the most work is the most important member of the team.
32. The teammate that performs the scheduling is the most important member of the team.
33. I prefer fetching.
34. I prefer building.

5.1.2 Objective Measures of Human-Robot Team Fluency

We considered the team's assembly time and the subjects' rescheduling time as a function of the subjects' decision-making authority. Recall that hypothesis **H1** predicted that the team would be more fluent, in terms of both assembly and rescheduling time, when the robot had more authority over task allocation. Rescheduling and assembly times are depicted in Figure 7.

Variance analysis demonstrated statistically significant differences in the distribution of rescheduling time as a function of decision-making authority ($F(2,69) = 55.1, p < 0.01$). Rescheduling time under the autonomous condition ($M = 30$, $SD = 0$) was lower than for the semi-autonomous condition ($M = 108$, $SD = 69$; $t(23) = 7.24$, $p < 0.01$ for comparison). Similarly, rescheduling time under the semi-autonomous condition was lower than for the manual condition ($M = 315$, $SD = 154$; $t(23) = 7.23$, $p < 0.01$ for comparison).

Repeated-measure analysis of variances indicated significant differences in assembly time as a function of condition ($F(2,46) = 3.84$, $p = 0.03$). Assembly time under the autonomous condition ($M = 520$, $SD = 60.6$) was faster than under either the semi-autonomous ($M = 564$, $SD = 83.9$; $t(23) = 2.37$, $p = 0.01$ for comparison) or manual conditions ($M = 582$, $SD = 115$; $t(23) = 2.18$, $p = 0.02$ for comparison).

Learning effects on assembly time were observed as a function of trial number (ANOVA $F(2,69) = 3.68$, $p = .03$). Specifically, assembly times during the third trial ($M = 519$, $SD = 85$) were lower than those observed during the first ($M = 585$, $SD = 49$; $t(23)$, $p = .002$ for comparison) and second trials ($M = 567$, $SD = 75$; $t(23)$, $p = .022$ for comparison). Nonetheless, the k–factorial design counterbalanced these learning effects, and results indicated significant differences in both assembly time ($F(2,46) = 3.84$, $p = .03$) and rescheduling time ($F(2,69) = 55.1$, $p < 0.01$) as a function of the level of automation, thereby supporting **H1**.

The participants did not repeat the rescheduling process. Each participant performed task allocation for the entire team in the manual condition, and each participant selected their own tasks in the semi-autonomous condition. It is possible that participants would have been able to perform task allocation more quickly with practice. However, we leave such an investigation of learning effects for future work.

5.1.3 Subjective Measure of Satisfaction as a Function of Task-Allocation Authority

Recall that our second hypothesis **H2** stated that workers would prefer partial authority over the task process

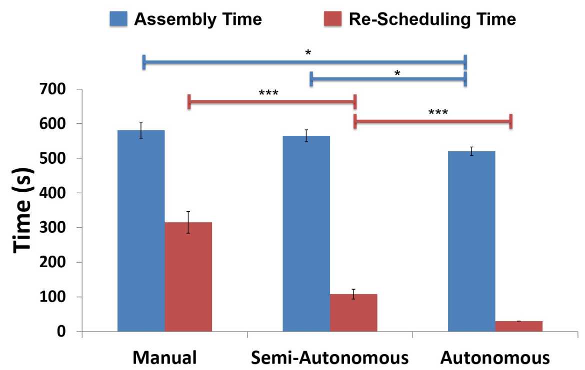


Fig. 7: This figure depicts the average and standard error for the assembly times for each condition.

Table 4: P-values for statistically significant post-trial questions (N=24). Statistically significant values are shown in bold.

Question	Omnibus	Auto v. Man.	Semi v. Man.	Auto. v. Semi.
1	$\boldsymbol{p = 0.033}$	$\boldsymbol{p = 0.007}$	$p = 0.089$	$p = 0.062$
5	$\boldsymbol{p = 0.008}$	$\boldsymbol{p = 0.001}$	$\boldsymbol{p = 0.011}$	$\boldsymbol{p = 0.036}$
10	$\boldsymbol{p = 0.006}$	$\boldsymbol{p < 0.001}$	$\boldsymbol{p = 0.034}$	$p = 0.012$
11	$\boldsymbol{p = 0.006}$	$\boldsymbol{p = 0.001}$	$\boldsymbol{p = 0.023}$	$\boldsymbol{p = 0.008}$
14	$\boldsymbol{p < 0.001}$	$\boldsymbol{p < 0.001}$	$\boldsymbol{p < 0.001}$	$\boldsymbol{p < 0.001}$
15	$\boldsymbol{p = 0.003}$	$\boldsymbol{p < 0.001}$	$\boldsymbol{p = 0.020}$	$\boldsymbol{p = 0.008}$
16	$\boldsymbol{p = 0.012}$	$\boldsymbol{p = 0.002}$	$p = 0.052$	$\boldsymbol{p = 0.016}$
17	$\boldsymbol{p = 0.001}$	$\boldsymbol{p < 0.001}$	$\boldsymbol{p = 0.005}$	$\boldsymbol{p = 0.003}$
18	$\boldsymbol{p = 0.001}$	$\boldsymbol{p < 0.001}$	$\boldsymbol{p = 0.011}$	$\boldsymbol{p < 0.001}$

rather than total control, and that having no control would be preferable to complete control. An omnibus Friedman test confirmed a statistically significant difference in the distribution of a subset of the Likert-scale responses for the three conditions, as shown in Table 4. A pair-wise Friedman test supported our hypothesis that the subjects were more satisfied while under the autonomous and semi-autonomous conditions compared with the manual condition for the questions listed in Table 4.

However, responses to no single question suggested that subjects favored the semi-autonomous condition over the autonomous condition. A post-hoc Friedman test with a requisite Bonferroni correction of $\frac{\alpha}{3}$ indicated that the subjects actually thought more highly of the robot and team interaction while under the autonomous condition rather than the semi-autonomous condition (i.e., questions 5, 11, 14, 15, 16, 17 and 18).

The post-test questionnaire included three questions designed to determine whether subjects would be likely to work with the robot again, given the same levels of decision-making authority allotted to the subject and the robot as had occurred in the experiment. Applying the omnibus Friedman test across Questions 22-24 from Table 2, we observed a statistically significant difference in the subjects' responses ($p < 0.001$). Post-hoc analysis using pair-wise comparison with a Bonferroni correction confirmed that subjects agreed they were more likely to work with the robot again if the robot performed task allocation autonomously, rather than if the subject and robot shared task allocation authority ($p < 0.001$) or the subject had complete task allocation authority ($p < 0.001$). Similarly, subjects were more likely to report that they would work with the robot again if the robot and human shared task allocation authority than if the subject had sole authority over task allocation ($p = 0.001$). Given the strong preference observed for the autonomous condition, we revised hypothesis **H2** for future experiments to state that subjects prefer working with a robot co-leader with more authority over task allocation.

Table 5: Representative open-ended responses from subjects preferring the manual, semi-autonomous, and autonomous conditions.

Manual
"There is something soul-sucking about taking the thinking away from the workers."
Semi-autonomous
"I prefer the scenario where I pick the tasks I want, because some tasks are more fun for me than others ... even if it might slightly increase completion time." "I got to schedule my work and the robot filled in the rest of the schedule with the purpose of optimizing time."
Autonomous
"It removes the possibility of scheduling being influenced by the ego of the team leader." "Comparing times and planning for three agents was a headache for me; the robot did a much better job. The whole operation felt more coordinated." "[It is] easier for the robot to deal with scheduling many complex tasks than it is for a human, because it can consider all at once [without] getting overwhelmed."

5.1.4 Analysis of Open-Ended Responses: Robot Co-Leader

Questions 25 and 26 of the post-trial questionnaire offered subjects the opportunity to provide open-ended responses to prompts regarding which condition they preferred, and to whom they would prefer to give control of task allocation to in a manufacturing setting. While the majority of subjects' responses were supportive of a robotic assistant that autonomously allocated work, we also provide representative responses from subjects who preferred the manual, semi-autonomous and autonomous conditions, as shown in Table 5. While most of the subjects' responses directly supported one of the three experimental conditions, some suggested a "blended" level of control, where the robot would "assign tasks but allow the person to override them (if, for example, they become overwhelmed or bored)."

5.1.5 Subject Idle Time and Satisfaction

Our third hypothesis **H3** stated that subjects would be more satisfied working on a human-robot team when they were less idle. Our post-test questionnaire prompted subjects to rate the degree to which they would want to work with the robot again, depending on the robot co-leader's role in the scheduling process, with one question for each of the three conditions. Both idle time and subject satisfaction were dependent variables in our experiment. The three experiment conditions (autonomous, semi-autonomous and manual) were ranked according to each subject's preference; the condition's rank was then plotted against the corresponding idle time for each subject-condition pair. The Pearson product-moment correlation coefficient of satisfaction and idle time ($r = 0.125$) was not statistically significant ($t(23) = 0.90$, $p = 0.367$).

5.1.6 Post-hoc Analysis

We conducted a *post-hoc* analysis to better understand the differences in the ways that people allocated work while under the various experiment conditions. We observed two important behaviors indicating statistical significance as a function of decision-making authority: the amount of work subjects allocated to themselves, and the number of times that the robot fetched part kits for the subject.

First, we measured the amount of work subjects allocated to themselves relative to the amount of work the robot allocated to subjects, as a function of time. As shown in Figure 8, we found that subjects allocated more work to themselves under both the semi-autonomous ($M = 12\%$, $SD = 5\%$; $t(23) = 2.31, p < .016$) and manual conditions ($M = 11\%$, $SD = 5\%$; $t(23) = 2.30, p < .016$). As such, we established a new hypothesis, **H6**: When subjects have more control over their work, they prefer to take on a disproportionate share of the work. However, this observed trend could be a novelty, and could wear off as the work becomes more routine.

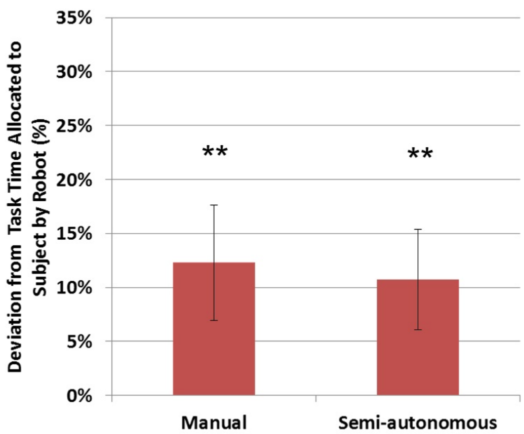


Fig. 8: The average and standard error of the percent difference between the time the subjects were assigned to work under the semi-autonomous and manual conditions relative to the autonomous condition.

Second, we sought to understand how the subjects were tasking the robot. Figure 9 depicts the number of trials in which the robot co-leader fetched more part kits for the human assistant than for the subject. A χ^2 test indicated that the number of trials in which the robot co-leader fetched more for the human assistant than for the subject in the autonomous condition (6 of 24 trials) was significantly different than the number of times the robot fetched for the assistant (20 of 24 trials), $\chi^2 = 5, p < 0.01$. Subsequently, we established a new hypothesis, **H7**: Subjects prefer the co-leader to fetch part kits for them. We note, however, that subjects could have been exhibiting lack of experience in how to best utilize the robotic teammate.

5.2 Comparing the Effects of a Robot versus a Human Co-Leader

In our first round of experiments, we observed statistically significant differences in team fluency and subject satisfaction when the subject worked with a robotic versus a human co-leader. However, we need to understand whether these differences are inherent when sharing decision-making authority over task allocation with any co-leader – whether human or robot – or if there is something intrinsic to a robotic co-leader that influences team dynamics. As such, we conducted a follow-up experiment and report the results here.

5.2.1 Objective Measures of Human-Only versus Human-Robot Team Fluency

We considered the team's assembly time and the subjects' rescheduling time as functions of the subjects'

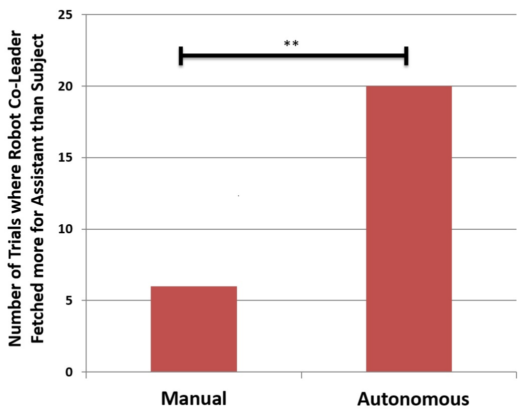


Fig. 9: The number of trials in which the robot co-leader fetched more parts kits for the human assistant than for the subject.

decision-making authority. Recall that hypothesis **H1** predicted that the team would be more fluent, in terms of both assembly and rescheduling time, when the human co-leader had more control authority over task allocation.

As in the first experiment, with a robot co-leader, variance analysis demonstrated statistically significant differences in the distribution of rescheduling time as a function of decision-making authority with a human co-leader ($F(2, 69) = 51.0$, $p < 0.01$). Rescheduling time under the autonomous condition ($M = 30$, $SD = 0$) was lower than for the semi-autonomous condition ($M = 106$, $SD = 60.8$; $t(23) = 6.11$, $p < 0.01$). Likewise, rescheduling time under the semi-autonomous condition was lower than for the manual condition ($M = 220$, $SD = 95.9$; $t(23) = 4.97$, $p < 0.01$). Figure 10 depicts the rescheduling time for both human-robot and human-only teams. Two-factor analysis of variance with replication showed no statistical significance between human-robot and human-only teams ($F(1, 142) = 0.429$, $p = 0.514$).

To compare assembly times, we considered the percent difference between the assembly time in the semi-autonomous and autonomous conditions versus the autonomous condition. This comparison is shown in Figure 11. Two-factor analysis of variance demonstrated significant differences in assembly time, as a function of the level of decision-making authority ($F(2, 46) = 3.72$, $p = .027$) and whether the co-leader was a human or robot ($F(1, 46) = 4.66$, $p = 0.033$). The interaction effect between decision-making authority and team composition was not significant ($F(1, 46) = 4.65$, $p = 0.628$).

Based on our *post-hoc* analysis of the experiment in which subjects worked with a robot co-leader, we established two new hypotheses: **H6** and **H7**. Recall that hypothesis **H6** states that subjects prefer to take on a disproportionate share of the work themselves, and hypothesis **H7** states that subjects are biased to direct the robot co-leader to fetch parts for them. We tested these hypotheses in a scenario where the subjects are working with a human co-leader, in order to determine whether the observed trends persist.

We performed an omnibus z-test for one mean to test hypothesis **H6**, and observed no statistically significant evidence supporting it ($M = -1.7\%, SD = 25.1\%$; $z = -0.388, p = 0.654$). Performing a post-hoc analysis with a Bonferroni correction $\frac{\alpha}{2}$ indicated that subjects allocated statistically significantly more work to themselves under the manual condition ($M = 7.274\%, SD = 27.927\%$) compared with the semi-autonomous condition ($M = -10.598\%, SD = 28.096\%$; $t(23) = 2.043, p = 0.013$ for comparison). Results from a one-sample t-test showed a statistical trend for subjects to allocate less work to themselves under the semi-autonomous condition ($M = -10.598\%, SD = 28.096\%$) than the autonomous condition ($t(23) = -1.859, p = 0.038$).

We also compared the amount of work allocated to the subjects when working with a human versus a robot co-leader (Figure 12). A two-factor analysis of variance with replication indicated a statistically significant difference between teams with a human co-leader and teams with a robot co-leader ($F(1, 69) = 6.005$, $p < 0.016$). Results from a t-test showed that subjects allocated statistically significantly less work to them-

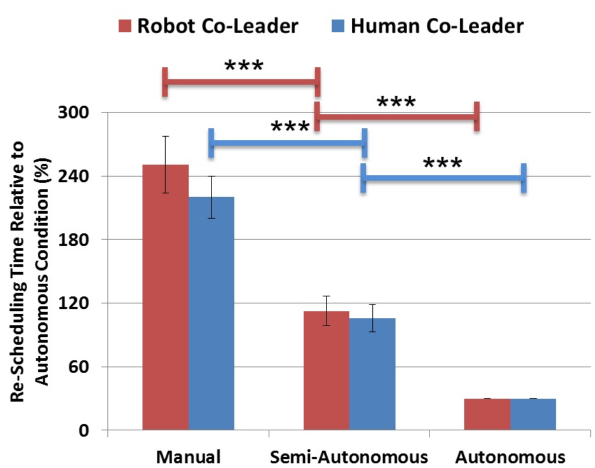


Fig. 10: The average and standard error of the rescheduling time across different conditions for decision-making authority and type of co-leader.

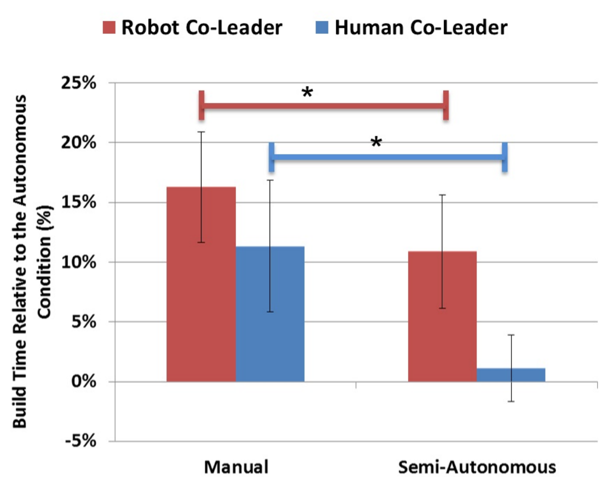


Fig. 11: The average and standard error of the assembly time across different conditions for decision-making authority and type of co-leader.

selves when working with a human co-leader ($M = -10.598\%, SD = 28.096\%$) than when working with a robot co-leader ($M = 12.186\%, SD = 26.065\%$) under the semi-autonomous condition ($t(23) = -3.077, p = 0.003$).

Next, we tested hypothesis **H7** to determine whether subjects preferred the human co-leader to fetch parts for them (Figure 13). Pearson's χ^2 test showed that the number of trials in which the human co-leader fetched more for the human assistant than for the subject in the autonomous condition (12 of 24 trials) was not statistically significantly different than in the manual condition (12 of 24 trials), $\chi^2 = 0.0, p = 1.0$.

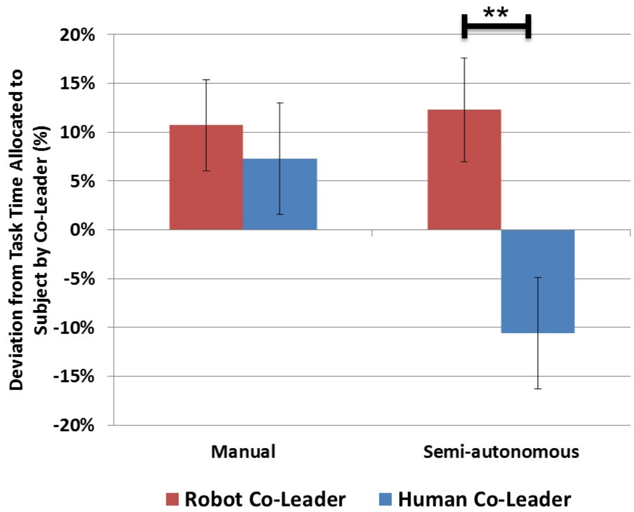


Fig. 12: The average and standard error of the percent difference between the time the subjects were assigned to work under the semi-autonomous and manual conditions relative to the autonomous condition.

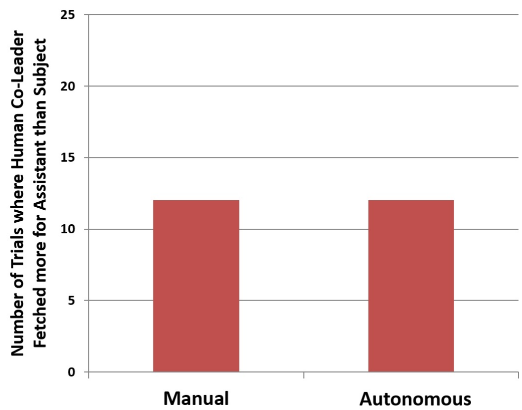


Fig. 13: The number of trials in which the human co-leader fetched more parts kits for the human assistant than for the subject.

5.2.2 Subjective Measure of Satisfaction When Working with Human and Robot Co-leaders

Based upon our findings when subjects worked with a robot co-leader, we revised our second hypothesis **H2** to state that workers prefer working with a co-leader with more decision-making authority. We now evaluate our revised hypothesis for teams with a human co-leader. An omnibus Friedman test for the post-trial (Table 1) and post-test (Table 2) questionnaires showed statistically significant differences only in the responses to Questions 14 ($p = 0.043$) and 18 ($p = 0.016$) from Table 1. The autonomous condition received the most positive Likert-scale responses, while the semi-autonomous condition garnered the most negative Likert-scale responses. These questions measured subjects' perceptions of how quickly the team completed the tasks and how well the team worked together. It is important to note that the time to completion for teams with human co-leaders was significantly faster under the semi-autonomous condition than the manual condition.

Next, we compare the effects of a human versus a robot co-leader on the subjective measures of participant satisfaction. Recall that hypothesis **H4** states that subjects would be more satisfied with a human co-leader. We used a Mann-Whitney U test to compare the Likert-scale responses from post-trial questionnaires following experiments with human co-leaders to those with a robot co-leader. Subjects agreed more strongly that the human co-leader liked, appreciated, and better understood the subject (Question 6, $z = 2.455$, $p = 0.018$; Question 8, $z = 3.256$, $p = 0.002$; and Question 5, $z = 2.8468$, $p = 0.007$); that the subject and human co-leader understood, trusted, and respected each other (Question 5, $z = 2.8468$, $p = 0.007$; Question 9, $z = 2.214$, $p = 0.032$; and Question 7, $z = 3.926$, $p < 0.001$); and that both the subject and human co-leader were necessary for the completion of the tasks (Question 21, $z = 2.451$, $p = 0.018$; and Question 19, $z = 5.674$, $p < 0.001$).

Hypothesis **H5** states that providing robots authority over team coordination more strongly improves the perceived value of these agents than giving similar authority to another human teammate. We evaluated this hypothesis using $x_{R,max}(i)$ and $x_{R,min}(i)$, defined as the maximum and minimum Likert-scale ratings associated with a given question for subject i working with a robot co-leader across all three experiment conditions. Similarly, $x_{H,max}(j)$ and $x_{H,min}(j)$ are the corresponding maximum and minimum Likert-scale ratings for subject j working with a human co-leader across all three experiment conditions. We evaluated the pairwise frequency with which subjects i working with the

Table 6: Differences between Likert-scale responses for the autonomous and manual conditions when working with a robot versus a human co-leader.

Question	Difference	Robot Co-Leader versus Human Co-Leader
1	$\Delta_R > \Delta_H$	$\chi^2 = 112.232, \boldsymbol{p < 0.001}$
2	$\Delta_R > \Delta_H$	$\chi^2 = 4.672, \boldsymbol{p < 0.031}$
5	$\Delta_R > \Delta_H$	$\chi^2 = 7.291, \boldsymbol{p = 0.007}$
6	$\Delta_R > \Delta_H$	$\chi^2 = 14.070, \boldsymbol{p < 0.001}$
8	$\Delta_R > \Delta_H$	$\chi^2 = 5.036, \boldsymbol{p = 0.025}$
9	$\Delta_R > \Delta_H$	$\chi^2 = 17.831, \boldsymbol{p < 0.001}$
10	$\Delta_R > \Delta_H$	$\chi^2 = 39.287, \boldsymbol{p < 0.001}$
11	$\Delta_R > \Delta_H$	$\chi^2 = 5.000, \boldsymbol{p = 0.250}$
12	$\Delta_R > \Delta_H$	$\chi^2 = 7.170, \boldsymbol{p = 0.007}$
13	$\Delta_R > \Delta_H$	$\chi^2 = 15.515, \boldsymbol{p < 0.001}$
14	$\Delta_R > \Delta_H$	$\chi^2 = 51.564, \boldsymbol{p < 0.001}$
15	$\Delta_R > \Delta_H$	$\chi^2 = 104.836, \boldsymbol{p < 0.001}$
16	$\Delta_R > \Delta_H$	$\chi^2 = 100.000, \boldsymbol{p < 0.001}$
17	$\Delta_R > \Delta_H$	$\chi^2 = 83.571, \boldsymbol{p < 0.001}$
18	$\Delta_R > \Delta_H$	$\chi^2 = 84.366, \boldsymbol{p < 0.001}$
19	$\Delta_R > \Delta_H$	$\chi^2 = 105.780, \boldsymbol{p < 0.001}$
22-24	$\Delta_R > \Delta_H$	$\chi^2 = 68.702, \boldsymbol{p < 0.001}$
4	$\Delta_H > \Delta_R$	$\chi^2 = 24.923, \boldsymbol{p < 0.001}$
20	$\Delta_H > \Delta_R$	$\chi^2 = 68.702, \boldsymbol{p < 0.001}$
21	$\Delta_H > \Delta_R$	$\chi^2 = 5.838, \boldsymbol{p = 0.016}$

robot co-leaders exhibited a larger difference in Likert-scale responses across experiment conditions than subjects j working with a human co-leader. We indicated that the difference in Likert-scale responses for subject i working with a robot co-leader were greater than for subject j with a human co-leader if $x_{R,max}(i) \geq x_{H,max}(j)$ and $x_{H,min}(j) > x_{R,min}(i)$, or else if $x_{R,max}(i) > x_{H,max}(j)$ and $x_{H,min}(j) \geq x_{R,min}(i)$. Similarly, the difference in Likert-scale responses for subject j working with a human co-leader were indicated to be greater than for subject i with a robot co-leader if $x_{H,max}(j) \geq x_{R,max}(i)$ and $x_{R,min}(i) > x_{H,min}(j)$, or else if $x_{H,max}(j) > x_{R,max}(i)$ and $x_{R,min}(i) \geq x_{H,min}(j)$.

The Pearson χ^2 test indicated that level of co-leader autonomy more strongly influenced the perceived value of the robot teammate than the human teammate across 16 of the 21 post-trial and post-experiment Likert-scale questions. The results are summarized in Table 6. Performing a post-hoc analysis with Bonferroni correction of $\frac{\alpha}{2}$ indicated that level of autonomy for the co-leader more strongly affected the Likert-scale response for subjects working with the robot co-leader than the human co-leader.

5.2.3 Subjective Measures of Subjects' Task Allocation Strategies, Preferences and Views

We administered a second post-test questionnaire (shown in Table 3) to subjects who worked with a human co-leader, in order to understand how subjects performed task allocation, their views on different team roles and their preferences for performing different types of tasks.

To determine whether subjects felt that they employed a different strategy when allocating tasks under the semi-autonomous and manual conditions (Question 27), we conducted a Pearson's χ^2 test, which indicated a statistically significant difference between the two ($\chi^2 = 7.347, p = 0.025$). Specifically, 75% of subjects agreed that they used a different strategy, while 21% of subjects disagreed. The Pearson's χ^2 test indicated no significant difference between groups in responses to Question 28.

We conducted a Friedman test to compare Likert-scale responses to Question 28 versus those given for Question 29, Question 31 versus Question 32 and Question 33 versus Question 34. Results from this test indicated no significant difference between these pairs of questions ($p = 0.647$, $p = 0.471$, and $p = 0.647$, respectively).

5.2.4 Analysis of Open-Ended Responses: Human Co-Leader

Questions 25 and 26 of the post-trial questionnaire offered subjects the opportunity to provide open-ended responses detailing which condition they preferred working under, and to whom they would give control of the task allocation in a manufacturing setting. While the majority of subjects' responses were supportive of a robotic assistant that autonomously allocated work, we also provide representative responses from subjects who preferred the manual, semi-autonomous and autonomous conditions, as shown in Table 7. While most of the subjects' responses directly supported one of the three experimental conditions, some participants suggested a "blended" level of control, where the robot was able to assign tasks, but the subject would be able to override the robot in the event that they became bored or overwhelmed.

6 Discussion

6.1 Guidance on Deploying Autonomous Robot Teammates

The aim of this study was to determine how much control a human member of a human-robot team should have over their robot counterpart in order to maximize team efficiency and worker satisfaction. We hypothesized that allowing workers some control over the task allocation process would increase their satisfaction without too great a sacrifice to team efficiency; however,

Table 7: Representative open-ended responses from subjects preferring the manual, semi-autonomous and autonomous conditions.

Manual
"I would keep the job [of scheduling] with myself, as it gives me greater flexibility to decide who does what" "The [manual condition] to be able to see the big picture."
Semi-autonomous
"Preferred the [semi-autonomous condition]. I still retained some agency in assigning tasks globally, but the ordering was optimized for me. "When I schedule the work for myself and the co-leader scheduled the work for the assistant. I knew which tasks I was good at, so I preferred this scenario."
Autonomous
"I prefer the co-leader scheduling because it is less work for me." "This scenario seemed to result in the least idle time amongst the three of us."

we observed that autonomous control yielded improvements to both objective and subjective measures compared with manual or semi-autonomous control. Specifically, hypothesis **H1**, that autonomy would increase productivity, was supported, yet hypothesis **H2**, that workers prefer partial control over task allocation decisions, was not supported. This finding is in keeping with anecdotal evidence that subjects prefer working with highly autonomous robots [2].

These results provide guidance for the successful introduction of robots into human teams. First, providing human teammates more decision-making authority over robot behavior is not sufficient to improve worker satisfaction, and may in fact degrade team performance. Also, team fluency does appear to positively correlate with a human worker's willingness to collaborate with robotic technology.

Second, results from these experiments have indicated potential efficiency-related issues following from enabling human workers to perform task allocation. For example, giving human workers the ability to task a teammate who can only fetch parts for the team can result in workers monopolizing the time of their fetching teammates (hypothesis **H7**). We also observed that people working with a robot co-leader tend to assign a disproportionate amount of work to themselves (hypothesis **H6**). This disparity is most significant when people only have control over which tasks they will perform. The finding that subjects allocated a disproportionate amount of work to themselves when working with the robot are in keeping with a well-known phenomenon called the *planning fallacy*, which states that people underestimate the amount of time they need to complete a set of tasks and overestimate the amount of time that others need to complete the same set of tasks [27]. Our findings also show that subjects perceived the human co-leader more favorably than the robotic co-leader, and subjects allocated less work to themselves when working with the human in the semi-autonomous control. This discrepancy suggests that the planning fallacy could be more profound when working with a robotic teammate.

Third, responses to post-experiment questionnaires indicated subjects felt that the human co-leader was a better teammate than the robotic teammate, which supports hypothesis **H4**. Specifically, the subjects felt more strongly that the human co-leader liked, appreciated and understood them; that subjects and the human co-leader understood, trusted, and respected each other; and that both the subjects and the human co-leader were important to the task.

Finally, although subjects attributed higher value to a human co-leader than a robotic one, supporting hypothesis **H4**, we did see evidence that greater robot authority over task allocation more strongly improved the perceived value of the agent than giving similar authority to another human teammate, supporting hypothesis **H5**. Interestingly, the subjects' rated desire to work with the robot again also grew as a function of robot autonomy. It may therefore be possible to bridge the gap in the perceived value of human and robotic teammates by further enhancing the robot's autonomy and authority in team decision-making.

Given these points, we recommend that roboticists provide robots with as much autonomy as possible to support human-robot team coordination. Robot teammates with the ability to autonomously allocate and schedule tasks can improve both task completion time and the desire of human workers to cooperate with their robotic teammates. Furthermore, people may not effectively understand how to utilize robotic teammates with specialized capability. For example, our study indicated subjects did not optimally employ a robot that was dedicated solely to fetching part kits for the primary build process. While human fetchers might be able to perceive and respond to a human builder monopolizing a fetcher's time, a robotic assistant would not unless given the ability to do so. Allowing robots more autonomy over their behavior may help to counteract these biases and guide people toward a better understanding of how to best utilize these robots.

6.2 Limitations and Future Work

There are limitations to our experimental findings. Our sample population consisted of college students and young professionals whose livelihoods were not threatened by the possibility of robots replacing them. Providing manufacturing workers with more control in the decision-making process may still influence the satisfaction of those workers. However, our findings suggest that team fluency is also likely to be an important component in the successful introduction of robotic teammates. To better understand the relative contributions of team fluency and decision-making authority to worker satisfaction in manufacturing, a future study in which manufacturing workers are specifically recruited will be necessary.

Also, each participant in our experiment worked as part of a human-robot team for a single, 90-minute period; however, those in the manufacturing field would be working with robots every day, possibly for years. Human workers may have strong preferences for some jobs over others, and may make different choices or have different preferences for task allocation when working with robots every day over the long term. We propose that a longitudinal study is necessary to observe the trajectory of human worker satisfaction over time, since the short- and long-term effects of decision-making authority may differ.

The social role of the robot is an interesting point for future study. In all control conditions, the robot teammate was responsible for the sequencing of tasks, tracking the status of the scheduling, and telling the human teammates when each subtask could be started and finished. In the autonomous condition, the robot had full control over the scheduling decisions, whereas, in the manual condition, the subject had control over the allocation of tasks to the team. In the semi-autonomous condition, the role of task allocation was shared between the subject and the robot co-leader. Thus, one might see the robot co-leader as more of a supervisor in the autonomous condition, more like a teammate when control was shared in the semi-autonomous condition, and more like a tool or resource in the manual condition. Furthermore, the robot was not better than the human teammates at fetching or building part kits. This phyiscal limitation may make the robot seem less of an equal member of the team. Providing the robot the ability to fetch and build as quickly as the human teammates may also change the perceived role of the robot on the team. These points about the perception of the robot's social role on the team would be an interesting area for further study.

7 Conclusion

With the increasing desire and ability to integrate autonomous robotic agents into manufacturing environments, it is important to understand how much decision-making authority human workers should have over their robotic counterparts when allocating tasks to both human and robot team members. While worker autonomy can improve team efficiency, providing a worker either too little or too much control may be alienating or overwhelming, respectively. We conducted a set of experiments with human subjects to determine how much control a worker should have over the task allocation process, and how working with a robotic teammate may change team dynamics, as opposed to working with only human teammates. We found that an autonomous robot can outperform a human worker when conducting part ($p < 0.001$) or all of the task allocation process ($p < 0.001$). However, rather than finding an ideal balance of control authority to maximize worker satisfaction, we observed that workers preferred to cede control authority to the robot ($p < 0.001$). Furthermore, we found that people were more likely to allocate a disproportionate amount of work to themselves while working with a robot ($p < 0.01$). Our results suggest that providing workers with a role in the allocation of tasks to their robotic counterparts may not be an effective method of improving worker satisfaction. Rather, team fluency may have a stronger influence on worker satisfaction than the individual's level of decision-making authority. Furthermore, people not experienced in working alongside robotic teammates may not know how best to utilize this technology. Allowing robots with advanced scheduling algorithms to control scheduling decisions may provide an ability to account for these biases and increase the degree to which human workers value their robotic teammates.

References

1. Julie A. Adams. Multiple robot-single human interaction: effects on perceived workload and performance. *Behavior and Information Technology*, 28(2):183–298, 2009.
2. Jennifer Alsever. Robot workers take over warehouses. *CNN Money*. `http://money.cnn.com/2011/11/09/smallbusiness/kiva_robots/`, November 9, 2011.
3. L. Ardissono, G. Petrone, G. Torta, and M. Segnan. M.: Mixed-initiative scheduling of tasks in user collaboration. In *In: Proc. of WEBIST 2012 - Eight International Conference on Web Information Systems and Technologies*, pages 342–351, 2012.
4. Michael J. Barnes, Jessie Y.C. Chen, Florian Jentsch, and Elizabeth S. Redden. Designing effective soldier-robot teams in complex environments: training, interfaces, and individual differences. In *Proceedinggs of the Interna-*

tional Conference on Engineering Psychology and Cognitive Ergonomics (EPCE), pages 484–493. Springer, 2011.
5. Mokhtar S. Bazaraa, John J. Jarvis, and Hanif D. Sherali. John Wiley & Sons, Inc., Hoboken, New Jersey, third edition, 2010.
6. Pauline Berry, Bart Peintner, Ken Conley, Melinda Gervasio, Tomás Uribe, and Neil Yorke-Smith. Deploying a personalized time management agent. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '06, pages 1564–1571, New York, NY, USA, 2006. ACM.
7. Dimitris Bertsimas and Robert Weismantel. *Optimization over Integers*. Dynamic Ideas, Belmont, 2005.
8. Elizabeth Blickensderfer, Janis A. Cannon-Bowers, and Eduardo Salas. Cross-training and team performance. making decisions under stress: Implications for individual and team training. *American Psychological Association*, pages 299–311, 1998.
9. Jennifer Casper and Robin Roberson Murphy. Human-robot interaction in rescue robotics. *IEEE Transaction on Systems, Man, and Cybernetics (SMCS)*, 34(2):138–153, 2004.
10. Jessie Y.C. Chen, Michael J. Barnes, and Zhihua Qu. Roboleader: an agent for supervisory control of mobile robots. In *Proceedings of the International Conference on Human-Robot Interaction (HRI)*, 2010.
11. Andrew S Clare, Mary (Missy) L. Cummings, Jonathan P. How, Andrew K. Whitten, and Olivier Toupet. Operator objective function guidance for a real-time unmanned vehicle scheduling algorithm. 9:161–173, 2012.
12. Mary (Missy) L. Cummings, Amy S. Brzezinski, and John D. Lee. Operator performance and intelligent aiding in unmanned aerial vehicle scheduling. *IEEE Intelligent Systems*, 22(2):52–59, March 2007.
13. Edmund H. Durfee, James C. Boerkoel Jr., and Jason Sleight. Using hybrid scheduling for the semi-autonomous formation of expert teams. *Future Generation Computer Systems*, July 2013.
14. Elliot E. Entin and Daniel Serfaty. Adaptive team coordination. *Human Factors*, 41:312–325, 1999.
15. Siyuan Feng, Eric Whitman, X. Xinjilefu, and Christopher G. Atkeson. Optimization-based full body control for the darpa robotics challenge. *Journal of Field Robotics*, 32(2):293–312, 2015.
16. Dieter Fox. Adapting the sample size in particle filters through kld-sampling. *International Journal of Robotics Research (IJRR)*, 22:985–1003, 2003.
17. Matthew C. Gombolay, Ronald J. Wilcox, and Julie A. Shah. Fast scheduling of multi-robot teams with temporospatial constrints. In *Proceedings of the Robots: Science and Systems (RSS)*, Berlin, Germany, June 24-28, 2013.
18. Michael A. Goodrich, Bryan S. Morse, Cameron Engh, Joseph L. Cooper, and Julie A. Adams. Towards using UAVs in wilderness search and rescue: Lessons from field trials. *Interaction Studies, Special Issue on Robots in the Wild: Exploring Human-Robot Interaction in Naturalistic Environments*, 10(3):453–478, 2009.
19. Masahiro Hamasaki, Hideaki Takeda, Ikki Ohmukai, and Ryutaro Ichise. Scheduling support system for academic conferences based on interpersonal networks. In *Proceedings of ACM Hypertext*, 2004.
20. Thomas Haynes, Sandip Sen, Neeraj Arora, and Rajani Nadella. An automated meeting scheduling system that utilizes user preferences. In *Proceedings of the First International Conference on Autonomous Agents*, AGENTS '97, pages 308–315, New York, NY, USA, 1997. ACM.
21. Paul Hebert, Max Bajracharya, Jeremy Ma, Nicolas Hudson, Alper Aydemir, Jason Reid, Charles Bergh, James Borders, Matthew Frost, Michael Hagman, John Leichty, Paul Backes, Brett Kennedy, Paul Karplus, Brian Satzinger, Katie Byl, Krishna Shankar, and Joel Burdick. Mobile manipulation and mobility as manipulationdesign and algorithms of robosimian. *Journal of Field Robotics*, 32(2):255–274, 2015.
22. Guy Hoffman. Evaluating fluency in human-robot collaboration. In *International Conference on Human-Robot Interaction (HRI), Workshop on Human Robot Collaboration*, 2013.
23. Guy Hoffman and Cynthia Breazeal. Effects of anticipatory action on human-robot teamwork: Efficiency, fluency, and perception of team. In *Proceedings of the International Conference on Human-Robot Interaction (HRI)*, pages 1–8, 2007.
24. Eli R. Hooten, Sean T.Hayes, and Julie A. Adams. A comparison of communicative modes for map-based tasking. In *IEEE Internation Conference on Systems, Man, and Cybernetics*, 2011.
25. Matthew Johnson, Brandon Shrewsbury, Sylvain Bertrand, Tingfan Wu, Daniel Duran, Marshall Floyd, Peter Abeles, Douglas Stephen, Nathan Mertins, Alex Lesman, John Carff, William Rifenburgh, Pushyami Kaveti, Wessel Straatman, Jesper Smith, Maarten Griffioen, Brooke Layton, Tomas de Boer, Twan Koolen, Peter Neuhaus, and Jerry Pratt. Team ihmc's lessons learned from the darpa robotics challenge trials. *Journal of Field Robotics*, 32(2):192–208, 2015.
26. Henry L. Jones, Stephen M. Rock, Dennis Burns, and Steve Morris. Autonomous robots in SWAT applications: Research, design, and operations challenges. *AUVSI*, 2002.
27. Daniel Kahneman and Amos Tversky. Intuitive prediction: biases and corrective procedures. Technical report.
28. Corey D. Kidd and Cynthia Breazeal. Effect of a robot on user perceptions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems IROS 2004*, volume 4, pages 3559–3564, Sept 2004.
29. Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in humanrobot interaction. *International Journal of Human-Computer Studies*, 64(10):962 – 973, 2006.
30. Santiago Macho, Marc Torrens, and Boi Faltings. A multi-agent recommender system for planning meetings. In *Proceedings of Workshop on Agent-based Recommender Systems, Autonomous Agents 2000, ACM*, 2000.
31. Colin F. Mackenzie, Yan Xiao, and Richard Horst. Video task analysis in high performance teams. *Cognition, Technology, and Work*, 6:139–147, 2004.
32. Robin R. Murphy. Meta-analysis of autonomy at the darpa robotics challenge trials. *Journal of Field Robotics*, 32(2):189–191, 2015.
33. Nicola Muscettola, Paul Morris, and Ioannis Tsamardinos. Reformulating temporal plans for efficient execution. In *Proceedings of the 6^{th} International Conference on Principles of Knowledge Representation and Reasoning (KR&R)*, Trento, Italy, June 2-5, 1998.
34. Stefanos Nikolaidis and Julie Shah. Human-robot cross-training: computational formulation, modeling and evaluation of a human team training strategy. In *Proceedings of the International Conference on Human-Robot Interaction (HRI)*, pages 33–40, 2013.

35. Raja Parasuraman and Dietrich H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52:381–410, 2010.
36. Raja Parasuraman, Mouloua Mustapha, and Brian Hilburn. Adaptive aiding and adpative task allocation enchance human-machine systems. *Automation technology and human performance: Current research and trends*, pages 119–123, 1999.
37. Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. Comparing a computer agent with a humanoid robot. In *2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 145–152, March 2007.
38. Jason C. Ryan, Ashis Gopal Banerjee, Mary L. Cummings, and Nicholas Roy. Comparing the performance of expert user heuristics and an integer linear program in aircraft carrier deck operations. *IEEE Transaction on Cybernetics*, PP(9), August 2013.
39. Eduardo Salas, Jennifer E. Fowlkes, Renee J. Stout, Dana M. Milanovich, and Carolyn Prince. Does CRM training improve teamwork skills in the cockpit?: Two evaluation studies. *Human Factors*, 41:326–343, 1999.
40. Julie Shah, James Wiken, Brian Williams, and Cynthia Breazeal. Improved human-robot team performance using chaski, a human-inspired plan execution system. In *Proceedings of the International Conference on Human-Robot Interaction (HRI)*, pages 29–36, 2011.
41. Anthony Stentz, Herman Herman, Alonzo Kelly, Eric Meyhofer, G. Clark Haynes, David Stager, Brian Zajac, J. Andrew Bagnell, Jordan Brindza, Christopher Dellin, Michael George, Jose Gonzalez-Mora, Sean Hyde, Morgan Jones, Michel Laverne, Maxim Likhachev, Levi Lister, Matt Powers, Oscar Ramos, Justin Ray, David Rice, Justin Scheifflee, Raumi Sidki, Siddhartha Srinivasa, Kyle Strabala, Jean-Philippe Tardif, Jean-Sebastien Valois, J. Michael Vande Weghe, Michael Wagner, and Carl Wellington. Chimp, the cmu highly intelligent mobile platform. *Journal of Field Robotics*, 32(2):209–228, 2015.
42. C. Volpe, J. Cannon-Bowers, E. Salas, and P. Spector. The impact of cross training on team functioning. *Human Factors*, 38:87–100, 1996.
43. Joshua Wainer, David J. Feil-Seifer, Dylan A. Shell, and Maja J. Matarić. Embodiment and human-robot interaction: A task-based perspective. In *The 16th IEEE International Symposium on Robot and Human interactive Communication*, pages 872–877, August 2007.
44. Ronald J. Wilcox, Stefanos Nikolaidis, and Julie A. Shah. Optimization of temporal dynamics for adaptive human-robot interaction in assembly manufacturing. In *Proceedings of Robotics: Science and Systems (RSS)*, Sydney, Australia, July 9-13 2012.
45. Haoqi Zhang, Edith Law, Rob Miller, Krzysztof Gajos, David Parkes, and Eric Horvitz. Human computation tasks with global constraints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 217–226, New York, NY, USA, 2012. ACM.

Matthew C. Gombolay is a PhD candidate and NSF Graduate Research Fellow in the Interactive Robotics Lab at MIT. He received his S.M. (2013) in aeronautics and astronautics at MIT and his B.S (2011) in mechanical engineering from Johns Hopkins University. Matthew studies the interaction of humans and automation and is developing computational methods for real-time and collaborative resource optimization.

Reymundo A. Gutierrez is a Masters of Engineering candidate in the Interactive Robotics Lab at MIT. He received an S.B. (2014) in computer science and engineering from MIT. He is developing methods for affordance-based interaction for manipulation and control of dexterous robots.

Shanelle G Clarke is currently pursuing her Bachelors of Science in aeronautics and astronautics at MIT. Her interests include adaptive control, modeling and simulation of autonomous aerial systems.

Giancarlo F. Sturla is pursuing his Bachelors of Science in computer science at the Massachusetts Institute of Technology. His interests include scheduling, data analysis and dynamic control of human-robot teams.

Julie A. Shah is an Assistant Professor in the Department of Aeronautics and Astronautics and leads the Interactive Robotics Lab at MIT. Shah received her SB (2004) and SM (2006) from the Department of Aeronautics and Astronautics at MIT, and her PhD (2010) in Autonomous Systems from MIT. Her lab draws on expertise in artificial intelligence, human factors, and systems engineering to develop interactive robots that emulate the qualities of effective human team members to improve the efficiency of human-robot teamwork.

Matthew C. Gombolay

Giancarlo F. Sturla

Reymundo A. Gutierrez

Julie A. Shah

Shanelle G. Clarke