Article

The International Journal of Robotics Research
2017, Vol. 36(5–7) 597–617
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0278364916688255
journals.sagepub.com/home/ijr
SAGE

Computational design of mixed-initiative human–robot teaming that considers human factors: situational awareness, workload, and workflow preferences

Matthew Gombolay, Anna Bair, Cindy Huang and Julie Shah

Abstract

Advancements in robotic technology are making it increasingly possible to integrate robots into the human workspace in order to improve productivity and decrease worker strain resulting from the performance of repetitive, arduous physical tasks. While new computational methods have significantly enhanced the ability of people and robots to work flexibly together, there has been little study of the ways in which human factors influence the design of these computational techniques. In particular, collaboration with robots presents unique challenges related to the preservation of human situational awareness and the optimization of workload allocation for human teammates while respecting their workflow preferences. We conducted a series of human subject experiments to investigate these human factors, and provide design guidelines for the development of intelligent collaborative robots based on our results.

Keywords

Human–robot teaming, human–robot interaction, planning and scheduling, situational awareness, workload, preference scheduling

1. Introduction

Human–robot teaming has the potential to increase the productivity of human labor and improve the ergonomics of manual tasks. Based on recent industry interest in fielding human–robot teams, researchers have been investigating how best to include a human participant in the decision-making loop as a member of a human–robot team (Adams, 2009; Ardissono et al., 2012; Barnes et al., 2011; Clare et al., 2012; Dragan and Srinivasa, 2012; Goodrich et al., 2009; Herlant et al., 2016; Hooten et al., 2011; Pierce and Kuchenbecker, 2012; Sanderson, 1989; Zhang et al., 2012). However, the intricate choreography required to coordinate human-robot teams safely and efficiently represents a challenging computational problem. Task allocation and sequencing with upper- and lower-bound temporal constraints is known to be NP-hard (Bertsimas and Weismantel, 2005).

Fully autonomous solutions to this problem have been recently proposed by both academic researchers (Bertsimas and Weismantel, 2005; Gombolay et al., 2013; Parker et al., 2015) and industry practitioners (Alsever, 2011). While these new computational methods have significantly enhanced the ability of people and robots to work flexibly together, there has been little study of the ways in which human factors must influence the design of these computational techniques. Specifically, we must consider how situational awareness changes as a function of the level of robot initiative during the decision-making process, the consequences of varying the workload assigned by the robot to human agents, and how to include the workflow preferences of human team members in decision-making. Improvements assessed through simple measures of efficiency, such as task time, do not guarantee the long-term productivity and viability of the human–robot team.

Researchers have shown that providing a machine or robotic agent with autonomous capabilities yields important benefits for human-robot team fluency (Dragan and Srinivasa, 2012, 2013; Dragan et al., 2013; Hooten et al., 2011; Gombolay et al., 2015; Pierce and Kuchenbecker, 2012; Tellex et al., 2014). For example, Hooten et al. (2011). For example, Hooten et al. (2011) have shown how autonomous mode control for input devices with a low degree of freedom can improve control of robotic agents with a high degree of freedom, such as manipulator arms. Tellex et al. (2014) imbued robots with the ability to generate queries for a human team member when necessary to resolve planning conflicts, and validated the benefits of

Massachusetts Institute of Technology, USA

Corresponding author:
Matthew Gombolay, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA.
Email: gombolay@mit.edu

such a capability through human subject experiments. However, these works typically either relegate the human role to that of a supervisor rather than a team member who must cooperate with the robot to plan and execute a schedule (as in works by Dragan and Srinivasa, 2012, 2013; Dragan et al., 2013; Hooten et al., 2011; Pierce and Kuchenbecker, 2012; Tellex et al., 2014), or focus on a human-robot dyad (Lasota and Shah, 2015; Nikolaidis and Shah, 2013), which does not require consideration of challenging scheduling constraints, owing to the relatively small size of the team.

Recent work, including studies we have conducted, has begun to explore the problem of human-robot team coordination (Bauer et al., 2008; Gombolay et al., 2014, 2015; Talamadupula et al., 2010, 2014). Like research in teleoperation and supervisory control, these studies have indicated that allowing the robotic agent a greater degree of autonomy yields immediate performance gains in mixedinitiative scheduling (Gombolay et al., 2014, 2015). In prior work, we showed that human participants preferred to cede control over scheduling decisions to a robotic teammate in such a setting (Gombolay et al., 2014). However, our preliminary work did not investigate key humanfactor concerns, such as how increased robot autonomy might reduce human situational awareness, the size of the workload that should be assigned to human teammates, or how to incorporate the scheduling preferences of human teammates in the event that they lead to suboptimal performance.

Human factors, such as situational awareness, workload influence, and workflow preferences, have long been studied, and a failure to consider these parameters of the human–machine interface problem has been shown to have serious consequences (Endsley, 1988, 1995; Endsley and Kaber, 1999; Kaber and Endsley, 2004; Riley et al., 2004). Kaber and Endsley (1997) succinctly enumerate the following potential issues following from increased automation: (1) a human operator failing to identify a problem and appropriately intervene, (2) over-trust in the automation (complacency), (3) a loss of situational awareness, and (4) degradation in the direct or manual control of the system.

Situational awareness has been defined as the ability to perceive, comprehend, and project the state of an environment (Endsley, 1995). Loss of situational awareness while operating highly autonomous systems has accounted for hundreds of deaths in commercial and general aviation (e.g. National Transportation Safety Board, 1973, 1979, 1981, 1988, 1990). Human workers must maintain their situational awareness in order to effectively take control of a job typically performed by an autonomous machine in the event that that machine fails.

Workload assignment is another key issue in human factors (Parasuraman et al., 2008; Stanton and Young, 2011; Tsang and Vidulich, 2006; Wickens, 2008). It has been shown in prior work that human performance is highly dependent on workload (Parasuraman et al., 2008; Proctor and Zandt, 2008; Stanton and Young, 2011; Tsang and Vidulich, 2006; Wickens, 2008): A workload that is too heavy or too light can degrade performance and contribute to a loss of situational awareness (Proctor and Zandt, 2008; Tsang and Vidulich, 2006).

Understanding and incorporating workflow preferences is also essential for safe, effective human–machine teaming (Alami et al., 2006; Hoffman and Breazeal, 2007; Kwon and Suh, 2012; Lasota and Shah, 2015; Nikolaidis and Shah, 2013). In manufacturing, human teams can develop individualized workflow preferences that are not shared by other teams in the same environment; consequently, a member of one team may be unable to replace a worker on another team effectively without a period of adjustment.

In this paper, we report the results from a series of human subject experiments in the context of human–robot team coordination. First, we investigated how situational awareness varies as a function of the degree of autonomy a robotic agent has during scheduling, and found that human participants' awareness of their team's actions decreased as the degree of robot autonomy increased. Given prior work indicating that humans typically prefer the robot to have greater autonomy (Baraglia et al., 2016; Gombolay et al., 2015; Hoffman and Breazeal, 2007; Huang and Mutlu, 2016; Liu et al., 2016), roboticists must balance the desire for increased automation and the performance improvements it yields with the risk for—and cost resulting from—reduced situational awareness.

Second, we studied how team fluency varies as a function of the workload (tasks not related to decision-making about scheduling) given to a human team member by a robotic agent, and the manner in which a robot should include the workflow preferences of its human teammates in the decision-making process.

A roboticist or practitioner of multi-agent coordination might take the most straightforward approach by including the preferences of each human team member and balancing work assignments according to a given fairness metric. However, we found that when the goal of including human team members "preferences is orthogonal to the goal of assigning each agent tasks in the way that most benefits the team's overall performance, people are usually amenable to relinquishing their preferred assignments for the sake of improved team fluency. We also observed a relationship between human team members" preferences, their utilization during task execution and their perception of team efficiency. Participants felt more strongly that their teams performed the assigned tasks using the least possible amount of time, even though the schedule duration (makespan) was constant across all trials within participants.

2. Background

As the complexity of human-operated machines has increased, so has the need for increased machine autonomy to aid human operators. As such, researchers in the fields of human factors and artificial intelligence, including

robotics, have sought to improve the fluency of the human–machine system. Here, we review related works and identify key gaps in the literature that demonstrate the need for the experimental investigation we present in this paper.

2.1. Aiding human workers via autonomy

There has been a flourish of recent work focused on the development of an improved human-machine interface (Barnes et al., 2011; Cummings et al., 2007; Dragan and Srinivasa, 2012; Dragan et al., 2013; Goodrich et al., 2009; Hooten et al., 2011; Jones et al., 2002). For example, Dragan et al. (2013) developed and explored an intelligent, customizable interface for teleoperation. This interface mediates the consequences of a human operator not being in close physical proximity to the action performed in order to make teleoperation more seamless, and leverages the autonomous capabilities of the robot to assist in accomplishing a given task. In such work, researchers often view the human operator as a vital component of the decision-making loop, particularly when this operator has knowledge of factors that are difficult to capture within a manually encoded, autonomous framework (Clare et al., 2012; Cummings et al., 2007; Durfee et al., 2014). Complementary to approaches that include the human operator in the loop, other work has focused on the development of computational methods able to generate scheduling solutions using information collected a priori from human experts (Ardissono et al., 2012; Hamasaki et al., 2004; Haynes et al., 1997; Macho et al., 2000; Zhang et al., 2012).

Researchers have proposed various mechanisms for distributed decision-making in the form of agents that can independently reason about their activities (Brunet et al., 2008; Gombolay et al., 2013; Jones et al., 2011; Korsah et al., 2013; Nunes and Gini, 2015; Parker et al., 2015; Tellex et al., 2014). For example, Tellex et al. (2014) developed a robotic system able to learn a mental model of how people perform various assembly manufacturing tasks and adapt workflow to improve fluency for a human–robot dyad.

While some researchers have focused on human-in-the-loop decision makers, others have investigated the complementary areas of teleoperation and blended autonomy, in which human and machine agents work jointly toward accomplishing a physical action (Dragan and Srinivasa, 2013; Herlant et al., 2016; Muelling et al., 2015; Pierce and Kuchenbecker, 2012) as opposed to cognitive tasks. For example, Pierce and Kuchenbecker (2012) developed a data-driven method for learning a mapping of the arm motions necessary to reach a specific physical state ("target pose") from a human' mental model and translating those motions to corresponding robot motions in a physical environment. The robot is then able to use this learned mapping to aid the operator in achieving the desired robot pose (Pierce and Kuchenbecker, 2012).

Herlant et al. (2016) investigated the challenges of controlling a robotic arm using a low-dimensional input device, such as a joystick. They showed that mode switching accounts for a significant time burden for the user, and developed an automatic mode selection algorithm that reduces this burden.

Muelling et al. (2015) developed an improved brain–computer interface to alleviate the challenges of latency, low-dimensional user commands, and asymmetric control inputs, all of which are common to robotic teleoperation. Their system relies on a combination of computer vision, user intent inference, and arbitration between human and robotic systems. In their work, Muelling et al. (2015) validated their system via experiments where participants used input from two intracortical implants to control a robotic manipulator with seven degrees of freedom. Muelling et al. (2015) found that their brain–computer interface enabled completion of tasks that were previously infeasible without arbitration.

There is also evidence that the manner in which people receive and interact with machine autonomy is influenced by a number of additional factors, including individual differences among operators and system embodiment (Hoffman and Ju, 2014; Ju and Sirkin, 2010; Klemmer et al., 2006; Lee et al., 2006a,b; Riek and Robinson, 2011; Takayama and Pantofaru, 2009). For example, Takayama and Pantofaru (2009) investigated proxemics in human–robot interaction and found differences based on patrticipants' sex and prior experiences in interacting with robots and animals. Ju and Sirkin (2010) studied the effect of embodiment in capturing attention and engendering a desire to interact with the system; Lee et al. (2006a) found embodiment with restrictions on tactile interaction to result in a null or negative effect. However, there has been little study of the ways in which human factors, including situational awareness, workload assignment, and workflow preferences must influence the design of computational techniques for mixed-initiative human–robot teaming.

2.2. Situational awareness

Within the field of human factors (Endsley, 1988, 1995; Endsley and Kaber, 1999; Kaber and Endsley, 2004; Riley et al., 2004) and, more recently, in human-robot interaction (Chen et al., 2007; Drury et al., 2006; Fong and Thorpe, 2001; Steinfeld et al., 2006) the study of situational awareness has been of utmost importance, particularly in the context of aviation (National Transportation Safety Board, 1973, 1979, 1981, 1988, 1990). In her seminal paper, Endsley (1995) defined a three-level model for situational awareness: perception (level 1 SA), comprehension (level 2 SA), and projection (level 3 SA). These levels require the operator of a complex system to perceive the state of the environment, understand the meaning of this state and project the state into the future in order to understand how that state must change (Endsley, 1995).

In subsequent work, Endsley and Kaber (1999) explored varying levels of automation in order to test situational

awareness. They found that higher automation resulted in improved performance if the implementation of that automation did not fail; however, if implementation did fail, automation resulted in much poorer performance by the human operator. Also, they wrote, collaboration on a task (as opposed to a human or robotic agent performing a task alone) can result in poorer performance and less situational awareness (Endsley and Kaber, 1999).

Kaber and Endsley (2004) attempted to address two design variables affecting situational awareness that had previously not been studied in conjunction: the level of automation and adaptive automation. In adaptive automation, the allocation of tasks to a human worker or a machine changes as a function of the state of the environment (Kaber and Endsley, 2004). Kaber and Endsley found that participants had higher situational awareness at lower levels of automation, and lower situational awareness at higher levels of automation. When adaptive automation changed such that participants experienced different automation levels at varied time spans, participants did not perceive the periods of higher automation as involving a smaller task load, as they were also monitoring the automated task execution (Kaber and Endsley, 2004).

While many researchers have focused on modeling situational awareness, understanding how situational awareness decreases, and evaluating the consequences of degraded situational awareness, few have developed interfaces specifically to augment situational awareness (Fong and Thorpe, 2001).

One major gap in the robotics and human factors literature is the study of situational awareness wherein humans plan and execute a sequence of actions collaboratively within a human–robot team. Much work has focused on the human in a supervisory control role (e.g. Endsley and Kaber, 1999; Tellex et al., 2014) or as part of a dyad, for which the coordination of actions is relatively simple (Nikolaidis and Shah, 2013).

2.3. Mental and physical workload

Workload is a key issue identified in the human subject literature, which has indicated that human performance is highly dependent on workload (Parasuraman et al., 2008; Proctor and Zandt, 2008; Stanton and Young, 2011; Tsang and Vidulich, 2006; Wickens, 2008). A combination of results from prior work has led to a model of the relationship between workload and performance: workload that is too heavy or too light can degrade both performance and situational awareness (Proctor and Zandt, 2008; Tsang and Vidulich, 2006). A consequence of a high workload is increased reliance on and compliance with automation (Parasuraman and Riley, 1997).

Researchers have previously sought effective means to reduce workload through the use of semi-autonomous decision support tools (Shah et al., 2015), particularly in the field of air traffic control, owing to the notoriously challenging nature of aircraft coordination (Loft et al., 2007; Lokhande and Reynolds, 2012; Niederée et al., 2012). Lokhande and Reynolds (2012) report that, even with the aid of a decision support tool, air traffic controllers spent 81.9% of their time in a head-down position looking at information displays, rather than visually monitoring traffic on the ground.

Loft et al. (2007) developed a model to predict the level of workload for air traffic controllers, and confirmed results from prior work indicating that mental workload increases with task difficulty. However, they also observed an unexpectedly stronger effect on mental workload as a function of the ability of air traffic controllers to prioritize tasks and manage resources.

To help evaluate mental workload, researchers have proposed various subjective and psycho-physiological metrics (Brookings et al., 1996; Hart and Staveland, 1988; Kramer, 1991; Steinfeld et al., 2006). The most well-known metric is the NASA Task Load Index (TLX): a subjective, multivariate means of evaluating perceived mental workload (Hart and Staveland, 1988).

While the relationship between workload and task performance has been studied extensively with regard to human factors, it remains uncharacterized in the context of human–robot teams in which a robotic agent plays a substantial role in coordinating physical work. Prior studies have shown that people prefer to delegate decision-making about scheduling to a robotic agent (Gombolay et al., 2015), yet there is a gap in the literature regarding the effects of varying physical workload on team fluency in such a scenario.

2.4. Scheduling preferences

Researchers in the fields of artificial intelligence and robotics have explored computational methods for incorporating preference-based constraints when coordinating human–robot teams (Alami et al., 2006; Berry et al., 2006; Hawkins et al., 2014; Kwon and Suh, 2012; Nikolaidis and Shah, 2013; Wilcox et al., 2012). Wilcox et al. (2012) developed an adaptive preferences algorithm to schedule human–robot teams dynamically in real time, according to the unique preferences of human workers, as human teams in a factory setting can vary greatly with regard to how they accomplish assembly tasks (Wilcox et al., 2012). Alami et al. (2006) encoded task-specific constraints and workflow preferences that allow for prediction of likely human actions. Berry et al. (2006) developed an artificial intelligence assistant, known as PTIME, to learn the preferences and schedule the activities of human operators via a mathematical programming technique. Bayesian networks (Kwon and Suh, 2012), first-order Markov logic networks, and AND–OR graphs (Hawkins et al., 2014) have also been used to predict human actions during human–robot collaboration.

Preferences for task scheduling have been the subject of much prior study (Grano et al., 2009; Haynes et al., 1997; Lottaz, 1996; Soomer and Franx, 2008), but the human factors of scheduling activities have not been as well assessed. Generally, research has focused on the implementation of fairness metrics (such as in the work of Zhang and Shah, 2015) and other mathematical formulations for optimally scheduling according to human team members' preferences (Grano et al., 2009; Haynes et al., 1997; Lottaz, 1996; Soomer and Franx, 2008). However, roboticists must also ask the fundamental question of whether these preferences should be included in robot decision-making, and, if so, how best to do so.

2.5. *Motivating the need for further investigation*

Although substantial bodies of work have made important contributions to the advancement of human–robot interaction, we have identified three key gaps in prior literature: First, human situational awareness as a function of robot initiative over decision-making in human–robot teaming has not yet been investigated, but the potential for degradation to situational awareness with increased robotic autonomy must be assessed. Second, this effect must also be studied in the context of collaboration during the performance of physical tasks. Finally, while mechanisms for preference scheduling have been developed, the human implications of an intelligent collaborative robot including (or not including) human workflow preferences into the scheduling process has not been addressed.

It is essential that we address these gaps in the literature. Human–robot teaming is in the process of transitioning from modes in which humans supervise automated systems to peer-to-peer and more collaborative modes of automation (Fong et al., 2003; Hoffman and Breazeal, 2004; Matthias et al., 2011; Reed and Peshkin, 2008; Unhelkar et al., 2014). We can observe this trend in the growing number of new applications within robotics that require the colocation of human and robotic work, such as Boeing's Fully Automated Upright Build (Boeing, 2014), BMW's collaborative manufacturing robots (Knight, 2013), and Amazon Robotics (formerly known as KIVA Systems) warehouse robots (Alsever, 2011).

To our knowledge, there have been no studies investigating changes to situational awareness in a mixed-initiative scheduling setting, varying mental workload during robotic scheduling of human teammates, or the way in which robotic agents should incorporate the workflow preferences of human team members. We have also not encountered any prior investigations that assess situational awareness in a mixed-initiative scheduling setting in which human and robotic agents are members of the same physical team.

Our work has significant new implications for the design of intelligent collaborative robots. Briefly, robotic agents must balance human participants' task type preferences with the workloads assigned to those participants. Scheduling participants to perform more highly preferred tasks at the cost of increased idle time can degrade team fluency. Also, providing a robotic agent with increased autonomy over scheduling decisions—while preferable from the human teammates' points of view—can degrade situational awareness. This degradation could have negative consequences in domains where the robotic agent is not highly reliable, such as new manufacturing applications or field robotics.

3. Aims of the experiment

Prior literature (Chen et al., 2007; Dragan and Srinivasa, 2012; Fong and Thorpe, 2001; Gombolay et al., 2015; Herlant et al., 2016) has shown the potential advantages of providing a robotic teammate with greater autonomy, and recent work in the realm of mixed-initiative scheduling has extended these findings. Such studies have indicated that a robot generates a schedule more quickly and that a team is able to complete assigned tasks more efficiently when the schedule is generated by the robotic agent alone as opposed to when a human team member assists in the scheduling process (Gombolay et al., 2013). Furthermore, participants in prior experiments have readily stated that they would prefer working with a robotic teammate with a greater degree of autonomy (Gombolay et al., 2015).

However, this recent work provides an incomplete picture. For example, we do not yet know the ramifications of conceding autonomy to a robotic agent in environments where human team members might have to reallocate work manually to account for an environmental disturbance that the robot is unable to consider. Also, we do not understand whether or how the *way* in which a robot schedules a team (e.g. whether the robot happens to assign tasks to participants who prefer them) affects the participants' experiences. Finally, we do not know whether the amount of work assigned by the robot results in a suitable workload for human teammates.

We conducted a series of experiments to better understand the following: (1) whether situational awareness degrades when the robotic agent has a greater degree of control over scheduling decisions, (2) how a robotic agent should schedule tasks for a human–robot team, given the human team members' workflow preferences, and (3) whether there is a trade-off between the degree to which human team members' scheduling preferences are included in the scheduling process and the effective utilization of those workers.

3.1. *Experiment: situational awareness in mixed-initiative human–robot teaming*

Prior work in human factors has also indicated that there are significant consequences associated with ceding decision-making initiative to an autonomous agent (Endsley, 1995,

1988; Kaber and Endsley, 2004; Kaber et al., 2000; Riley et al., 2004): Chiefly, the human counterpart can experience a decline in situational awareness. This phenomenon has been observed in a variety of domains, including telerobotics (Kaber et al., 2000). We proposed an experiment to serve as the first such investigation in the setting of human–robot teaming using mixed-initiative scheduling, with the human and robot workers sharing scheduling responsibilities.

3.1.1. Independent variable. To determine the potential consequences of providing a robotic teammate with greater autonomy over scheduling decisions, we conducted a novel human subject experiment consisting of three team members: a robot, a human subject, and a human assistant (i.e. a confederate), who were required to complete a series of fetching and building tasks. In this experiment, the independent variable was the allocation of authority over scheduling decisions; this independent variable had three levels, or conditions.

- *Manual control.* The human subject decides who will perform each of the tasks.
- *Semi-autonomous control.* The human subject decides which tasks he or she will perform, and the robot assigns the remaining tasks to itself and the human assistant.
- *Autonomous control.* The robot decides who will perform each of the tasks.

3.1.2. Hypothesis. We established the following hypothesis.

Hypothesis 1. *Participants' situational awareness will be poorer when the robotic teammate has greater autonomy over scheduling decisions.*

3.1.3. Dependent variables. To test Hypothesis 1, we conducted an experiment using the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1988), which was designed to measure the situational awareness of a pilot in an aircraft cockpit. During an experiment in which a pilot operated a simulated aircraft, the experimenter blanked out the information displays and the pilot was required to recall vital information about the state of the aircraft.

This protocol has disadvantages. For example, halting the experiment to question the subject is highly intrusive and could lead to a decline in performance when the subject must resume flying the aircraft. Also, the responses are highly dependent on the subject's ability to remember information, which decays as a function of time—over the course of a long test, the subject may begin to forget important pieces of information about the system's state. In our experimental design, we applied the same test, in the same manner, to all participants; therefore, any such negative effect would be balanced across experimental conditions. Furthermore, we did not repeat the SAGAT test; it was only administered once during the experiment, which concluded after the test administration.

For our SAGAT test, we employed a set of objective and subjective measures, as shown in Table 1. The objective measures evaluated the accuracy of the participants' perceptions of the state of the human–robot team; the subjective measures were paired with the objective measures to evaluate the participants' confidence in their answers to the objective questions.

Participants responded to the objective questions (i.e. questions 1, 5, 9, 13, and 17) using the template-based response format shown in Table 2, and responded to subjective questions (i.e. questions 2–4, 6–8, 10–12, 14–16, and 18–20) according to a five-point Likert response format ("strongly disagree," "weakly disagree," "neutral," "weakly agree,"and "strongly agree"). We included the later questions to gain insight into the participants' subjective perception of their situational awareness.

Table 2 depicts each individual subtask that could be assigned to each team member. (We describe the nature of these subtasks in the subsequent description of the experiment design.) However, we note for clarity that the task set consisted of fetching and building tasks A, B, C1, C2, D, E, F, and G, where the fetch and build subtasks for C1 were required to be completed before the fetch and build subtasks for C2 could begin. The table does not include fetch operations for A and D because the experiment began with kits A and D already fetched. This condition increased the number of possible actions that the human agents could take at the start of the experiment.

To test our hypothesis via objective measures, we defined a metric, called the "SA score," that assesses how well participants are able to provide the desired information for each question in Table 1. We computed the SA score for each team member according to equation (1), and we computed the overall SA score for the whole team according to equation (2). In these equations, S^a_{response} is the set of tasks that the subject reported for agent a for a given question, and S^a_{correct} is the correct set of tasks for agent a for that same question. In this manner, we have sets S^a_{response} and S^a_{correct} for each agent and for each of the objective questions 1, 5, 9, 13, and 17

$$\text{SA score for agent } a := |S^a_{\text{response}} \backslash S^a_{\text{correct}}| + |S^a_{\text{correct}} \backslash S^a_{\text{response}}| \quad (1)$$

$$\text{SA score for team} := \sum_{a=1}^{n} (\text{SA score for agent } a) \quad (2)$$

In essence, equation (1) counts the number of mistakes, false positives (incorrect tasks identified in the response) and false negatives (correct tasks not identified in the response). The team's SA score is a sum of the individual SA scores. We assume that the subject's situational awareness of each team member is equally important. A perfect score is equal to zero, and the worst possible score is equal to the total number of fetch and build tasks (14 tasks).

Let us consider an example in which the correct answers are as follows: subject, $S^{\text{participant}}_{\text{correct}} = \{\text{Fetch B, Build C1}\}$;

Table 1. Post-trial questionnaire administered to participants for the experiment to measure situational awareness as a function of the level of autonomy over scheduling decisions given to the robotic teammate. Participants responded to questions 1, 5, 9, 13, and 17 using the response form shown in Table 2. Participants responded to questions 2–4, 6–8, 10–12, 14–16, and 18–20 using a Likert response format ("strongly disagree," "weakly disagree," "neutral," "weakly agree," and "strongly agree").

Current actions
1. What is each team member currently doing? *(Circle nothing if the team member is idle.)*
2. I am aware of what the robot co-leader is doing.
3. I am aware of what the human assistant is doing.
4. I am aware of what I am doing.
Preceding action
5. Which task did each team member last complete prior to the current task? *(Circle nothing if the team member has not yet completed a task.)*
6. I am aware of which task the robot co-leader just did.
7. I am aware of which task the human assistant just did.
8. I am aware of what I just did.
Past schedule
9. Please list tasks each team member has completed. List the tasks in the order in which they were completed by writing 1 for the first task, 2 for the second task, and so forth.
10. I am aware of which tasks the human or robot co-leader has completed.
11. I am aware of which tasks the human assistant has completed.
12. I am aware of which tasks I have completed.
Future schedule
13. Which tasks will each team member complete in the future? (Circle one task in each row to show which team member will complete which task in the future.)
14. I am aware of which tasks the human or robot co-leader will do in the future.
15. I am aware of which tasks the human assistant will do in the future.
16. I am aware of which tasks I will do in the future.
Dynamic rescheduling
17. Given the work that has already been completed, who do you think will complete the remaining tasks if the human or robot co-leader was no longer available?
18. I am aware of the team's schedule.
19. If I had to produce a new schedule for the team, I would know enough.
20. If I had to produce a new schedule for the team, I would do a good job.

Table 2. Response format for post-test questionnaire (Table 1).

Team leader (you)	Human assistant	Robot co-leader
Fetch B	Fetch B	Fetch B
Fetch C1	Fetch C1	Fetch C1
Fetch C2	Fetch C2	Fetch C2
Build A	Build A	Build A
Build B	Build B	Build B
Build C1	Build C1	Build C1
Build C2	Build C2	Build C2
Fetch E	Fetch E	Fetch E
Fetch F	Fetch F	Fetch F
Fetch G	Fetch G	Fetch G
Build D	Build D	Build D
Build E	Build E	Build E
Build F	Build F	Build F
Build G	Build G	Build G

human assistant, $S^{\text{asst.}}_{\text{correct}} = \{\text{Fetch C1, Build A}\}$; and robotic agent, $S^{\text{robot}}_{\text{correct}} = \{\emptyset\}$. Let us say the subject provided the following answer: subject, $S^{\text{subject}}_{\text{response}} = \{\text{Fetch B, Build A}\}$; human assistant, $S^{\text{asst.}}_{\text{correct}} = \{\text{Fetch C1, Build D, Build G}\}$; and robotic agent, $S^{\text{robot}}_{\text{correct}} = \{\text{Fetch E}\}$. The SA score would then be calculated as

$$\begin{aligned}
&\text{SA score for subject}\\
&\quad = |S^{\text{participant}}_{\text{response}} \backslash S^{\text{participant}}_{\text{correct}}| + |S^{\text{participant}}_{\text{correct}} \backslash S^{\text{participant}}_{\text{response}}|\\
&\quad = |\{\text{Build A}\}| + |\{\text{Build C1}\}|\\
&\quad = 2\\
&\text{SA score for assistant}\\
&\quad = |S^{\text{asst.}}_{\text{response}} \backslash S^{\text{asst.}}_{\text{correct}}| + |S^{\text{asst.}}_{\text{correct}} \backslash S^{\text{asst.}}_{\text{response}}|\\
&\quad = |\{\text{Build D, Build G}\}| + \{\text{Build A}\}|\\
&\quad = 3\\
&\text{SA score for robot}\\
&\quad = |S^{\text{robot}}_{\text{response}} \backslash S^{\text{robot}}_{\text{correct}}| + |S^{\text{robot}}_{\text{correct}} \backslash S^{\text{robot}}_{\text{response}}|\\
&\quad = |\{\text{Build A}\}| + |\{\emptyset\}|\\
&\quad = 1\\
&\text{SA score for team}\\
&\quad = \sum_{a=1}^{n} (\text{SA score for agent } a)\\
&\quad = 2 + 3 + 1\\
&\quad = 6
\end{aligned}$$

3.2. Experiment: workflow preferences

We sought to understand how the robot's inclusion of human team members' preferences for completing particular tasks affects the relationship between the human and robotic agents.

3.2.1. Independent variable. Our independent variable was the degree to which participants' preferences were

respected by the robotic teammate when scheduling. We established three experimental conditions for this variable using a within-participants experiment design.

- *Positive.* The robot generates a schedule incorporating the preferences of the subject.
- *Neutral.* The robot ignores the preferences of the subject.
- *Negative.* The robot schedules according to the opposite of the preferences stated by the subject.

3.2.2. Hypothesis. We established the following hypothesis.

Hypothesis 2. *Participants would prefer to work with a robotic teammate that incorporates their scheduling preferences than with one that is unaware of their preferences, and participants would prefer to work with a robotic teammate that is ignorant of their preferences than with one that actively schedules against their preferences.*

3.2.3. Dependent variables. To test our hypothesis, we conducted a within-participants experiment in which all participants experienced each of the three conditions once, and completed a post-trial questionnaire after experiencing each condition. This questionnaire consisted of 21 Likert statements, as shown in Table 3. Hoffman (2013) previously developed and validated the questions drawn from the "Robot Teammate Traits" and "Working Alliance for Human–Robot Teams" surveys. The later survey is a derivative of the "Working Alliance Inventory," originally developed and validated by Horvath and Greenberg (1989).

Participants also completed a questionnaire on completing the tasks under each condition, as shown in Table 4. This questionnaire gathered demographic information and included three additional Likert statements summarizing the experience of the participants, along with two open-ended questions.

We note that this questionnaire is not balanced, in that the number of positive prompts (e.g. "I believe the robot likes me.") outweighed the number of negative prompts (e.g. "I feel uncomfortable with the robot."). However, potential bias arising from an unbalanced survey is mitigated since the same questionnaire is administered in each condition.

Table 3. 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 toward 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 team's 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.

Table 4. Subjective measures—post-test questionnaire.

Overall preference
22. If the robot scheduled me to complete the tasks I preferred, I would want to work with the robot again.
23. If the robot did not know my preferences when scheduling, I would want to work with the robot again.
24. If the robot scheduled me to do different tasks from what I preferred, 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?

3.3. Experiment: workload vis-à-vis workflow preferences

In this experiment, we studied how team fluency varies as a function of the size of the workload assigned to a human teammate by a robotic teammate. We focused exclusively on modulating the degree to which scheduling preferences were included, and did not control for workload—rather, we controlled for overall team efficiency (makespan). We discuss how including participants' preferences in the scheduling process can decrease their workload—and, in turn, lead to decreased team fluency—in Section 7.

To isolate the effects of variation in a subject's workload, we separated the inclusion of scheduling preferences and increase of the subject's workload into two independent variables. We posit that decoupling workload from preferences results in a clearer understanding of the effects of

Table 5. Four experimental conditions varying two independent variables (degree to which scheduling preferences are included and participants' utilization), each of which have two levels (high, low).

2 × 2 **design**	High utilization	Low utilization
High preference	High preference–high utilization	High preference–low utilization
Low preference	Low preference–high utilization	Low preference–low utilization

varying workload and, in turn, the inclusion of workflow preferences.

3.3.1. Independent variables. We considered two independent variables: (1) the degree to which the robot respected participants' preferences during scheduling, and (2) the participants' utilization, defined as the total amount of time for which the subject was occupied during execution of a particular schedule. We identified a subject as having high utilization if the majority of the subject's time was spent working, as opposed to being idle, and vice versa for low utilization. We employed a 2 × 2 within-participants design with the following four conditions, as shown in Table 5.

- *High preference–high utilization.* The robot generates a schedule incorporating the preferences of the participant and highly utilizes the participant.
- *High preference–low utilization.* The robot generates a schedule incorporating the preferences of the participant and minimally utilizes the participant.
- *Low preference–high utilization.* The robot generates a schedule according to the opposite of the preferences of the participant and highly utilizes the participant.
- *Low preference–low utilization.* The robot generates a schedule according to the opposite of the preferences of the participant and minimally utilizes the participant.

3.3.2. Hypotheses. We established the following hypotheses.

Hypothesis 3A. *A participant's subjective assessment of a robotic teammate is favorably influenced by working with a robot that makes allocation decisions that incorporate the participant's scheduling preferences, as opposed to decisions that contradict the participant's preferences.*

(In contrast with Hypothesis 2, this hypothesis was assessed while controlling for the workload utilization of the participant.)

Hypothesis 3B. *A participant's subjective assessment of a robotic teammate is favorably influenced by working with a robot that makes work allocation decisions that result in high utilization of the participant's time, as opposed to low utilization.*

3.3.3. Dependent variables. To test our hypotheses, we conducted a within-participants experiment in which each participant experienced each condition once. As in the previous experiment, we administered a post-trial questionnaire after each of the conditions, as well as a post-test questionnaire after each participant had completed all conditions. The investigating agent workload included four conditions instead of three; as such, participants responded to a total of four post-trial questionnaires. We employed the same design for the post-trial (Table 3) and post-test questionnaires (Table 4).

4. Formal problem definition

The problem of scheduling a team of heterogeneous agents to complete a set of tasks with upper- and lower-bound temporal constraints and shared resources (e.g. spatial locations) falls within the XD [ST-SR-TA] class of scheduling problems, according to the comprehensive taxonomy defined by Korsah et al. (2013). This class is one of the most computationally challenging in the field of scheduling. The [ST-SR-TA] class of problems is composed of tasks requiring one robot or agent at a time (single-robot tasks [ST]), robots or agents that perform one task at a time (single-task robots [SR]) and a time-extended schedule of tasks that must be built for each robot or agent (time-extended allocation [TA]). This time-extended schedule includes cross-schedule dependencies (XD) among the individual schedules of the agents; such dependencies arise, for example, when agents must share limited-access resources (e.g. physical locations).

We formulated an instance of this problem to develop an experiment task as a mixed-integer linear program, as depicted in equations (3) to (13). This formulation serves as a common basis to model each of the three experiments. We subsequently discuss experiment-specific extensions

$$\min z, z = g\left(\{A^a_{\tau_i^j} | \tau_i^j \in \boldsymbol{\tau}, a \in A\}, \{J_{\langle \tau_i^j, \tau_x^y \rangle} | \tau_i^j, \tau_x^y \in \boldsymbol{\tau}\}, \{s_{\tau_i^j}, f_{\tau_i^j} | \tau_i^j \in \boldsymbol{\tau}\}\right) \tag{3}$$

subject to

$$\sum_{a \in A} A^a_{\tau_i^j} = 1, \forall \tau_i^j \in \boldsymbol{\tau} \tag{4}$$

$$ub_{\tau_i^j} \geq f_{\tau_i^j} - s_{\tau_i^j} \geq lb_{\tau_i^j}, \forall \tau_i^j \in \boldsymbol{\tau} \tag{5}$$

$$f_{\tau_i^j} - s_{\tau_i^j} \geq lb^a_{\tau_i^j} - M\left(1 - A^a_{\tau_i^j}\right), \forall \tau_i^j \in \boldsymbol{\tau}, a \in A \tag{6}$$

$$s_{\tau_x^y} - f_{\tau_i^j} \geq W_{\langle \tau_i^j, \tau_x^y \rangle}, \forall \tau_i^j, \tau_x^y \in \boldsymbol{\tau} |, \forall W_{\langle \tau_i^j, \tau_x^y \rangle} \in \boldsymbol{TC} \tag{7}$$

$$f_{\tau_x^y} - s_{\tau_i^j} \leq D^{\text{rel}}_{\langle \tau_i, \tau_j \rangle}, \forall \tau_i^j, \tau_x^y \in \boldsymbol{\tau} | \exists D^{\text{rel}}_{\langle \tau_i^j, \tau_x^y \rangle} \in \boldsymbol{TC} \tag{8}$$

$$f_{\tau_i^j} \leq D^{\text{abs}}_{\tau_i^j}, \forall \tau_i \in \boldsymbol{\tau} | \exists D^{\text{abs}}_{\tau_i^j} \in \boldsymbol{TC} \tag{9}$$

$$s_{\tau_x^y} - f_{\tau_i^j} \geq M\left(A^a_{\tau_i^j} + A^a_{\tau_x^y} - 2\right) + M\left(J_{\left\langle \tau_i^j, \tau_x^y \right\rangle} - 1\right), \forall \tau_i^j, \tau_x^y \in \boldsymbol{\tau}, \forall a \in A \tag{10}$$

$$s_{\tau_i^j} - f_{\tau_x^y} \geq M\left(A^a_{\tau_i^j} + A^a_{\tau_x^y} - 2\right) - M\left(J_{\left\langle \tau_i^j, \tau_x^y \right\rangle}\right), \forall \tau_i^j, \tau_x^y \in \boldsymbol{\tau}, \forall a \in A \tag{11}$$

$$s_{\tau_x^y} - f_{\tau_i^j} \geq M\left(J_{\left\langle \tau_i^j, \tau_x^y \right\rangle} - 1\right), \forall \tau_i^j, \tau_x^y \in \boldsymbol{\tau} | R_{\tau_i^j} = R_{\tau_x^y} \tag{12}$$

$$s_{\tau_i^j} - f_{\tau_x^y} \geq -M\left(J_{\left\langle \tau_i^j, \tau_x^y \right\rangle}\right) \forall \tau_i, \tau_j \in \boldsymbol{\tau} | R_{\tau_i^j} = R_{\tau_x^y} \tag{13}$$

In this formulation, $A^a_{\tau_i^j} \in \{0, 1\}$ is a binary decision variable for the assignment of agent a to subtask τ_i^j (i.e. the jth subtask of the ith task); $A^a_{\tau_i^j}$ equals 1 when agent a is assigned to subtask τ_i^j and 0 otherwise. $J_{\left\langle \tau_i^j, \tau_x^y \right\rangle} \in \{0, 1\}$ is a binary decision variable specifying whether τ_i^j comes before or after τ_x^y, and $s_{\tau_i^j}, f_{\tau_i^j} \in [0, \infty)$ are the start and finish times of τ_i^j, respectively. $\boldsymbol{TC}$ is the set of simple temporal constraints relating task events. M is a large, positive constant used to encode conditional statements as linear constraints.

Equation (3) is a general objective that is a function of the decision variables $\{A^a_{\tau_i^j} | \tau_i^j \in \boldsymbol{\tau}, a \in A\}$, $\{J_{\left\langle \tau_i^j, \tau_x^y \right\rangle} | \tau_i^j, \tau_x^y \in \boldsymbol{\tau}\}$ and $\{s_{\tau_i^j}, f_{\tau_i^j} | \tau_i^j \in \boldsymbol{\tau}\}$. Equation (4) ensures that each τ_i^j is assigned to a single agent. Equation (5) ensures that the duration of each $\tau_i^j \in \boldsymbol{\tau}$ does not exceed its upper- and lower-bound durations. Equation (6) requires that the duration of τ_i^j, $f_{\tau_i^j} - s_{\tau_i^j}$ is no less than the time required for agent a to complete τ_i^j. Equation (7) requires that τ_x^y occurs at least $W_{\left\langle \tau_i^j, \tau_x^y \right\rangle}$ units of time after τ_i^j (i.e. $W_{\left\langle \tau_i^j, \tau_x^y \right\rangle}$ is a lower-bound on the amount of time between the start of τ_x^y and the finish of τ_i^j).

Equation (8) requires that the duration between the start of τ_i^j and the finish of τ_x^y is less than $D^{\text{rel}}_{\left\langle \tau_i^j, \tau_x^y \right\rangle}$ (i.e. $D^{\text{rel}}_{\left\langle \tau_i^j, \tau_x^y \right\rangle}$ is an upper-bound on the finish time of τ_x^y *relative* to the start of τ_i^j). Equation (9) requires that τ_i^j finishes before $D^{\text{abs}}_{\tau_i^j}$ units of time have expired since the start of the schedule (i.e. $D^{\text{abs}}_{\tau_i^j}$ is an upper-bound on the latest *absolute* time τ_i^j can be finished). Equations (10) and (11) enforce that agents can only execute one subtask at a time. Equations (12) and (13) enforce that each resource R_i can only be accessed by one agent at a time.

The worst-case time complexity of a complete solution technique for this problem is dominated by the binary decision variables for allocating tasks to agents ($A^a_{\tau_i^j}$) and sequencing ($J_{\left\langle \tau_i^j, \tau_x^y \right\rangle}$), and the complexity is given by $O\left(2^{|A||\boldsymbol{\tau}|^3}\right)$, where $|A|$ is the number of agents and $|\boldsymbol{\tau}|$ is the number of tasks. Agent allocation contributes $O\left(2^{|A||\boldsymbol{\tau}|}\right)$, and sequencing contributes $O\left(2^{|\boldsymbol{\tau}|^2}\right)$.

5. Scheduling mechanism

For all three experiments, we adapted a dynamic scheduling algorithm, called Tercio, to schedule the human–robot teams (Gombolay et al., 2013). Tercio is an empirically fast, high-performance dynamic scheduling algorithm designed for coordinating human–robot teams with upper- and lower-bound temporospatial constraints. The algorithm is designed to operate on a simple temporal network (Muscettola et al., 1998) with set-bounded uncertainty. If the schedule's execution exceeds its set bounds, Tercio reschedules the team (Gombolay et al., 2013).

As shown in Figure 1, the algorithm takes as input a temporal constraint problem, a list of agent capabilities (i.e. the lower-bound, upper-bound, 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 (14). In this equation, ***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

$$\boldsymbol{A}^* = \min_{\{A\}} \max_{\textbf{Agents}} \sum_{\tau_i^j} A^a_{\tau_i^j} \times C^a_{\tau_i^j}, \forall a \in \textbf{Agents} \tag{14}$$

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 test returns that this 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.

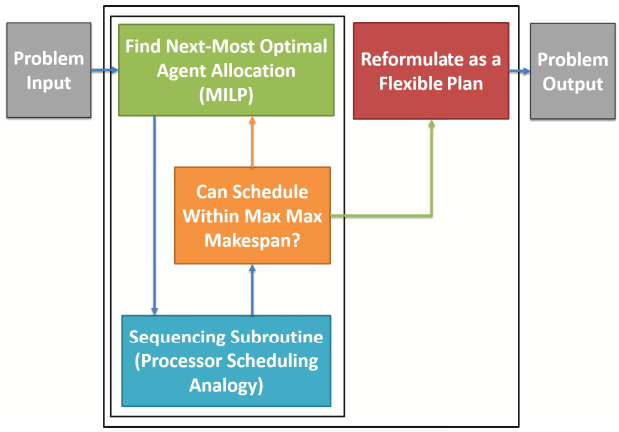

Fig. 1. Tercio takes as input a temporal constraint problem and finds a satisficing, flexible schedule by utilizing an analytical schedulability test to ensure a feasible solution. MILP: mixed-integer linear programming.

We employed Tercio because it allows for easy altering of task allocation within its task allocation subroutine. Here, we describe the specific Tercio alterations incorporated in each experiment. Note that only the task allocation subroutine within Tercio was modified for our three experiments; the sequencing subroutine remained unaltered.

5.1. Algorithm modifications for mixed-initiative scheduling

In the situational awareness experiment, we sought to determine whether situational awareness degrades as a robotic agent is allowed greater autonomy over scheduling decisions. We considered three conditions: autonomous, semi-autonomous, and manual control. Under the autonomous condition, the robotic teammate performed scheduling for the entire team; as such, the robot could use Tercio without modifications.

Under the semi-autonomous condition, in which the human participant decides which tasks to perform and the robotic agent decides how to allocate the remaining tasks between itself and a human assistant, Tercio was required to consider the tasks allocated by the participant. After the participant specified which tasks he or she would perform, the experimenter provided these assignments to the robot, which encoded the allocation as an assignment to the decision variables. Specifically, Tercio set $A^{\text{participant}}_{\tau^j_i} = 1$, $A^{\text{asst.}}_{\tau^j_i} = 0$, $A^{\text{robot}}_{\tau^j_i} = 0$ for subtasks τ^j_i assigned to the participant, and $A^{\text{participant}}_{\tau^y_x} = 0$ for subtasks τ^y_x that the participant left unassigned. Thus, the robot (via Tercio) only needed to solve for the allocation variables not already allocated by the participant.

Under the autonomous condition, the participant specified all task allocation assignments. As such, the robotic agent set $A^a_{\tau^j_i} = 1$ for all subtasks τ^j_i assigned to agent a, and $A^a_{\tau^y_x} = 0$ for all subtasks τ^y_x not assigned to agent a, for all agents a.

5.2. Algorithm modifications for scheduling with preferences

We focused on the effect of incorporating the preferences of human team members when generating a team's schedule. Preferences can exist in a variety of forms. For example, human team members may have preferences about the duration of events (how long it takes to complete a given task) or the duration between events (the lower-bound or upper-bound on the time between two tasks) (Wilcox et al., 2012). In our investigation, we considered preferences related to task types—for example, a worker may prefer to complete a drilling task rather than a painting task. Such preferences can be included in the mathematical formulation in equations (3) to (13) as an objective function term where one seeks to maximize the number of preferred tasks assigned to the participant, as shown in equation (15). In this equation, the objective function term for maximizing preferences is balanced with the established criteria (i.e. function

$$g\big(\{A^a_{\tau^j_i} | \tau^j_i \in \boldsymbol{\tau}, a \in A\}, \{J_{\langle \tau^j_i, \tau^y_x \rangle} | \tau^j_i, \tau^y_x \in \boldsymbol{\tau}\}, \{s_{\tau^j_i}, f_{\tau^j_i} | \tau^j_i \in \boldsymbol{\tau}\}\big)$$

from equation (3)) via a weighting parameter α

$$\min\ z, z = \alpha \times g\Big(\{A^a_{\tau^j_i} | \tau^j_i \in \boldsymbol{\tau}, a \in A\}, \{J_{\langle \tau^j_i, \tau^y_x \rangle} | \tau^j_i, \tau^y_x \in \boldsymbol{\tau}\}, \{s_{\tau^j_i}, f_{\tau^j_i} | \tau^j_i \in \boldsymbol{\tau}\}\Big) - (1-\alpha) \times \left(\sum_{\tau^j_i \in \boldsymbol{\tau}_{\text{preferred}}} A^{\text{participant}}_{\tau^j_i} \right) \tag{15}$$

Alternatively, one could incorporate preferences as a set of constraints on enforcement of a minimum or maximum level of preferred work assigned to the participant, as shown in equations (16) and (17). In these equations, $k^{\text{pref}}_{\text{ub}}$ and $k^{\text{pref}}_{\text{lb}}$ are upper- and lower-bounds on the number of preferred tasks allocated to the participant, and $k^{\text{pref}^c}_{\text{ub}}$ and $k^{\text{pref}^c}_{\text{lb}}$ are upper- and lower-bounds on the number of non-preferred tasks allocated to the participant.

$$k^{\text{pref}}_{\text{lb}} \le \sum_{\tau^j_i \in \boldsymbol{\tau}_{\text{pref}}} A^{\text{participant}}_{\tau^j_i} \le k^{\text{pref}}_{\text{ub}} \tag{16}$$

$$k^{\text{pref}^c}_{\text{lb}} \le \sum_{\tau^j_i \in \boldsymbol{\tau}_{\text{pref}}{}^c} A^{\text{participant}}_{\tau^j_i} \le k^{\text{pref}^c}_{\text{ub}} \tag{17}$$

We chose to model the inclusion of preferences as a set of constraints, which we added to Tercio's task allocation subroutine. For the purpose of human participant experimentation, where one must control for confounders, this

approach offers greater control over schedule content, as opposed to including a preference term within the objective function. The challenge of using an objective function model lies in the need to tune one or more coefficients (e.g. α in equation (15)) in the objective function to balance the contribution of the schedule efficiency (i.e. makespan) with the importance of adhering to preferences. We found this tuning to be difficult for a variety of participants.

For all three conditions, we set $k_{\text{lb}}^{\text{pref}} = k_{\text{lb}}^{\text{pref}^{\text{c}}} = 0$. Under the positive condition, participants could be assigned only one task that did not align with their preferences (i.e. $k_{\text{ub}}^{\text{pref}} = \infty$ and $k_{\text{ub}}^{\text{pref}^{\text{c}}} = 1$). Participants preferring to build could be assigned one fetching task at most, and vice versa. Under the negative condition, participants could be assigned a maximum of one task that aligned with their preferences (i.e. $k_{\text{ub}}^{\text{pref}} = 1$ and $k_{\text{ub}}^{\text{pref}^{\text{c}}} = \infty$); for example, participants preferring to build could be assigned one build task at most. Under the neutral condition, Tercio's task allocation subroutine would run without alteration (i.e. $k_{\text{ub}}^{\text{pref}} = k_{\text{ub}}^{\text{pref}^{\text{c}}} = 1$, $\tau_{\text{preferred}} = \emptyset$).

Based on results from previous studies indicating the importance of team efficiency (Gombolay et al., 2015, 2014), we sought to control for the influence of schedule duration on team dynamics. For the experiment studying scheduling preferences, we ran 50 iterations of Tercio for each participant under the positive, neutral, and negative parameter settings, generating a total of 150 schedules. We then identified a set of three schedules, one from each condition, for which the makespans were approximately equal. (We did not control for the workload of the individual agents.) The robot then used these schedules to schedule the team under the respective conditions.

5.3. Algorithm modifications for constraints based on workload and scheduling preference

In this experiment, we needed to control for makespan across all four conditions while varying the participants' workloads and the types of tasks they were assigned.

To control for the degree to which preferences were included in the schedule, we again added equations (16) and (17) to Tercio's task allocation subroutine. Under conditions with high preference, all tasks assigned to the participant were preferred tasks (i.e. $k_{\text{ub}}^{\text{pref}} = \infty$ and $k_{\text{ub}}^{\text{pref}^{\text{c}}} = 0$); under conditions with low preference, all tasks assigned to the participant were non-preferred tasks (i.e. $k_{\text{ub}}^{\text{pref}} = 0$ and $k_{\text{ub}}^{\text{pref}^{\text{c}}} = \infty$). Under all conditions, we set $k_{\text{lb}}^{\text{pref}} = k_{\text{lb}}^{\text{pref}^{\text{c}}} = 0$.

To control for the utilization of the participant, we added an objective function term to Tercio's task allocation subroutine that minimized the absolute value of the difference between the desired utilization of the participant U^{target} and the actual utilization of the participant $\sum_{\tau_i^j \in \boldsymbol{\tau}} A_{\tau_i^j}^{\text{participant}} \times lb_{\tau_i^j}$. Since the absolute value function is non-linear and cannot be handled by a linear program solver, we linearized the term as

$$z_{\text{utility}} \geq U^{\text{target}} - \sum_{\tau_i^j \in \boldsymbol{\tau}} A_{\tau_i^j}^{\text{participant}} \times lb_{\tau_i^j} \tag{18}$$

$$z_{\text{utility}} \geq -U^{\text{target}} + \sum_{\tau_i^j \in \boldsymbol{\tau}} A_{\tau_i^j}^{\text{participant}} \times lb_{\tau_i^j} \tag{19}$$

We generated schedules for each condition in three steps. First, we ran Tercio without any alterations to the task allocation subroutine for 100 iterations. Tercio works by iteratively generating task allocations and then sequencing the task set given the corresponding task allocation. Each iteration takes approximately one-third of a second. By running Tercio for several iterations, we allowed it to explore the search space so that it could then identify a candidate schedule with given characteristics (e.g. a specific degree of utilization of a particular agent). From these iterations, we recorded the median utilization U^{median} of the participant.

Next, we ran four additional sets of 100 iterations of Tercio—one set for each of the four conditions listed above. As before, we used equations (16) and (17) to control for the degree to which the robot included the participant's preferences while scheduling. When the preference variable was set to high, we set $k_{\text{ub}}^{\text{pref}} = \infty$ and $k_{\text{ub}}^{\text{pref}^{\text{c}}} = 0$; we set $k_{\text{ub}}^{\text{pref}} = 0$ and $k_{\text{ub}}^{\text{pref}^{\text{c}}} = \infty$ for the low preference condition. In both conditions, $k_{\text{lb}}^{\text{pref}} = k_{\text{lb}}^{\text{pref}^{\text{c}}} = 0$.

In the experiment studying workload, we controlled for the participant's utilization via equations (18) and (19). When the utilization variable was set to high, we set $U^{\text{target}} = U^{\text{median}}$. When the utilization variable was set to low, we set $U^{\text{target}} = U^{\text{median}}/2$.

We then identified one schedule from each of the four sets of 100 Tercio iterations to generate a set of schedules with short, approximately equal, makespans and utilizations close to their respective targets. To generate this set, we employed equation (20), which minimizes the difference between the longest and shortest makespans across the four conditions (i.e. $\max_{i,j}\left(m_i - m_j\right)$), the longest makespan (i.e. $\max_i m_i$), and the maximum difference between each schedule's target utilization U_i^{target} and its actual utilization U_i. In our experimental procedure, we set $\alpha_1 = \alpha_2 = 1, \alpha_3 = 2$

$$z_{\text{tuning}} = \alpha_1 \max_{i,j \in \text{ schedules}} \left(m_i - m_j\right) + \alpha_2 \max_{i \in \text{ schedules}} m_i + \alpha_3 \max_{i \in \text{ schedules}} \left(U_i^{\text{target}} - U_i\right) \tag{20}$$

6. Experimental design

We conducted a series of three human participant experiments ($n = 17$, $n = 18$, $n_3 = 20$) that required the fetching and assembly of Lego part kits. The goal of these experiments was to assess: (1) how a robotic teammate's inclusion of the preferences of its human teammates while

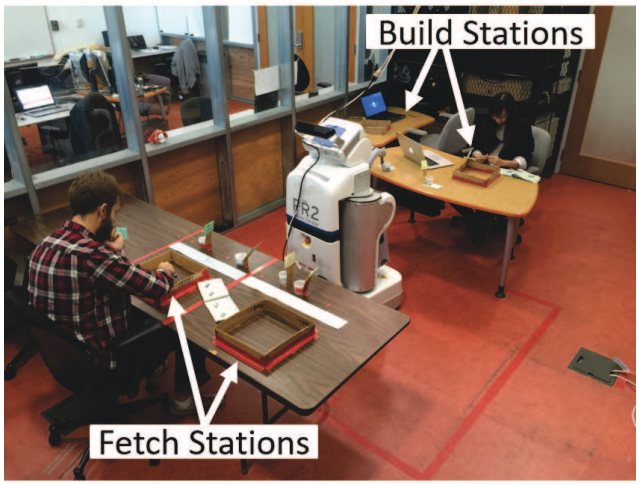

Fig. 2. Laboratory room where experiment took place. There were two locations where the human and robot workers could inspect part kits during a fetching task, and two locations where the human workers built part kits.

scheduling affects team dynamics, (2) how the benefits of including these scheduling preferences varies as a function of the degree to which the robot utilizes the human participant, and (3) how situational awareness degrades as a function of the level of autonomy afforded to the robot over scheduling decisions. We used the same basic experimental set-up for all three experiments, which we describe next.

6.1. Materials and set-up

Our human–robot manufacturing team consisted of a human participant, a robotic assistant, and a human assistant. The human participant was capable of both fetching and building, while the robot assistant was only capable of fetching. One of the experimenters played the role of a third teammate (the human assistant) for all participants and was capable of both fetching and building. This human assistant was included to represent the composition of a human–robot team more realistically within a manufacturing setting. We used a Willow Garage PR2 platform, depicted in Figure 2, as the robotic assistant for our human–robot team. The robot used adaptive Monte Carlo localization (Fox, 2003) and the standard *Gmapping* package in the Robot Operating System for navigation.

6.2. Procedure

The scenario included two types of task: fetching and assembling part kits. As shown in Figure 2, the experiment environment included two fetching stations and two build stations, with four part kits located at each fetching station.

Fetching a part kit involved moving to one of two fetching stations where the kits were located, inspecting the part kit, and carrying it to the build area. The architecture of our fetching task is analogous to actions required in many manufacturing domains. To adhere to strict quality assurance standards, fetching a part kit required verification from one or two people that all of the correct parts were present in the kit, as well as certification from another person that the kit had been verified. We also imposed additional constraints to better mimic an assembly manufacturing environment. A part kit had to be fetched before it could be built, and no two agents were able to occupy the same fetching or build station at the same time.

Agents were required to take turns using the fetching stations, as allowing workers to sort through parts from multiple kits at the same location risked the participants mixing the wrong part with the wrong kit. Furthermore, 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 cease until the missing item has been found. As such, we imposed a 10 min deadline from the time that the fetching of a part kit began until that kit had been built.

Assembly of the Lego model involved eight tasks $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_8\}$, each of which consisted of a *fetch* and *build* subtask $\tau_i = \{\tau_i^{\text{fetch}}, \tau_i^{\text{build}}\}$. The amount of time each participant took to complete each subtask $C_i^{\text{participant-fetch}}$ and $C_i^{\text{participant-build}}$ was measured during a training round. The timings for the robot $C_i^{\text{robot-fetch}}$ and human assistant $C_i^{\text{assist-fetch}}$ and $C_i^{\text{assist-build}}$ (as measured by an experimenter) were collected prior to the experiments.

In all three experiments, the robotic agent employed Tercio as a dispatcher, communicating to the participant and human assistant when to initiate their next subtasks. Tercio would tell agents when they were each able to initiate or complete each subtask, and each agent would send a message acknowledging initiation or completion via simple, text-based messages over a TCP/IP GUI (SocketTest v3.0.0 ©2003–2008 Akshathnkumar Shetty, `http://sockettest.sourceforge.net/`).

6.2.1. Modifications for the experiment studying situational awareness. For the study evaluating the effects of mixed-initiative scheduling on the situational awareness of the human team members, we performed a between-participants experiment, where each participant experienced only one of three conditions: autonomous, semi-autonomous, or manual.

As stated previously, under the autonomous condition, the robot scheduled the three members of the team using Tercio with the default task allocation subroutine. Under the semi-autonomous condition, participants each selected which tasks they would perform and the robot allocated the remaining tasks to itself and the human assistant. Under the manual condition, the participant allocated tasks to each of the team members. The robot sequenced the tasks under all conditions.

After the human participant or the robot completed the task allocation and sequencing process, the participants were allowed 3 min to review the schedule. We found in

prior work that participants required ≈3 min to perform task allocation (Gombolay et al., 2015); therefore, we wanted to allow participants at least this much time to review a robot-generated schedule under the autonomous condition. Participants were not told that they would be later questioned about their experiences because we did not want to unduly bias them to focus on preparing for such a test. Instead, we wanted participants to attend fully to the task at hand.

After the participants reviewed the schedule, the team executed their tasks according to that schedule. At approximately 200 s into task execution, the experimenter halted the process and administered the post-trial questionnaire (Table 1) according to the SAGAT. The timing of the intervention was tuned to allow each team member to have been assigned at least one task on average. The team did not complete the schedule after the SAGAT test; the experiment concluded following administration of the questionnaire.

6.2.2. Extensions for the experiment to study scheduling preferences. For the experiment to study scheduling preferences, we employed a within-participants design. Participants experienced all three experimental conditions: positive, neutral, and negative. The order in which participants experienced these conditions was randomized. Participants were randomly assigned to these conditions. At the beginning of each condition, participants were told that their robot teammate wanted to know whether they preferred to complete fetch tasks or build tasks, and the participants responded accordingly.

Deference to the participants with regard to their preferred tasks is in keeping with a pseudo-experiment. We did not attempt to balance participants according to the number in our sample who preferred fetching vs. building, as 14 of 18 participants (78%) preferred building tasks. Participants were not informed a priori of the different conditions; as such, subjective evaluations of team dynamics under each condition would not be influenced by the expectation that the robot would or would not cater to the participants' preferences.

The preferences, along with task completion times for each of the three team members, were provided to the robot, which scheduled the team. The team then performed the tasks to completion. After the schedule was completed, participants completed the post-trial questionnaire given in Table 3. This process was repeated once for each condition, as indicated previously. After completing the tasks under all three conditions, participants completed the post-test questionnaire shown in Table 4. The experiment concluded after completion of this questionnaire.

6.2.3. Extensions for the experiment to study workload. For the experiment to study workload influence, we employed an experimental design that mirrored the procedure for the experiment studying workflow preferences, with one primary difference: We varied the workload and the degree to which human preferences were considered during scheduling, rather than preferences alone. Participants were not informed about whether the robot was varying their utilization, and the schedule itself was not reported to the participant; participants had to infer changes to their degree of utilization based only on their subjective experience.

7. Results

In this section, we report the results from statistical analysis of our experiments. Statistical significance is measured at the $\alpha = 0.05$ level.

7.1. Participants

We recruited participants for all three experiments from a local university. The cohort for the situational awareness study consisted of 20 participants (6 men and 14 women) with an average age of 19.5 ± 1.95 years (range, 18 to 25 years). The cohort for the study of scheduling preferences included 18 participants (10 men and 8 women) with an average age of 27 ± 7 years (range, 19 to 45 years). The cohort for the workload study consisted of 18 participants (10 men and 8 women) with an average age of 21 ± 3 years (range, 18 to 30 years). In all experiments, participants were assigned to the various experimental conditions via random sampling without replacement, so as to balance participants across the conditions.

7.2. Results for situational awareness

Recall that the associated hypothesis (Hypothesis 1) states that human participants' situational awareness would decline as the robot's autonomy over scheduling decisions increased.

We administered a SAGAT-based test in which participants completed a questionnaire consisting of both objective and subjective measures. We observed statistically significant decreases in situational awareness and participants' confidence in their situational awareness while under the autonomous condition, when the robot had full control over scheduling decisions.

Figure 3 depicts the team situational awareness score for questions 1, 5, 9, and 13 from the post-trial questionnaire (Table 1). For clarity when comparing the results from each question, we have normalized the values for each question in Figure 3, such that the maximum team score for each question is equal to 1.

We conducted a mixed-factor analysis of variance (ANOVA) for question 1, and observed a statistically significant difference for participants' responses to question 1 ($F(2, 17) = 3.894$, $p < 0.041$) across the three conditions. Results from a pairwise comparison with Student's t test indicated that participants were statistically significantly more accurate when recalling which action team members performed under the semi-autonomous condition ($M = 0.67$, $SD = 0.48$) than the autonomous condition

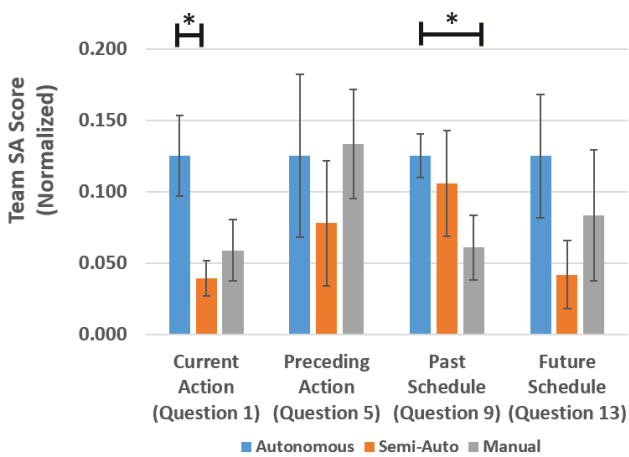

Fig. 3. Participants' average SA scores for questions 1, 5, 9, and 13 in the post-trial questionnaire of Table 1, as a function of the degree of automation over scheduling decisions. Whisker bars indicate standard error of the mean. A lower score indicates a better situational awareness.

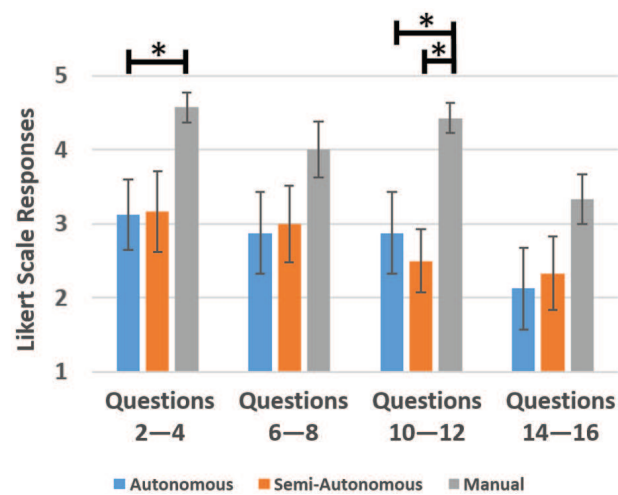

Fig. 4. Average of medians of participants' responses to Likert response questions 2–4, 6–8, 10–12, and 14–16 under autonomous, semi-autonomous, and manual conditions. Whisker bars indicate standard error of the mean.

($M = 2.13$, $SD = 1.36$), with $t(12), p < 0.014$. The manual condition ($M = 1.00$, $SD = 0.89$) was not statistically significantly different from the other two conditions.

We also applied a set of pairwise t tests with a Bonferroni correction $\alpha' = \alpha/3 = 0.05/3 = 0.01\bar{6}$ for responses to question 9, and found that participants were less accurate when recalling all previous actions of each agent under the autonomous condition ($M = 7.88$, $SD = 2.75$) than the manual condition ($M = 3.38$, $SD = 3.49$), with $p < 0.0158$. There was no statistically significant difference with regard to the semi-autonomous ($M = 6.67$, $SD = 5.71$) condition.

Next, we considered participants' responses to the set of participant questions from Table 1: questions 2–4, 6–8, 10–12, and 14–16. Choosing the correct test was challenging for our design. We wanted a composite measure for confidence, which combines the responses to questions 2–4, and likewise for questions 6–8, 10–12, and 14–16, as a repeated measure. However, we could not immediately apply an ANOVA because the data were on an ordinal rather than an interval scale.

We performed two types of analysis for these data. First, we used a non-parametric analysis, which assumes ordinal, non-normally distributed data. To measure the confidence of an individual participant under a given condition for the current actions of agents (referring to questions 2–4), we used the median of the answers to the relevant questions as our single data point. We then compared the set of medians, which is notionally more robust to within-participant variance, for participants under each condition, with the medians for other participants. For a qualitative description, a histogram of the medians is given in Figure 4.

Results from an omnibus Kruskal–Wallis test indicated significant differences across the conditions with regard to participants' confidence in their situational awareness for the current activities (questions 2–4; $\chi^2(2) = 6.09$, $p = 0.0476$), and past activities of their team members (questions 10–12; $\chi^2(2) = 7.98$, $p = 0.018$). A pairwise Kruskal–Wallis test indicated that participants were statistically significantly more confident in their situational awareness for the current activities of team members (questions 2–4) when under the manual condition than the semi-autonomous ($\chi^2(1) = 5.61$, $p < 0.018$) or autonomous ($\chi^2(1) = 4.04$, $p < 0.044$) conditions. Likewise, we found that participants were statistically significantly more confident in their situational awareness for the current activities of team members (questions 10–12) under the manual condition than the semi-autonomous ($\chi^2(1) = 7.93$, $p < 0.005$) or autonomous ($\chi^2(1) = 4.15$, $p = 0.0416$) conditions.

In our second analysis, we treated the data as interval data. Prior work has included extensive analysis suggesting that one can reasonably approximate a symmetric Likert response format as interval data (Carifio, 1976, 1978), and that the F test is quite robust with respect to breaking the assumptions of normality with regard to interval data (Glass et al., 1972). We applied a mixed-factor ANOVA and used this test to measure the composite confidence for the sets of questions corresponding to current, preceding, past, and future actions of team members.

After applying a mixed-factor ANOVA, we found that the level of robotic autonomy over scheduling decisions affected participants' confidence in their knowledge of the current actions of their team (questions 2–4, $F(2, 18) = 4.228$, $p < 0.031$), as well as their confidence in their knowledge of the team's previous actions (questions 6–8, $F(2, 18) = 6.293$, $p < 0.008$). These findings support the results of the Kruskal–Wallis test.

Table 6. Correlation between utilization and participants' perception of the team ($N = 17$).

Question	Correlation coefficient, r	t	p
7	0.285	2.084	0.021
10	0.311	2.287	0.013
14	0.286	2.086	0.021
15	0.269	1.957	0.028

On performing pairwise comparisons of the autonomous, semi-autonomous, and manual conditions using the mixed-factor ANOVA, we again observed statistically significantly greater confidence in situational awareness among participants with regard to the current activities of team members (questions 2–4) under the manual condition than in the autonomous condition ($F(1,13) = 11.377$, $p = 0.005$). Likewise, we found that participants were statistically significantly more confident in their situational awareness about the current activities of team members (questions 10–12) when under the manual condition than the semi-autonomous ($F(1,11) = 18.615$, $p = 0.001$) or autonomous ($F(1,11) = 8.960$, $p = 0.010$) conditions. These findings corroborate those obtained from non-parametric testing and strongly suggest that participants have less confidence in their situational awareness when under the autonomous condition.

7.3. Results for scheduling preferences

Recall that Hypothesis 2 states that human participants would prefer to work with a robot when it includes their workflow preferences in scheduling decisions. Based on responses to questions 22–24 in Table 4, we observed statistically significant evidence that human participants preferred working with a robot that included their preferences when scheduling ($p < 0.001$). Participants reported that they would rather work with a robotic teammate that included their preferences than with one that was unaware of their preferences ($p < 0.001$). Furthermore, participants reported that they would prefer to work with a robot that was unaware of their preferences than a robot that scheduled according to the opposite of their preferences ($p < 0.001$). We also found that participants felt the robot liked them more (question 6 in Table 4) under the neutral condition, when the robot was unaware of the participants' preferences, than under the negative condition ($p < 0.05$). These results support our hypothesis that the preferences of human workers are important for a robotic teammate to include when making scheduling decisions.

Surprisingly, we also found that the amount of work allocated to participants had a strong impact on their subjective perceptions of their teams' interactions. In post-hoc analysis, we computed the Pearson product–moment correlation coefficient for the rank of participants' Likert-scale responses to questions from the post-trial questionnaire

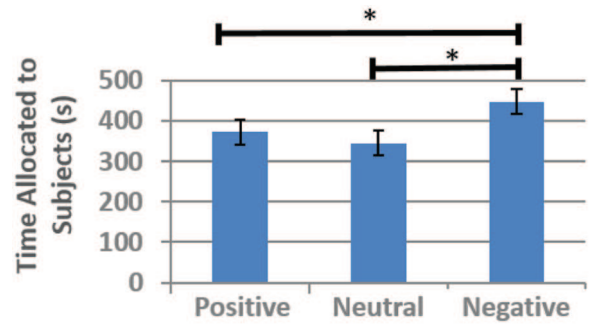

Fig. 5. Mean and standard error of the amount of work, in seconds, assigned to the participant by the robotic teammate. Horizontal bars with an asterisk denote statistical significance ($p < 0.01$).

(Table 3) for each condition as a function of the amount of work assigned to each participant; we found that a statistically significant proportion of responses (Likert-scale responses on 16 of the 21 questions) positively correlated with the amount of time assigned to participants ($\chi^2 = 5.762, p = 0.016$). (Questions 1–7, 9–10, 12–16, 18, and 20 from Table 3 showed that participants' responses were positively correlated with their utilization.) Furthermore, 4 of the 16 responses with a positive correlation (those to questions 7, 10, 14, and 15 from Table 3) were statistically significantly correlated, as shown in Table 6 ($p < 0.05$). We did not observe a statistically significant negative correlation with the amount of time assigned to participants.

To investigate this finding further, we conducted a variance analysis and found that the robot allocated a statistically significantly different amount of work to the participant as a function of how the robot included the participant's preferences when scheduling, as shown in Figure 5 (ANOVA $F(2,48) = 5.16$, $p = 0.009$). Interestingly, we also found that participants were allocated statistically significantly more work, as measured in seconds, when under the negative condition ($M = 448$, $SD = 113$) compared with the positive ($M = 373$, $SD = 92$; $t(16) = 1.86$, $p = 0.04$) or neutral ($M = 345$, $SD = 82$; $t(17) = 2.14$, $p = 0.03$) conditions.

In collecting participants' preferences for the types of tasks they would rather complete, we found that 14 of 18 participants (78%) reported that they preferred to build the part kits rather than fetch them. Under the positive condition, participants received a maximum of one fetching task; under the negative condition, they received a maximum of one building task. The third teammate (the human assistant) was typically more proficient at building than the average participant; consequently, the optimal work allocation would typically assign the majority of building tasks to the assistant, with the participant providing support by performing more fetching tasks. (The robot teammate was only able to fetch part kits.) Thus, the negative condition afforded participants the opportunity to complete a larger

Table 7. Values of p for statistically significant post-trial questions ($N = 18$). Statistically significant values are in bold.

Question	Omnibus	High vs. low utility	High vs. low preference
2	**$p = 0.013$**	**$p = 0.002$**	$p = 0.096$
5	**$p = 0.010$**	**$p = 0.003$**	**$p = 0.020$**
7	**$p = 0.026$**	**$p = 0.035$**	**$p = 0.016$**
9	**$p < 0.001$**	**$p < 0.001$**	$p = 0.170$
10	**$p < 0.001$**	**$p = 0.007$**	$p = 0.061$
11	**$p = 0.026$**	**$p = 0.027$**	**$p = 0.029$**
13	**$p < 0.001$**	**$p = 0.001$**	**$p = 0.001$**
14	**$p < 0.001$**	**$p < 0.001$**	**$p = 0.004$**
15	**$p < 0.001$**	**$p = 0.005$**	**$p = 0.001$**
16	**$p = 0.010$**	**$p = 0.011$**	**$p = 0.003$**
17	**$p < 0.001$**	**$p = 0.011$**	**$p < 0.001$**
18	**$p = 0.004$**	**$p = 0.012$**	**$p = 0.012$**
20	**$p = 0.026$**	$p = 0.052$	**$p = 0.013$**

share of the work. Based on this result, we propose that participants' preferences for task types must be balanced with an innate desire on the part of a human worker to be an important contributor to his or her team.

7.4. Results of varying workload

Recall that Hypothesis 3A states that participants would prefer to work with a robotic agent that included their workflow preferences, and Hypothesis 3B states that participants would prefer working with a robotic agent that provided them with a relatively high workload. To test our hypotheses, we designed and conducted an experiment controlling for both the degree to which preferences were included and the degree to which participants were utilized, and report the results here.

Across 12 measures of the participants' perceptions about the human–robot team, we observed statistically significant evidence that participants preferred working with a robot that included their preferences when scheduling and, across 10 measures, that participants preferred working with a robot that utilized them more frequently. Results are given in Table 7. For the findings reported here, we first used an omnibus Friedman test to determine that a statistically significant difference existed across all conditions, and then applied a pairwise Friedman test to examine the differences between the conditions.

Our experiment in which a robotic agent included the preferences of a human team member when scheduling provided initial evidence that human participants would prefer to work with a robotic agent when it considered their workflow preferences and more frequently utilized their time. We can state that these data statistically significantly support our hypotheses. When a robot schedules for a human–robot team, the human team members' perceptions of the robot and of the team as a whole are significantly improved when the robot considers the preferences of the human worker and utilizes more of the workers' time.

In addition to these findings, we discovered a surprising trend between preferences, utilization, and the participants' perception of team efficiency in post-hoc analysis. Under the high preference–high utilization condition, participants felt more strongly that the team performed tasks using the least possible amount of time, even though the schedule duration (i.e. makespan) was constant across all trials within participants ($p < 0.004$). In the interests of further investigation, we propose a follow-on study, examining how human team members' perceptions of the passage of time and team efficiency is affected by the way in which a robot schedules the team.

8. Discussion

8.1. Design guidance for roboticists

We investigated key gaps in prior literature by assessing how situational awareness is affected by the level of autonomy in mixed-initiative scheduling for human–robot teams, the effects of increased or decreased workload in human–robot team fluency, and the role of workflow preferences in robotic scheduling. Based on our findings, we can provide design guidance for roboticists developing intelligent collaborative robots that engage in mixed-initiative decision-making with human participants.

Human situational awareness is poorer when the robotic agent has full autonomy over scheduling decisions, as assessed by both objective and subjective measures. However, prior research has indicated that decreasing robotic autonomy over scheduling decisions reduces efficiency and decreases the desire of the human participant to work with a robotic agent. Therefore, the positive and negative effects of increasing the robot's role in decision-making must be carefully weighed. If there is a high probability that the human agent will have to intervene to adjust work allocations, or if the potential cost of poorer human performance due to reduced situational awareness is high, then we recommend that the human participant retain the primary decision-making authority. If human intervention is unlikely, or the cost of poorer human performance is low, then the benefits of improved team efficiency can be safely achieved by allowing the robot to retain primary decision-making authority. In many applications, a mixed-initiative approach in which the participant and robot collaborate to make decisions offers a suitable middle ground between the two ends of this spectrum.

In addition, a human participant's perception of a robotic teammate scheduling a team's activities may improve when the human participant is scheduled to complete tasks that he or she prefers. However, human team members' perception of the robot may be negatively impacted when they are scheduled to be idle for much of the time. Providing human team members with more highly preferred tasks at the cost of decreasing the total amount of work assigned to them may, in fact, have more of a negative impact than assigning human team members less-preferred tasks. Although

the degree to which these variables interact is likely to be application-specific, it cannot be assumed that increasing one criterion at the cost of the other will improve team fluency.

Collaborations with robots that participate in decision-making related to the planning and scheduling of work present unique challenges with regard to preserving human situational awareness and optimizing workload allocation to human teammates while also respecting their workflow preferences. Careful consideration is necessary in order to design intelligent collaborative robots that effectively balance the benefits and detriments of maintaining an increased role in the decision-making process.

8.2. Limitations and future work

There are limitations to our findings. Our sample population consisted of young adults enrolled from a local university campus, whereas the target population consists of older, working adults in the fields of manufacturing and search-and-rescue, among other domains. Impressions of robotic teammates, in general, may differ significantly between these populations.

Workers may also use different criteria to evaluate a human–robot team. For example, if chronic fatigue is an issue in a given setting, workers may prefer a greater amount of idle time. Also, we limited the expression of preferences to a binary choice between two types of task; however, the preferences of real workers may be more nuanced and difficult to encode computationally. For these reasons, we recommend a follow-on study, conducted in multiple factories across a variety of industries and work environments, to confirm the results of our experiments.

We studied one robot form factor (i.e. a PR2) in our investigation. It is possible that other form factors could elicit a different response from participants. Further, we used a specific scheduling technique, Tercio, well-suited for human–robot teaming. It is possible that alternate scheduling algorithms could alter participants' experience.

When manipulating the degree to which participants are utilized and the amount of preferred work assigned to those participants, we used "high" and "low" settings. We found that increasing the setting of these independent variables from low to high positively affected participants' experience working with the robot. It is possible, however, that the relationship between utilization and participants' subjective experience is not linear. For example, an "extremely high" utilization could be less desirable even than low utilization. Future work should investigate utilization and workflow preferences across the entire spectrum.

9. Conclusions

While new computational methods have significantly enhanced the ability of people and robots to work flexibly together, there has been little study of the ways in which human factors must influence the design of these computational techniques. In this work, we investigated how situational awareness varies as a function of the degree of autonomy a robotic agent has during scheduling, and found that human participants' awareness of their team's actions decreased as the degree of robot autonomy increased. This indicates that the desire for increased autonomy and accompanying performance improvements must balanced with the risk of—and cost resulting from—reduced situational awareness. We also studied how team fluency varies as a function of the workload given to a human team member by a robotic agent, and the manner in which a robot should include the workflow preferences of its human teammates in the decision-making process. Results indicate a complex relationship between preferences, utilization, and the participants' perception of team efficiency. The three study results provide guidelines for the development of intelligent collaborative robots, and a framework for weighing the positive and negative effects of increasing the robot's role in decision-making.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Foundation Graduate Research Fellowship Program (grant number 2388357).

References

Adams JA (2009) Multiple robot-single human interaction: effects on perceived workload and performance. *Behavior and Information Technology* 28(2): 183–298.

Alami R, Chatila R, Clodic A, et al. (2006) Towards human-aware cognitive robots. In: *The fifth international cognitive robotics workshop (the AAAI-06 workshop on cognitive robotics)*, Boston, MA, 16–17 July 2006. Palo Alto, CA: AAAI.

Alsever J (2011) Robot workers take over warehouses. *CNN Money*. Available at: http://money.cnn.com/ 2011/11/09/small-business/kiva_robots/.

Ardissono L, Petrone G, Torta G, et al. (2012) Mixed-Initiative Management of Online Calendars. In: Cordeiro, José, Krempels and Karl-Heinz (eds) *Web Information Systems and Technologies 8th International Conference, WEBIST 2012*. Porto, Portugal, 18–21 April 2012. Available at: http://www.springer.com/us/book/9783642366079.

Baraglia J, Cakmak M, Nagai Y, et al. (2016) Initiative in robot assistance during collaborative task execution. In: *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*, Christchurch, New Zealand, 7–10 March 2016, pp.67–74. Piscataway, NJ: IEEE.

Barnes MJ, Chen JYC, Jentsch F, et al. (2011) Designing effective soldier-robot teams in complex environments: training, interfaces, and individual differences. In: Harris D (ed.) *Engineering Psychology and Cognitive Ergonomics*, Berlin: Springer, pp.484–493.

Bauer A, Wollherr D, and Buss M (2008) Human–robot collaboration: a survey. *International Journal of Humanoid Robotics* 5(1): 47–66.

Berry P, Peintner B, Conley K, et al. (2006) Deploying a personalized time management agent. In: *Proceedings of the fifth international joint conference on autonomous agents and multiagent systems*, Hakodate, Japan, 8–12 May 2006, pp.1564–1571. New York: ACM.

Bertsimas D and Weismantel R (2005) *Optimization Over Integers*. Belmont: Dynamic Ideas.

Boeing (2014) A futuristic view of the 777 fuselage build. Available at: http://www.boeing.com/features/2014/07/bca-777-fuselage-07-14-14.page. (accessed 28 February 2016)

Brookings JB, Wilson GF, and Swain CR (1996) Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology* 42(3): 361–377.

Brunet L, Choi HL, and How JP (2008) Consensus-based auction approaches for decentralized task assignment. In: *Proceedings of the AIAA guidance, navigation, and control conference (GNC)*, Honolulu, HI, 18–21 August 2008. Reston, VA:L AIAA.

Carifio J (1976) Assigning students to career education programs by preference: scaling preference data for program assignments. *Career Education Quarterly* 1: 7–26.

Carifio J (1978) Measuring vocational preferences: ranking versus categorical rating procedures. *Career Education Quarterly* 1978 3: 34–66.

Chen J, Haas E, and Barnes M (2007) Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 37(6): 1231–1245.

Clare AS, Cummings ML, How JP, et al. (2012) Operator objective function guidance for a real-time unmanned vehicle scheduling algorithm. *AIAA Journal of Aerospace Computing, Information and Communication* 9: 161–173.

Cummings ML, Brzezinski AS, and Lee JD (2007) Operator performance and intelligent aiding in unmanned aerial vehicle scheduling. *IEEE Intelligent Systems* 22(2): 52–59.

Dragan AD and Srinivasa SS (2012) Assistive teleoperation for manipulation tasks. In: *Proceedings of the seventh annual ACM/IEEE international conference on human-robot interaction*, Boston, MA, 5–8 March 2012, pp.123–124. New York: ACM.

Dragan A and Srinivasa S (2013) A policy-blending formalism for shared control. *International Journal of Robotics Research* 32(7): 790–805.

Dragan A, Lee K, and Srinivasa S (2013) Teleoperation with intelligent and customizable interfaces. *Journal of Human-Robot Interaction* 2(2): 33–57.

Drury JL, Riek L, and Rackliffe N (2006) A decomposition of UAV-related situation awareness. In: *Proceedings of the ACM SIGCHI/SIGART conference on human-robot interaction*, Salt Lake City, UT, 2–3 March 2006, pp.88–94. New York: ACM.

Durfee EH, Boerkoel JC, and Sleight J (2014) Using hybrid scheduling for the semi-autonomous formation of expert teams. *Future Generation Computer Systems* 31: 200–212.

Endsley MR (1988) Situation awareness global assessment technique (SAGAT). In: *Proceedings of the IEEE 1988 national aerospace and electronics conference*, Dayton, OH, 23–27 May 1988, vol. 3, pp.789–795. Piscataway, NJ: IEEE.

Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37(1): 32–64.

Endsley MR and Kaber DB (1999) Level of automation effects on performance, situation awareness, and workload in dynamic control task. *Ergonomics* 42(3): 462–492.

Fong T and Thorpe C (2001) Vehicle teleoperation interfaces. *Autonomous Robots* 11(1): 9–18.

Fong T, Thorpe C, and Baur C (2003) Collaboration, dialogue, human-robot interaction. In: Jarvis RA and Zelinsky A (eds.) *Robotics Research*. Berlin: Springer, pp.255–266.

Fox D (2003) Adapting the sample size in particle filters through KLD-sampling. *International Journal of Robotics Research* 22: 985–1003.

Glass GV, Peckham PD, and Sanders JR (1972) Consequences of failure to meet assumptions underlying the analyses of variance and covariance. *Review of Educational Research* 42(3): 237–288.

Gombolay MC, Gutierrez RA, Clarke SG, et al. (2015) Decision-making authority, team efficiency and human worker satisfaction in mixed human-robot teams. *Autonomous Robots* 39(3): 293–312.

Gombolay MC, Gutierrez RA, Sturla GF, et al. (2014) Decision-making authority, team efficiency and human worker satisfaction in mixed human-robot teams. In: *Robotics: science and systems* (ed D Fox, LE Kavraki, and H Kurniawati), Berkeley, CA, 12–16 July 2014. Cambridge, MA: RSS foundation.

Gombolay MC, Wilcox RJ, and Shah JA (2013) Fast scheduling of multi-robot teams with temporospatial constraints. In: *Robots: science and systems)* (ed P Newman, D Fox, and D Hsu), Berlin, Germany, 24–28 June 2013. Cambridge, MA: RSS Foundation.

Goodrich MA, Morse BS, Engh C, et al. (2009) 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.

Grano MLD, Medeiros DJ, and Eitel D (2009) Accommodating individual preferences in nurse scheduling via auctions and optimization. *Healthcare Management Science* 12: 228–242.

Hamasaki M, Takeda H, Ohmukai I, et al. (2004) Scheduling support system for academic conferences based on interpersonal networks. In: *Proceedings of the fifteenth ACM conference on hypertext and hypermedia*, Santa Cruz, CA, 9–13 August 2004. New York: AMC.

Hart SG and Staveland LE (1988) Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*, 52: 139–183.

Hawkins KP, Bansal S, Vo NN, et al. (2014) Anticipating human actions for collaboration in the presence of task and sensor uncertainty. In: *2014 IEEE international conference on robotics and automation (ICRA)*, Hong Kong, 31 May–7 June 2014, pp.2215–2222. Piscataway, NJ: IEEE.

Haynes T, Sen S, Arora N, et al. (1997) An automated meeting scheduling system that utilizes user preferences. In: *Proceedings of the first international conference on autonomous agents, AGENTS '97*, Marina del Rey, CA, 5–8 February 1997, pp.308–315. New York: ACM.

Herlant L, Holladay R, and Srinivasa S (2016) Assistive teleoperation of robot arms via automatic time-optimal mode switching. In: *Eleventh ACM/IEEE international conference on human robot interaction*, Christchurch, New Zealand, 7–10 March 2016, pp.35–42. Piscataway, NJ: IEEE.

Hoffman G (2013) Evaluating fluency in human-robot collaboration. In: *Robotics: Science and systems workshop on human-robot collaboration*, Berlin, Germany, 24–28 June 2013. Cambridge, MA: RSS Foundation.

Hoffman G and Breazeal C (2004) Collaboration in human-robot teams. In: *Proceedings of the AIAA 1st intelligent systems technical conference*, Chicago, IL, 20–22 September 2004. Reston, VA: AIAA.

Hoffman G and Breazeal C (2007) Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team. In: *Proceedings of the ACM/IEEE international conference on human-robot interaction*, Arlington, VA, 10–12 March 2007, pp.1–8. New York: ACM.

Hoffman G and Ju W (2014) Designing robots with movement in mind. *Journal of Human-Robot Interaction* 3(1): 89–122.

Hooten ER, Hayes ST, and Adams JA (2011) A comparison of communicative modes for map-based tasking. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 55(1): 505–509.

Horvath AO and Greenberg LS (1989) Development and validation of the working alliance inventory. *Journal of Counseling Psychology* 36(2): 223.

Huang CM and Mutlu B (2016) Anticipatory robot control for efficient human-robot collaboration. In: *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*, Christchurch, NZ, 7–10 March 2016, pp.83–90. Piscataway, NJ: IEEE.

Jones E, Dias M, and Stentz A (2011) Time-extended multi-robot coordination for domains with intra-path constraints. *Autonomous Robots* 30(1): 41–56.

Jones HL, Rock SM, Burns D, et al. (2002) Autonomous robots in SWAT applications: Research, design, and operations challenges. *Proceedings of the 2002 symposium for the association of unmanned vehicle systems international (AUVSI '02*, 9–11 July 2002, Orlando, FL, 2002. Arlington, VA: Association for Unmanned Vehicle Systems International.

Ju W and Sirkin D (2010) Animate objects: How physical motion encourages public interaction. In: Ploug T, Hasle P, and Oinas-Kukkonen H (eds.) *Persuasive Technology*. Berlin: Springer, pp.40–51.

Kaber DB and Endsley MR (1997) Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress* 16(3): 126–131.

Kaber DB and Endsley MR (2004) The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science* 5(2): 113–153.

Kaber DB, Onal E, and Endsley MR (2000) Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing & Service Industries* 10(4): 409–430.

Klemmer SR, Hartmann B, and Takayama L (2006) How bodies matter: five themes for interaction design. In: *Proceedings of the 6th conference on designing interactive systems*, University Park, PA, 26–28 June 2006, pp.140–149. New York: ACM.

Knight W (2013) Smart robots can now work right next to auto workers. *MIT Technology Review* Available at: https://www.technologyreview.com/s/518661/smart-robots-can-now-work-right-next-to-auto-workers/. (accessed 28 February 2016)

Korsah GA, Stentz A, and Dias MB (2013) A comprehensive taxonomy for multi-robot task allocation. *The International Journal of Robotics Research* 32(12): 1495–1512.0

Kramer AF (1991) Physiological metrics of mental workload: A review of recent progress. In: D Damos (ed.) *Multiple-Task Performance*. London: Taylor & Francis, 279–328.

Kwon WY and Suh IH (2012) A temporal Bayesian network with application to design of a proactive robotic assistant. In: *2012 IEEE international conference on robotics and automation (ICRA)*, St Paul, MN, 14–18 May 2012, pp.3685–3690. Piscataway, NJ: IEEE.

Lasota PA and Shah JA (2015) Analyzing the effects of human-aware motion planning on close-proximity human–robot collaboration. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57(1): 21–33.

Lee KM, Jung Y, Kim J, et al. (2006a) Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. *International Journal of Human-Computer Studies* 64(10): 962–973.

Lee KM, Peng W, Jin SA, and Yan C (2006b) Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of Communication* 56(4): 754–772.

Liu C, Hamrick JB, Fisac JF, et al. (2016) Goal inference improves objective and perceived performance in human-robot collaboration. In: *Proceedings of the 2016 international conference on autonomous agents & multiagent systems*, Singapore, 9–13 May 2016, pp.940–948. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Loft S, Sanderson P, Neal A, et al. (2007) Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 49(3): 376–399.

Lokhande K and Reynolds HJD (2012) Cognitive workload and visual attention analyses of the air traffic control tower flight data manager (TFDM) prototype demonstration. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56(1): 105–109.

Lottaz C (1996) Constraint solving, preference activation and solution adaptation in idiom. Technical report. Report no. LIA-REPORT-1996-005). Swiss Federal Institute of Technology (EPFL).

Macho S, Torrens M, and Faltings B (2000) A multi-agent recommender system for planning meetings. In: *Proceedings of the ACM conference on autonomous agents, workshop on agent-based recommender systems*. 4 June 2000. New York, NY: ACM.

Matthias B, Kock S, Jerregard H, et al. (2011) Safety of collaborative industrial robots: Certification possibilities for a collaborative assembly robot concept. In: *Proceedings of the IEEE international symposium on assembly and manufacturing (ISAM)*, Tampere, Finland, 25–27 May 2011. Piscataway, NJ: IEEE.

Muelling K, Venkatraman A, Valois JS, et al. (2015) Autonomy infused teleoperation with application to BCI manipulation. In: *Robotics: science and systems*, Rome, Italy, 13–17 July 2015. Cambridge, MA: RSS Foundation.

Muscettola N, Morris P, and Tsamardinos I (1998) Reformulating temporal plans for efficient execution. In: *Proceedings of the 6th international conference on principles of knowledge*

representation and reasoning (KR & R) (ed AG Cohn, L Schubert, and SC Shapiro), Trento, Italy, 2–5 June 1998, pp.444–452. San Francisco, CA: Morgan Kaufmann.

National Transportation Safety Board (1973) Aircraft accident report: Eastern Airlines 401/l-1011, Miami, FL, 29 December 1972. NTSB/AAR-73-14. Washington, DC.

National Transportation Safety Board (1979) Aircraft accident report: United Airlines, Inc., McDonnell Douglas DC-8-61, n8082u, Portland, OR, 28 December 1978. NTSB/AAR-79-07. Washington, DC.

National Transportation Safety Board (1981) Aircraft separation incidents at Hartsfield Atlanta International Airport, Atlanta, GA. NTSB/SIR-81-6. Washington, DC.

National Transportation Safety Board (1988) Aircraft accident report: Northwest Airlines, Inc. McDonnell-Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne County Airport, 16 August 1987. NTSB/AAR-99-05. Washington, DC.

National Transportation Safety Board (1990) Aircraft accident report: US Air Flight 105, Boeing 737-200, N282AU, Kansas International Airport, MO, 8 September 1989. NTSB/AAR-90-04. Washington, DC.

Niederée U, Jipp M, Teegen U, et al. (2012) Effects of observability, mood states, and workload on human handling errors when monitoring aircraft automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56(1): 1481–1485.

Nikolaidis S and Shah J (2013) 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)*, Tokyo, Japan, 3–6 March 2013, pp.33–40. Piscataway, NJ: IEEE.

Nunes E and Gini M (2015) Multi-robot auctions for allocation of tasks with temporal constraints. In: *29th AAAI conference on artificial intelligence*, Austin, TX, 25–29 January 2015. Palo Alto, CA: AAAI.

Parasuraman R and Riley V (1997) Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39(2): 230–253.

Parasuraman R, Sheridan TB, and Wickens CD (2008) Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making* 2(2): 140–160.

Parker J, Nunes E, Godoy J, et al. (2015) Exploiting spatial locality and heterogeneity of agents for search and rescue teamwork. *Journal of Field Robotics* 33(7): 877–900.

Pierce R and Kuchenbecker K (2012) A data-driven method for determining natural human-robot motion mappings in teleoperation. In: *2012 4th IEEE RAS EMBS international conference on biomedical robotics and biomechatronics (BioRob)*, Rome, Italy, 24–27 June 2012, pp.169–176. Piscataway, NJ: IEEE.

Proctor RW and Zandt TV (2008) *Human Factors in Simple and Complex Systems*. Boca Raton, FL: CRC Press.

Reed KB and Peshkin MA (2008) Physical collaboration of human-human and human-robot teams. *IEEE Transactions on Haptics* 1(2): 108–120.

Riek LD and Robinson P (2011) Challenges and opportunities in building socially intelligent machines. *IEEE Signal Processing Magazine* 28(3): 146–149.

Riley JM, Kaber DB, and Draper JV (2004) Situation awareness and attention allocation measures for quantifying telepresence experiences in teleoperation. *Human Factors and Ergonomics in Manufacturing & Service Industries* 14(1): 51–67.

Sanderson PM (1989) The human planning and scheduling role in advanced manufacturing systems: an emerging human factors domain. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 31(6): 635–666.

Shah SJ, Bliss JP, Chancey ET, et al. (2015) Effects of alarm modality and alarm reliability on workload, trust, and driving performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 59(1): 1535–1539.

Soomer M and Franx G (2008) Scheduling aircraft landing using airlines' preferences. *European Journal of Operational Research* 190: 277–291.

Stanton NA and Young MS (2011) Automatic intelligent cruise control. *Journal of Intelligent Systems* 15: 357–388.

Steinfeld A, Fong T, Kaber D, et al. (2006) Common metrics for human-robot interaction. In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction, HRI '06*, Salt Lake City, UT, 2–3 March 2006, pp.33–40. New York: ACM.

Takayama L and Pantofaru C (2009) Influences on proxemic behaviors in human-robot interaction. In: *IEEE/RSJ international conference on intelligent robots and systems, IROS 2009*, St. Louis, MO, 10–15 October 2009, pp.5495–5502. Piscataway, NJ: IEEE.

Talamadupula K, Benton J, Kambhampati S, et al. (2010) Planning for human-robot teaming in open worlds. *ACM Transactions on Intelligent Systems and Technology (TIST)* 1(2): 14.

Talamadupula K, Briggs G, Chakraborti T, et al. (2014) Coordination in human-robot teams using mental modeling and plan recognition. In: *2014 IEEE/RSJ international conference on intelligent robots and systems (IROS 2014)*, Chicago, IL, 14–18 September 2014, pp.2957–2962. Piscataway, NJ: IEEE.

Tellex S, Knepper RA, Li A, et al. (2014) Asking for help using inverse semantics. In: *Robotics science and systems (RSS) Conference* (ed D Fox, LE Kavraki and H Kurniawati), Berkeley, CA, 12–16 July 2014. Cambridge, MA: RSS Foundation.

Tsang PS and Vidulich MA (2006) *Mental Workload and Situation Awareness*. Hoboken, NJ: John Wiley & Sons, Inc., pp.243–268.

Unhelkar VV, Siu HC, and Shah JA (2014) Comparative performance of human and mobile robotic assistants in collaborative fetch-and-deliver tasks. In: *Proceedings of the ACM/IEEE international conference on human-robot interaction*, Bielefeld, Germany, 3–6 March 2014, pp.82–89. New York: ACM.

Wickens CD (2008) Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50(3): 449–455.

Wilcox RJ, Nikolaidis S, and Shah JA (2012) Optimization of temporal dynamics for adaptive human-robot interaction in assembly manufacturing. In: *Robotics: Science and Systems (RSS)* (ed. N Roy, P Newman, and S Srinivasa), Sydney, Australia, 9–13 July 2012, pp.441–448. Cambridge, MA: MIT Press.

Zhang C and Shah J (2015) On fairness in decision-making under uncertainty: Definitions, computation, and comparison. In: *29th AAAI conference on artificial intelligence*, Austin, TX, 25–29 January 2015. Palo Alto, CA: AAAI.

Zhang H, Law E, Miller R, et al. (2012) Human computation tasks with global constraints. In: *Proceedings of the SIGCHI conference on human factors in computing systems, CHI '12*, Austin, TX, 5–10 May 2012, pp.217–226. New York: ACM