Team Performance and User Satisfaction in Mixed Human-Agent Teams

Sami Abuhaimed
The University of Tulsa
Tulsa, United States
saa8061@utulsa.edu

Sandip Sen
The University of Tulsa
Tulsa, United States
sandip@utulsa.edu

ABSTRACT

Human-agent teams, consisting of at least one human and one agent teaming together to achieve a common objective, are increasingly prevalent and effective in both social and industrial spheres. Associated changes in human preferences and expectations from autonomous teammates will continue to shape and alter collaboration opportunities and dynamics within human-agent teams. New environments are emerging, including Ad Hoc teams where teammates collaborate without pre-coordination or prior knowledge of other teammates capabilities. Team members in ad hoc human-agent teams have to collaborate to find tasks allocations to effectively leverage teammate capabilities to improve team performance and human satisfaction. In this paper, we investigate ad hoc team dynamics under different team compositions, including those comprised of only humans or of human and agent team members, as well as teams consisting of more than two members. Experiments are run with MTurk workers and several hypotheses are evaluated on the effects of teammate type and team size on team performance and human satisfaction using a collaborative Human-Agent Taskboard (CHATboard) platform where teams repeatedly collaborate to complete assigned tasks.

KEYWORDS

Mixed Human-Agent Teams; Team Performance; User Satisfaction

ACM Reference Format:
Sami Abuhaimed and Sandip Sen. 2024. Team Performance and User Satisfaction in Mixed Human-Agent Teams. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 9 pages.

1 INTRODUCTION

The study of peer-level human-agent teams is relatively new and exciting and is increasingly the focus of many research programs. These teams consist of at least one human and one agent teaming together to achieve a common goal. Collaboration between humans and agents is becoming prevalent in various domains, such as smart rooms that include agents that can help business teams make decisions [12], medical teams that include agents such as discharge nurses [52], agents that motivate people in therapeutic activities [6], guide emergency evacuations [43] and assist in disaster relief [40, 46].

In addition, there are several forces that continually change the nature and functioning of these teams, including agent capabilities as well as ease of use and access. Agent capabilities and performance have been improving in areas such as optimization methods, efficient algorithms, accessible interface and are generally becoming more competent and reliable. With changing agent capabilities and human preferences, the dynamic of collaboration within human-agent teams will continue to evolve. This evolution will engender different team designs depending on the context. Humans and agents will coordinate to focus their efforts on complementary set of tasks to maximally leverage their respective strengths, expertise, and knowledge.

It is likely that this changing dynamic will create new applications and team roles, in addition to existing ones, that need to be shared by agents and humans. For example, in a disaster response scenario, one human team member may collaborate with an agent task allocator to allocate different tasks related to search and rescue operations, while other humans and agents (snake robots, robot dogs, drones, etc.) to quickly locate survivors and carry out safe retrieval, transportation, and medical treatment procedures while reducing the risk to the human actors.

We can envision these teams consisting of humans and agents collaborating to enrich human lives, support human self-efficacy, increase satisfaction and well-being, improve decision making, and increase team performance and productivity. For example, when human-agent team collaborate in search and rescue scenario, the agent can augment human teammate by doing new tasks, such as searching for survivors by "seeing" things humans cannot see, e.g. thermal imaging, or from different vantage points (aerial, inaccessible, or physically challenging spaces, etc.). As these teams become more prevalent in our societies, researchers have begun studying the interactions and dynamics within these teams to understand their functioning and improve their design [17, 23, 38].

Due to the increased connectivity, changing nature of work, and increased levels of crises, e.g., pandemics and natural disasters, ad hoc environments are becoming prevalent, and hybrid or heterogeneous teams of humans and autonomous agent are increasing in these settings. In an *ad hoc teamwork*, teammates, whether humans or agents, collaborate without pre-coordination: *"An ad hoc team setting is one in which teammates must work together to obtain a common goal, but without any prior agreement regarding how to work together"* [14]. Collaboration in ad hoc teams is more challenging due to the absence of prior knowledge and established relationships.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand.

It is important for these teams to be flexible, agile and to quickly adapt and learn about the team, to reliably achieve team goals.

Task allocation and the process of distributing tasks to team members have been studied separately in both human and agent teams. The challenge of task allocation within human-agent teams is leveraging complementary capabilities of team members to find allocations to improve team effectiveness, measured in terms of team performance and satisfaction of human team members. We consider scenarios in which the expertise distribution of the agent over the task types is fixed, known, and is simulated. It is necessary to first estimate the expertise levels of the current human team member(s) in a given human-agent team and accordingly adapt task allocations to team members to optimize team performance.

The allocation problem is exacerbated as a team member does not know the levels expertise of its partner *a priori* in ad hoc teams. Although we allow human and agent partners to share their estimated expertise over different task types, the accuracy and consistency of these estimates expressed by humans are often unreliable [28].

It is unclear how the presence of agent teammates will change the motivation, effort, and efficiency of human teammates in ad hoc human-agent teams compared to human-human teams. The two critical questions on the efficacy of human-agent ad hoc teams that we study in this paper are the following: *(a) Are human satisfaction and team performance depend on whether their teammates are humans or agents? (b) How do these dynamics change if the team consists of more than two team members?* We present results and analysis from experiments conducted to understand how the type of team member, agent or human, and team size can influence the effectiveness of ad hoc human-agent teams.

To facilitate experimentation with ad hoc human-agent teams, we use a collaboration framework for task allocation and performance analysis: the Collaborative Human-Agent Taskboard (CHATboard) [3]. We use it for repeated human-agent team task completion scenarios, where human workers were recruited from the Amazon Mechanical Turk (Mturk) platform. We ran experiments involving repeated collaboration using Agent and Human Allocator Protocols, which differ based on the identity of the allocator. Within each protocol, we compare teams with *deceptive agents* (pretending to be human) and *non-deceptive agents* (do not use deception) to allow comparative experiments with human-agent and human-"human" (where the "human" is an agent pretending to be human). We also compare the protocols with more than two team members. We present experimental results and analysis, highlighting team performance and human satisfaction, with different aspects of team work. Finally, we identify future research directions.

2 RELATED WORK

Human-agent teams have been studied in different domains such as space robotics [17], therapy [1], programming [42] and decision-making [5]. Human-agent studies typically examine human perceptions of team and agent teammates, which include trust [23], satisfaction [38] and humanness [18].

There has been recent interest in negotiation protocols in which humans and agents interact over different rounds [38]. The objective of the agent and the human is to maximize their own utilities.

Existing research has focused on agents who play supportive roles for human teammates [11, 27, 31], and mainly in robotic and simulation settings [10, 44]. We focus on ad hoc environments, while other studies [17, 38] incorporate training or interaction sessions with the agent and environment prior to experimentation. We are also interested in autonomous agents; DeChurch and Larson view an autonomous agent as a "team member fulfilling a distinct role in the team and making a unique contribution" [32]. Some studies in the space robotics [17] and negotiation [38] domains incorporate self-interested autonomous agents maximizing local utility. We focus on collaborative human-agent teamwork in which teammates make their decisions based on what is best for the team.

Task allocation has been extensively studied in multi-agent teams [15, 22, 29, 35, 36]. The focus is on designing efficient mechanisms to distribute tasks within their systems. An approach studies ways to better allocate the task allocator role to the team. Role allocation has been studied in the context of agent systems [9], and a role can generally be understood as an "abstract description of an agent's responsibilities and expected functions [50]." Roles also allow agents to specialize in different responsibilities and behaviors. Multi-agent role allocation has been studied in domains such as RoboCup Soccer [16] and pursuit games [2]. The other approach is to study ways to allocate tasks to agent team members. A task describes a specific unit of work undertaken by an agent [9]. Allocating tasks to agents is a major research area in the multi-agent systems literature, and it is studied in domains such as surveillance [7] and Search and Rescue [51]. Moreover, there is a lack of investigating environments that include human teammates. The presence of humans in agent teams may require new approaches, as we do not know whether the same mechanisms would produce required performance.

Task allocation is studied in the literature on human teams and organizations. The mechanism of task allocation, which includes capability identification, role specification, and task planning, is considered an important component of teamwork [13, 33, 34]. For To achieve its goals, any organization needs to solve four universal problems, including task allocation [39]. In human teams, the focus is on understanding the characteristics of human teams to design the best possible task allocation mechanism. However, there is little investigation of the effects of autonomous agents on human teams when they are included in the team's task allocation mechanism.

Thus, the study of task allocation with combined human and agent team members is promising. Few existing work in this area examines different directions [40, 44]. For example, [20] investigates an agent that helps humans control robots in simulations and experiments, with a focus on supporting operators.

In addition to performance, human satisfaction with the team is recognized as a major component of team effectiveness in teamwork [19, 49]. Team member satisfaction has been found to have positive effects on different team dimensions [26]. [48] studied the influence of different factors, such as team familiarity, on member satisfaction in virtual human teams.

Furthermore, most recognized human-team frameworks, such as the Input-Process-Output (IPO) model, examine teamwork through different dimensions, which include members, the process through which they interact, and teamwork outcome [33]. Other researchers in different domains have measured the satisfaction of team members with a similar perspective, that is, viewing human satisfaction with more than one dimension [21, 41]. One may be more satisfied with the team members than with the interaction protocol or the outcome of teamwork. Most studies of human-agent teams overlook or do not fully recognize these distinctions when measuring human perceptions. Furthermore, to our knowledge, there has been little investigation of team member satisfaction in the human-agent team literature. [38] studies human satisfaction, among other factors, with only teamwork outcome in a negotiation domain. Others, such as [45], investigate the role of agent failures in user satisfaction with the agent only.

3 RESEARCH HYPOTHESES

We now motivate and present the research hypotheses related to the influence on some key team characteristics of the effectiveness of ad hoc human-agent teams that we later experimentally evaluate. In this study, human and agent team members engage in teamwork where the goal is to complete the tasks assigned to the team.

Task allocation protocols divide team tasks between team members. Human and Agent Allocator Protocols differ in who allocates tasks to team members: the human in the former and the agent in the latter case. Previous work has observed that agent allocators produce better team performance compared to human allocators [3], and human satisfaction tends to be indifferent or low when agents allocate tasks [4]. This work considered teams with one agent and one human, without evaluating how these human-agent teams can differ from human-human teams.

Agent allocators have several advantages over human ones in effectively distributing team tasks: (a) lack of preference for task types that are not performance motivated (e.g., humans may enjoy doing tasks even though they may not be proficient in it), (b) agents will have better estimates of their own capabilities, (c) agents can consistently follow optimal allocation procedures, and (d) agents can more reliably learn from task performance of teammates in early episodes and adapt task allocations to improve performance.

Consistent with this previous work with teams of one human and one agent, we expect agent allocators to produce higher team performance than their human counterparts when the team consists of more than two members [3]. Also, because we expect agent allocators to produce a high team performance, the human teammate will perceive the team outcome as satisfactory [4]. The first four hypotheses presented below, **H1** to **H4**, refer to Human-agent teams with more than two members.

H1*: The performance of Human-agent teams is higher when using the Agent rather than the Human Allocator Protocol.*

H2*: Human teammates' satisfaction with Team Outcome will be higher with agent allocators rather than with human ones.*

When the agent is the task allocator, we expect a higher team performance. However, the human may feel less in control of the task allocation process and, hence, may be less satisfied in the protocol where they are not assigned the allocator role.

H3*: The satisfaction of human teammates with the Protocol will be higher when they, rather than the agent, allocate tasks.*

Although human team members prefer being the allocator they might still view agent team members favorably [4].

H4*: Human satisfaction with their agent teammates will be similar regardless of who is allocating tasks, agents, or humans.*

Human-agent teams formed in ad hoc environments face issues that could influence their effectiveness of collaboration. Agents within these teams have different characteristics, such as capabilities, that human teammates might not be aware of and that might affect their work quality, especially when they are assigned to pivotal team roles, such as that of allocating tasks to team members. It is important to understand if human team members react differently when agents, rather than humans, assume such pivotal team roles. We expect humans to feel more responsive and motivated to perform well if another human, rather than an agent, is allocating them tasks. In our experiments, we use deception, which was approved by our university's IRB: the user is told that the allocator is another human even when the allocator is actually an agent.

H5a*: Deceptive allocator agents, pretending to be humans, will produce higher team performance than non-deceptive agents.*

Similarly, we expect humans to be more attentive and responsive to the strengths and weaknesses of human (agent pretending to be human) team members, enhancing the ability to allocate tasks to the team, which then produces better team performance.

H5b*: Human (deceptive) allocators will produce better team performance when assigning tasks to deceptive agents rather than non-deceptive ones.*

Because we expect humans to perform better when the allocator is a deceptive agent, we also expect humans to be more satisfied with team Outcome than when partnering with non-deceptive agents. Similarly, we expect human allocators to perform better, and thus human satisfaction with Outcome to be higher as well.

H6a*: Humans will be more satisfied with Outcome when deceptive agents, rather than non-deceptive ones, allocate tasks.*

H6b*: Satisfaction with team Outcome will be higher when humans allocate tasks to deceptive agents rather than non-deceptive ones.*

Humans can be more comfortable with teammate characteristics that are familiar, i.e., they can recognize their capabilities quickly, rather than when working with unfamiliar agent teammates. Thus, we expect them to be more satisfied with the deceptive agent, who they believe is another human, than the non-deceptive one.

H7*: Humans will be more satisfied with the deceptive Teammate agent rather than the non-deceptive one.*

Lastly, previous work has shown that when agents are assigned the task allocator role in ad hoc human-agent teams, human satisfaction with the Agent Allocator Protocol tends to be lower with the Protocol than when humans are assigned to that role [4]. Humans like to have control over or provide key input to the allocation process. We expect to also observe that humans are more satisfied with Protocol when they are allocating compared to when an agent or another human allocates.

H8*: Human allocators will be more satisfied with Protocol when they are allocators than when another teammate, whether human (deceptive agent), or non-deceptive one, is assigned to allocator role.*

4 METHODOLOGY

We now present the experimental testbed, interaction protocols, agent behavior, evaluation metrics, and the experimental design adopted in our study.

4.1 Collaborative Human-Agent Taskboard (CHATboard)

We use CHATboard, an environment that facilitates human-agent and human-human team collaboration [3]. CHATboard contains a graphical interface that supports human-agent team coordination to complete a set of tasks (see Figure 1). CHATboard can display the task sets to be completed by the team. It also supports multiple task allocation protocols, communication between team members for expressing confidence levels, displaying task allocations and performance by team members on assigned tasks, etc.

The framework utilizes the concept blackboards on which tasks are posted to facilitate a human team member perceiving an agent as a distinct team member. Blackboards have been used effectively in agent teams as a common repository for information sharing between agents [25]. We incorporate three task boards into the task sharing frame: one shared board includes the set of team tasks organized by task types, and two other boards for tasks assigned to human and agent team members, respectively. These task boards facilitate coordination and are easily navigable repositories of team information, allowing team members to share and view information on the status of pending and completed tasks.

We assume a set of n team members N, $\{p_1, p_2,...,p_n\}$, a set of m task types M: $\{y_1, y_2,...,y_m\}$, a set of r tasks, T_{jr}:$\{t_{j1}, t_{j2},...,t_{jr}\}$, for each task type y_j. Team member i can share their confidence levels $p_i(y_j)$ over task types y_j, which represents their confidence or probability of successfully completing a task of that type. - Yes The set C_i:$\{p_i(y_1),p_i(y_2),....,p_i(y_m)\}$ represent expertise profile, i.e., the confidence levels for different task types, for team player, p_i. The team members interact over E episodes, where the episode numbers range from $1 \dots E$. $A_{i,e}$ denotes the set of tasks allocated to player i in episode e and we assume that all available tasks are exhaustively allocated, i.e., $\bigcup_i A_{i,e} = \bigcup_j T_{jr}$. The performance of the player p_i for a task t_{jk} in episode e is referred to as $o_{ijke} \in \{0, 1\}$, i.e., a team member either succeeds or fails in completing an assigned task. We define the performance of p_i on task type y_j in episode e as $\mu_{i,y_j,e} = \sum_{t_{jk} \in A_{i,e}} o_{ijke}$.

4.2 Interaction Protocols

We describe the protocols that manage ad hoc human-agent teamwork. In order for ad hoc human-agent teams to be effective, flexible protocols are needed to support the process of task allocation. An important protocol dimension is the allocator role: How is it assigned and *who* allocates the tasks? Two interaction protocols are designed to guide the task allocation process in an ad hoc environment: (i) Human Allocator Protocol and (ii) Agent Allocator Protocol. The latter assigns the task allocation role to agent teammate (see Figure 2), while the former, described below, assigns the task allocator role to the human teammate:

(1) *The protocol assigns task allocator role to human.*
(2) *The protocol asks agent teammate for its confidence levels over task types.*
(3) *The protocol passes the agent's confidence levels to the human.*
The following steps comprise an episode and are repeated N *times*
$e \leftarrow 1$
Episode starts:
(4) *The protocol asks human to provide task allocations for the team.*
(5) *Allocated tasks are communicated to the team members who then attempt to complete assigned tasks.*
(6) *The protocol receives human and agent task performance measures.*
(7) *The protocol computes statistics and displays overall team performance as well as individual team member performances for the episode on their respective task boards.*
Episode ends
$e \leftarrow e + 1$; *if* $(e < N)$, *Go to step 4.*

The two protocols differ on who allocates the tasks but are similar on other dimensions. It is natural for human teams to share that informaion. Hence, both protocols ask teammates to share how confident they are about completing tasks of different types with the allocator teammate. Although these protocols provide a framework for team interaction and task allocation, they do not determine the allocation strategy used by the allocator.

4.3 Agent Characteristics

4.3.1 Expertise. An agent expertise profile is represented as a vector of successful completion probabilities for task types Agent expertise is simulated by flipping a coin with success probability equal to the expertise level of the agent for that task type. An agent is cognizant of its expertise levels and communicates it, when required, as its confidence levels to its teammates. Collective expertise levels of team members reflect team capacity.

4.3.2 Agent Allocator Strategy. The primary goal of task allocation is to maximize the utilization of the available team capacity given the expertise of the team members. As the expertise levels are specified for task types, the agent allocator allocates by *task types* instead of considering individual *task items* for task allocation. The agent uses task completion success rates to update the performance estimates for human team members. These adapted performance estimates are used in the task allocation procedure. For one human, one agent teams, the agent objective is:

$$\max \sum_{y \in M} (x_y a(y) + (1 - x_y)h(y)); s.t. \forall y, x_y \in 0, 1, \tag{1}$$

where x_y are binary variables indicating whether a task type, y, is assigned to the human or agent, based on the current performance estimate of the human, $h(y)$, and agent, $a(y)$, on that task type.

This is an *unbalanced assignment problem* since the number of task types is greater than the number of team members ($m > n$). It can be solved by transforming it into a *balanced* formulation, e.g., adding dummy variables, and using the Hungarian algorithm [30]. We utilize the SCIP mixed-integer programming solver [37], represented by the `getAllocations()` procedure in Line 6 of Algorithm 1, to find the allocation that maximizes utilization of team performance estimates.

In many task allocation formulations, e.g., matching markets, assignment problems, etc., it is assumed that participants' preferences or confidence levels are accurately known [47]. However, in our formulation, learning is needed as we expect, and find in our experiments, that human participant estimates of their capabilities to be inaccurate.

Since we study ad hoc environments, the second goal of our agent is to quickly learn about its partner's expertise levels and adapt the task allocations accordingly for improved team performance. In the first episode, the agent strategy incorporates exploration of team members' capabilities by partitioning task items within each task

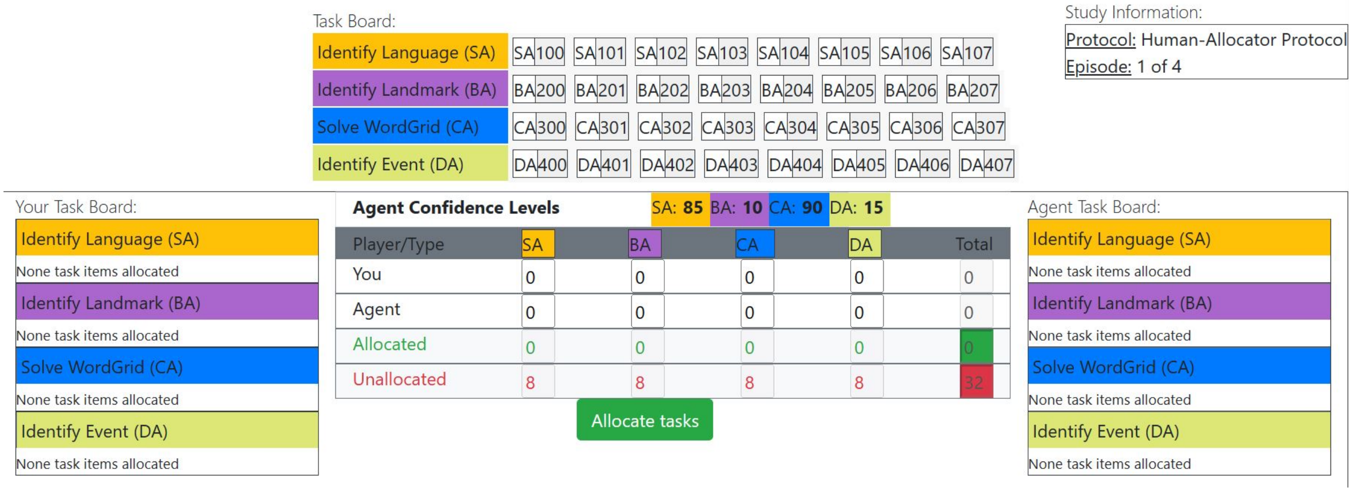

Figure 1: Allocation phase of Human Allocation Protocol

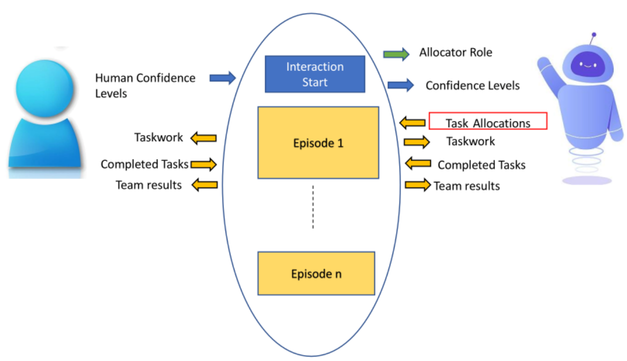

Figure 2: Agent Allocator Protocol Steps.

Algorithm 1 Agent Allocator Strategy

Input: $N = \{p_h, p_g\}, M = \{y_1, \ldots, y_m\}, E$

1: **for** $e = 1....E$ **do**
2: **if** $e = 1$ **then**
3: $Q_{i,y_j} \leftarrow p_i(y_j), \forall p_i \in N, y_j \in M$
4: each T_{y_j} is partitioned into n equal size subsets, which are randomly allocated to agent i to form $A_{i,1}$, for each $p_i \in N$
5: **else**
6: $A_{i,e} \leftarrow$ `getAllocations`$(Q_{i,e})$
7: **end if**
8: **if** y_j is allocated to p_i **then**
9: $Q_{i,y_j} \leftarrow (1 - \alpha) \cdot Q_{i,y_j} + \alpha \cdot \mu_{i,y_j,e}$
10: **end if**
11: **end for**

type, T_{y_j}, as shown in Line 4 in Algorithm 1, and then randomly allocates them. After the first episode, the agent updates the human expertise estimates based on their stated confidence levels over task types, provided at the start of interaction, along with actual teammate performance, for different task types in the first episode.

After each interaction, e, the agent updates the expertise model, Q_{i,y_j}, of the team member, p_i, for each task type, y_j as follows:

$$Q_{i,y_j} \leftarrow (1 - \alpha) \cdot Q_{i,y_j} + \alpha \cdot \mu_{i,y_j,e},$$

where $\mu_{i,y_j,e}$ is the observed performance and $0 < \alpha < 1$ is a learning parameter.

4.4 Evaluation Metrics

4.4.1 Team Performance. A team member either successfully completes or fails on an allocated task. Team overall performance is measured as the percentage of successful completion of assigned tasks over all episodes: $\frac{1}{E}\sum_{e=1}^{E} R_{team,e}$, where $R_{team,e}$ is the team performance in episode e, which is the average performance, μ, of all team members over all task types in that episode

$$R_{team,e} \leftarrow \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m} \mu_{i,y_j,e}.$$

Allocator	
A) Agent	B) Human
A.1 Human-agent A.2 Human-human (deception)	B.1 Human-agent B.2 Human-human (deception)

Table 1: A and B are two conditions for second set of experiments (*Teammate Type*). Within each condition, two experiments are conducted (see Section 4.5 for details).

4.4.2 Human Satisfaction. We measure satisfaction with three team dimensions: interaction protocol, teamwork outcome, and agent teammate. We adapt the satisfaction survey proposed by [21] and [41] with four questions for each dimension. The survey follows a five-point Likert scale setting administered at the end of the study. We present sample questions for each survey dimension:
Task Allocation Protocol: *"I feel satisfied with the processes used in the team's task allocation protocol."*
Agent teammate: *"I am satisfied with my agent teammate."*
Team Outcome: *"I feel satisfied with the things we achieved in the team's task allocation protocol."*

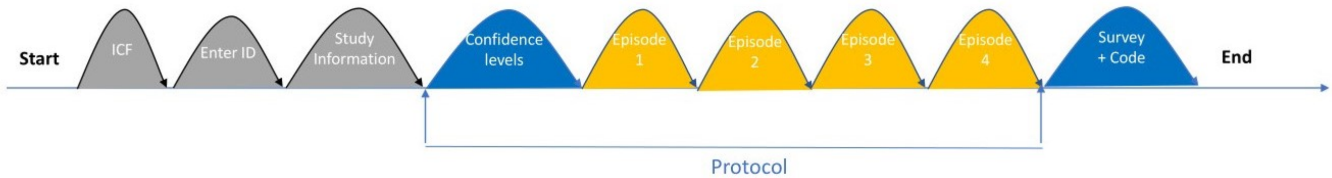

Figure 3: Timeline of the study.

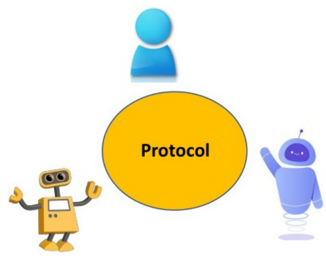

Figure 4: ***Team Size Effects:*** **One human and two agents.**

Evaluation Criteria \ Allocator	Human Mean	Human SD	Agent Mean	Agent SD
Protocol Satisfaction	4.14	0.64	4.27	0.60
Outcome Satisfaction	3.9	0.58	**4.22***	0.75
Agent Satisfaction	4.01	0.72	3.96	0.62
Team Performance	0.71	0.16	**0.82***	0.05

Table 2: Satisfaction and Team Performance for Protocols (Team Size Effects); $^*p < 0.05$**.**

4.5 Experiment Configuration

Two sets of experiments are conducted: (i) teams with more than two members and (ii) teams with deceptive agents.

In the first set, we performed experiments to assess the influence of *team size* with teams of one human and two agents ($n = 3$), $N = \{p_{a_1}, p_{a_2}, p_h\}$ (see Figure 4). We compare two conditions: A) agent allocator, B) human allocator. We used a between-subject study and recruited 130 participants from Amazon Turk with 65 subjects for each of the two conditions (Human and Agent Allocator Protocols), as recommended for a medium-sized effect [8].

For the second set, we performed experiments to evaluate the influence of the *teammate type* with teams of one human and one agent ($n = 2$), $N = \{p_a, p_h\}$. As Table 1 shows, we experiment with the following two conditions (A and B): (A) For agent allocators, we compare 1) human-agent with 2) human-human (deception) teams. (B) For human allocators, we compare 1) human-agent with 2) human-human (deception). Both the A and B experiments use between-subjects experiment design. To evaluate our hypotheses, we compare A.1 with A.2, B.1 with B.2, and B.1/B.2 with A.1/A.2. Thus, in the Human Allocator Protocol, human allocators are paired with both deceptive agents as well as non-deceptive ones. Similarly, for the Agent Allocator Protocol, humans collaborate with both deceptive agent allocators as well as non-deceptive ones. We recruited 260 participants from Amazon Turk with 130 subjects for each of the two conditions.

For the two sets of experiments, we use four task types ($m = 4$), M: $\{y_1, y_2, y_4, y_4\}$, which are the following:

Identify Language: A piece of text in a script is presented and the user is asked to identify the language of that text from a given set of options.

Solve WordGrid: A letter grid is presented and the user is asked if a given word is present in any horizontal, vertical, or diagonal order in that grid.

Identify Landmark: A picture of a prominent landmark is presented and the user is asked to select the country where the landmark was located from a given set of options.

Identify Event: A picture of a famous event is presented and the user is asked to select the date range when that event occurred from a given set of options.

Instances of these task types are shown in Figure 5. The task types are selected so that for each type, sufficient expertise variations are likely in recruited human subjects. For example, *Identify Language* is a task type in which the team is asked to identify the language, e.g., Japanese, in a text message from a number of options such as Japanese, German, Hindi, Arabic.

We conducted experiments with ad hoc teams that interact over four episodes ($E = 4$), and 32 ($r = 8$) task instances are assigned to the team in each episode. Confidence levels for tasks are stated in a [1,100] range, which are scaled by agents to a [0,1] range and interpreted as task completion probabilities.

Participants start the first episode after agreeing to the Informed Consent Form and reading the study description. Each episode contains three phases: task allocation, task completion, and task results presentation. Both overall and per-task type performance is shared with the human and agent teammates after each episode. After four episodes have been completed, participants are asked to complete a survey that includes satisfaction with various aspects of teamwork. We incorporate random comprehension attention checks to ensure the fidelity of the result [24]. Participants may receive a bonus payment based on team performance (See Figure 3 for an overview of how the study progresses).

5 RESULTS

5.1 Team Size Effects

We ran experiments to see if previously observed relative performance and satisfaction levels in teams with human and agent allocators were replicated when team size was increased to $n = 3$. We then compare the two protocols with respect to performance and

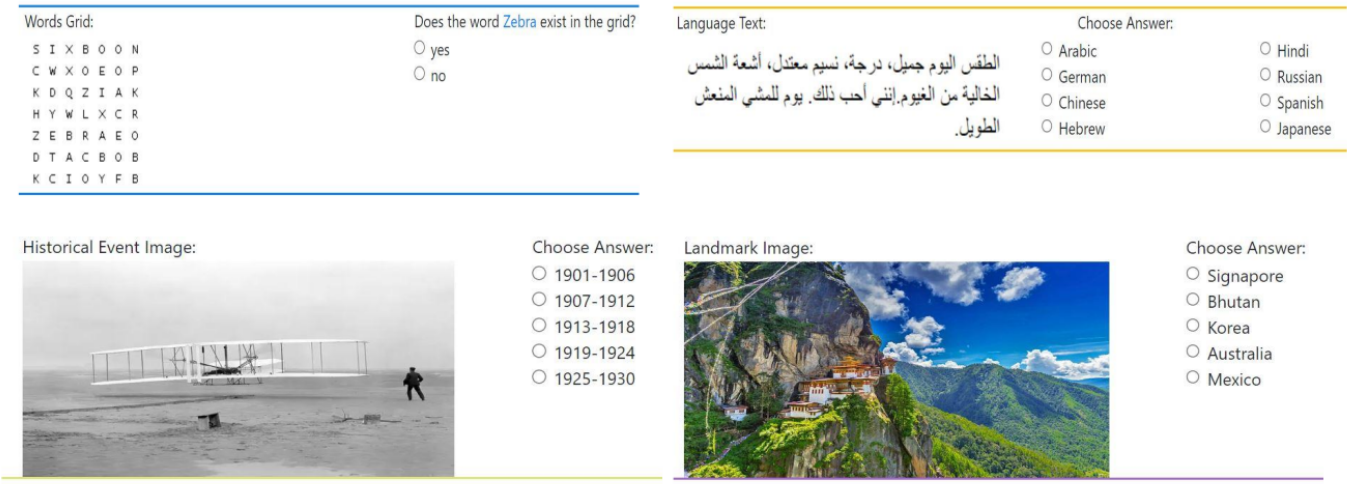

Figure 5: Instances of Solve wordgrid (top left), Identify Language (top right), Identify Landmark (bottom right), and Identify Event (bottom left) task types.

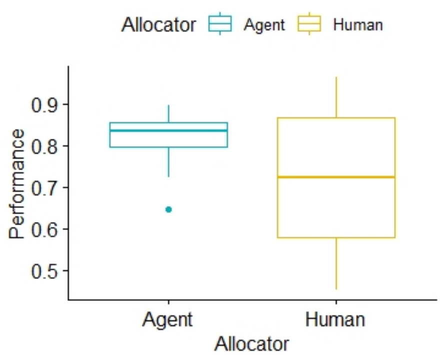

Figure 6: Team Performance for the Protocols (Team Size Effects).

satisfaction with Protocol, team Outcome, and agent Teammate. See Table 2 for a summary of the results.

5.1.1 Team Performance: We compare the performance of the two allocator protocols. The teams using Agent Allocator Protocol (M = 0.82, SD = 0.05) compared to ones using Human Allocator Protocol (M = 0.71, SD = 0.16) demonstrated significantly higher team performance, $t = 5.1, p < 0.01$. We plot the team performance distribution for the two protocols in Figure 6.

5.1.2 Satisfaction: We compare human satisfaction with the **Allocation Protocol** in the Human ($M = 4.14, SD = 0.64$) and Agent ($M = 4.27, SD = 0.60$) Protocols, and found no significant difference: $t = 1.18, p > 0.05$.

We compare human satisfaction with **Team Outcome** when humans ($M = 3.9, SD = 0.58$) and agents ($M = 4.22, SD = 0.75$) allocate. We found a significant difference in human satisfaction with Team outcome: $t = 2.71, p < 0.01$. Participants are more satisfied with the team Outcome when they are in the Agent Allocator Protocol rather than the Human one.

	Satisfaction			
	Protocol	Outcome	Teammate	Perf
Human-Human (deception)	3.60	3.71	3.47	0.69
Human-Agent (no deception)	3.54	3.59	3.45	0.70

Table 3: Satisfaction and Team Performance for *Agent Allocator (Teammate Type Effects).*

We compare human satisfaction with the **Agent** in the Human ($M = 4.01, SD = 0.72$) and Agent ($M = 3.96, SD = 0.62$) Allocator Protocols, and found no significant difference: $t = 0.41, p > 0.05$.

5.2 Teammate Type Effects

5.2.1 Team Performance: We analyze team performance for two Human-human (deception) conditions with Human-agent (non-deception) ones. We compare the performance of the team when the agent is assigned an allocator role (Condition A in Table 1): team performance for deceptive agent allocator ($M = 0.69, SD = 0.08$) is similar for the non-deceptive one ($M = 0.70, SD = 0.08$), and the difference is not statistically significant, $t = 0.59, p > 0.05$ (See Table 3). Similarly, we compare team performance when a human is assigned to the allocator role (Condition B in Table 1): performance of the team when assigned to deceptive agent ($M = 0.66, SD = 0.10$) is similar to ones assigned non-deceptive one($M = 0.67, SD = 0.09$), and the difference is not statistically significant, $t = 0.29, p > 0.05$ (See Table 4).

5.2.2 Satisfaction: We compare human satisfaction when an agent is assigned to allocator role (Condition A in Table 1) in deception and no-deception conditions. We find that human satisfaction with **Outcome** is little higher for deceptive agent allocators ($M = 3.71, SD = 0.86$) than non-deceptive ones (M=$3.59, SD = 82$), but the difference is not statistically significant, $t = 0.75, p > 0.05$. We find that the deceptive agent allocators have a similar satisfaction with **Teammate** ($M = 3.47, SD = 1.1$) than non-deceptive ones ($M = 3.45, SD = 0.91$), and the difference is not statistically significant, $t = 0.87, p > 0.05$ (See Table 3)

	Satisfaction			
	Protocol	Outcome	Teammate	Perf
Human-Human(deception)	4.03	3.77	3.63	0.66
Human-Agent(no deception)	3.87	3.73	3.83	0.67

Table 4: Satisfaction and Team Performance for *Human Allocator (Teammate Type Effects).*

We compare human satisfaction when a team is assigned to human allocators (Condition B in Table 1) in deception and no-deception conditions: satisfaction with **Outcome** is similar for deceptive agents ($M = 3.77, SD = 0.8$) and non-deceptive ones ($M = 3.73, SD = 0.8$), and the difference is not statistically significant, $t = 0.3, p > 0.05$. Lastly, we find that human satisfaction with **Teammate** is lower for deceptive agents ($M = 3.63, SD = 0.98$) and non-deceptive ($M = 3.83, SD = 0.80$) condition, and the difference is not statistically significant, $t = 1.2, p > 0.05$ (See Table 4).

We analyze satisfaction with the **Protocol** between agent and human (Conditions A and B in Table 1) allocators. We compare the satisfaction between the two deception conditions. We find that human satisfaction with the Protocol is higher when they are allocating ($M = 4.03, SD = 0.70$) rather than when a deceptive agent, i.e., human, is allocating ($M = 3.60, SD = 0.99$), and the difference is statistically significant $t = 2.8, p < 0.05$. Similarly, we find that the satisfaction of the Protocol is higher when humans are allocating ($M = 3.87, SD = 0.79$) rather than when the non-deceptive agent is allocation ($M = 3.54, SD = 1.1$), and the difference is statistically significant $t = 1.9, p < 0.05$. Lastly, the difference in human satisfaction with the Protocol between two deception conditions is higher than no-deception condition ($Diff_{Dec} = 0.43, Diff_{no} = 0.33$).

6 DISCUSSION

We use CHATboard, a flexible task allocation framework between human and agent team members, for ad hoc scenarios. We showed its efficacy in supporting collaboration between multiple humans and agents, as well as the ability to configure deceptive agents who pretend to be humans. We utilize the Agent and Human Allocator Protocols to guide team interactions.

Prior work has found that agent task allocators produce higher team performance for teams consisting of one human and one agent, and human satisfaction tends to be indifferent with Agent and low with Protocol when they are not allocators [3, 4]. In this work, agent allocators still produce higher team performance for teams consisting of more than two team members. Human satisfaction with teamwork Outcome is also higher for teams assigned to agent allocators as they produce higher team performance There was also an indifference in satisfaction with the teammate.

Although team performance was higher, there was no difference in human satisfaction with the Protocol, as we observed in previous work [3, 4]. This may suggest that when users have more than one teammate, they tend to become indifferent to task allocation role. An interesting future work would be to design different protocols, e.g., agent-guided, that would increase human satisfaction and preserve high team performance, and evaluate them with different team sizes. *In general, it is reassuring to find that most of the trends obtained in binary teams with 1-1 human-agent interactions are replicated in larger teams with unequal human-agent team compositions.*

We proposed that the effectiveness of human-agent teams might be influenced by whether the human's teammate is human or agent. In particular, humans may perceive other human teammates favorably compared to their agent counterparts. We design experiments to evaluate deceptive agents within the Agent and Human Allocator protocols. In the former, we compare teams assigned to deceptive agent allocators, i.e., pretending to be human, with teams that have non-deceptive agents, i.e., agents who identify themselves truthfully. In the latter, we compare teams with human allocators but the teammates vary between deceptive and non-deceptive agents.

Most of our hypotheses, involving differential performance and satisfaction of humans collaboarting with human vs. agent teammates, were not supported. In **H5a** and **H6a**, we conjectured that humans will have higher team performance and Outcome satisfaction with deceptive agent allocators than non-deceptive ones, but we found no difference. We observe the same result for the hypotheses **H5b** and **H6b**, with human allocators: There was no difference between deceptive and non-deceptive agents. Similarly, we state in **H7** that humans will have greater satisfaction with teammates who are human (deceptive agent) than actual agents, but we did not find any difference. We did expect the last result, as mentioned in **H8**, that humans will be more satisfied with the Protocol when they allocate, regardless of whether the teammate is a deceptive agent or an actual agent. In summary, when humans are allocators, there is no difference in team performance and human satisfaction when their teammate is agent or human (deceptive agent). Likewise, when the team is assigned to agent task allocators, human teammates satisfaction and team performance are similar irrespective of whether the agent task allocator is deceptive, pretending to be human, or non-deceptive agent.

What makes the aforementioned results unexpected is that, contrary to our hypotheses, humans view agents on a par with human teammates. *This may have an interesting effect on current and future human agent teams, as these results suggest that users are increasingly comfortable and willing to view agent teammates as peers!*

The observation that humans are satisfied with the Protocol in which they are given more input or control of the task allocation process, that is, assigned to the allocator role, is not new, as shown in previous work, but what is interesting is that the satisfaction of the protocol is higher when the teammate is human (deceptive agent) rather than agent; in other words, humans prefer the allocator role *more* when the other teammate is human rather than agent.

The observations from this paper can inform the design of ad hoc human-agent teams. It is worth considering agents in task allocation roles since agents outperform their humans counterparts. Furthermore, the fact that humans are indifferent to whether a teammate is a human or an agent with respect to team performance and satisfaction is informative. It motivates further work to understand what factors human teammates prioritize other than teammate type. We also plan to investigate the dynamics of ad hoc teams with different allocation protocols, such as when an agent provides guidance to human allocators.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to the Saudi Arabian Ministry of Education and the University of Tulsa for their support.

REFERENCES

[1] Amal Abdulrahman, Deborah Richards, and Ayse Aysin Bilgin. 2021. Reason Explanation for Encouraging Behaviour Change Intention. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems.* 68–77.

[2] Myriam Abramson, William Chao, and Ranjeev Mittu. 2005. Design and evaluation of distributed role allocation algorithms in open environments. In *International Conference on Artificial Intelligence.*

[3] Sami Abuhaimed and Sandip Sen. 2022. Effective Task Allocation in Ad Hoc Human-Agent Teams. In *HHAI2022: Augmenting Human Intellect.* IOS Press, 171–183.

[4] Sami Abuhaimed and Sandip Sen. 2023. Human Satisfaction in Ad Hoc Human-Agent Teams. In *International Conference on Human-Computer Interaction.* Springer, 207–219.

[5] Ashton Anderson, Jon Kleinberg, and Sendhil Mullainathan. 2017. Assessing human error against a benchmark of perfection. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 11, 4 (2017), 1–25.

[6] Timothy W Bickmore, Rukmal Fernando, Lazlo Ring, and Daniel Schulman. 2010. Empathic touch by relational agents. *IEEE Transactions on Affective Computing* 1, 1 (2010), 60–71.

[7] Giulio Binetti, David Naso, and Biagio Turchiano. 2013. Decentralized task allocation for surveillance systems with critical tasks. *Robotics and Autonomous Systems* 61, 12 (2013), 1653–1664.

[8] Willem-Paul Brinkman. 2009. Design of a questionnaire instrument. In *Handbook of mobile technology research methods.* Nova Publishers, 31–57.

[9] Adam Campbell and Annie S Wu. 2011. Multi-agent role allocation: issues, approaches, and multiple perspectives. *Autonomous agents and multi-agent systems* 22 (2011), 317–355.

[10] Jessie YC Chen and Michael J Barnes. 2014. Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems* 44, 1 (2014), 13–29.

[11] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A Smith. 2018. Creative writing with a machine in the loop: Case studies on slogans and stories. In *23rd International Conference on Intelligent User Interfaces.* 329–340.

[12] Robert G Farrell, Jonathan Lenchner, Jeffrey O Kephjart, Alan M Webb, MIchael J Muller, Thomas D Erikson, David O Melville, Rachel KE Bellamy, Daniel M Gruen, Jonathan H Connell, et al. 2016. Symbiotic cognitive computing. *AI Magazine* 37, 3 (2016), 81–93.

[13] David M Fisher. 2014. Distinguishing between taskwork and teamwork planning in teams: Relations with coordination and interpersonal processes. *Journal of applied Psychology* 99, 3 (2014), 423.

[14] Katie Genter, Noa Agmon, and Peter Stone. 2011. Role-Based Ad Hoc Teamwork. In *Proceedings of the Plan, Activity, and Intent Recognition Workshop at the Twenty-Fifth Conference on Artificial Intelligence (PAIR-11).*

[15] Brian P Gerkey and Maja J Matarić. 2004. A formal analysis and taxonomy of task allocation in multi-robot systems. *The International journal of robotics research* 23, 9 (2004), 939–954.

[16] Brian P Gerkey and Maja J Matarić. 2004. On role allocation in RoboCup. In *RoboCup 2003: Robot Soccer World Cup VII 7.* Springer, 43–53.

[17] Felix Gervits, Dean Thurston, Ravenna Thielstrom, Terry Fong, Quinn Pham, and Matthias Scheutz. 2020. Toward Genuine Robot Teammates: Improving Human-Robot Team Performance Using Robot Shared Mental Models.. In *AAMAS.* 429–437.

[18] Moojan Ghafurian, Neil Budnarain, and Jesse Hoey. 2019. Role of emotions in perception of humanness of virtual agents. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems.* 1979–1981.

[19] Deborah L Gladstein. 1984. Groups in context: A model of task group effectiveness. *Administrative science quarterly* (1984), 499–517.

[20] Michael A Goodrich, Timothy W McLain, Jeffrey D Anderson, Jisang Sun, and Jacob W Crandall. 2007. Managing autonomy in robot teams: observations from four experiments. In *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 25–32.

[21] Stephen G Green and Thomas D Taber. 1980. The effects of three social decision schemes on decision group process. *Organizational behavior and human performance* 25, 1 (1980), 97–106.

[22] Tyler Gunn and John Anderson. 2013. Effective task allocation for evolving multi-robot teams in dangerous environments. In *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT),* Vol. 2. IEEE, 231–238.

[23] Feyza Merve Hafızoğlu and Sandip Sen. 2018. The Effects of Past Experience on Trust in Repeated Human-Agent Teamwork. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems.* 514–522.

[24] David Hauser, Gabriele Paolacci, and Jesse Chandler. 2019. Common concerns with MTurk as a participant pool: Evidence and solutions. (2019).

[25] Barbara Hayes-Roth. 1985. A blackboard architecture for control. *Artificial intelligence* 26, 3 (1985), 251–321.

[26] Guido Hertel, Susanne Geister, and Udo Konradt. 2005. Managing virtual teams: A review of current empirical research. *Human resource management review* 15, 1 (2005), 69–95.

[27] Toshiyuki Inagaki. 2008. Smart collaboration between humans and machines based on mutual understanding. *Annual reviews in control* 32, 2 (2008), 253–261.

[28] Daniel Kahneman. 2011. *Thinking, fast and slow.* Macmillan.

[29] G Ayorkor Korsah, Anthony Stentz, and M Bernardine Dias. 2013. A comprehensive taxonomy for multi-robot task allocation. *The Intl Journal of Robotics Research* 32, 12 (2013), 1495–1512.

[30] Harold W Kuhn. 1955. The Hungarian method for the assignment problem. *Naval research logistics quarterly* 2, 1-2 (1955), 83–97.

[31] Vivian Lai and Chenhao Tan. 2019. On human predictions with explanations and predictions of machine learning models: A case study on deception detection. In *Proceedings of the conference on fairness, accountability, and transparency.* 29–38.

[32] Lindsay Larson and Leslie A DeChurch. 2020. Leading teams in the digital age: Four perspectives on technology and what they mean for leading teams. *The Leadership Quarterly* 31, 1 (2020), 101377.

[33] John E Mathieu, John R Hollenbeck, Daan van Knippenberg, and Daniel R Ilgen. 2017. A century of work teams in the Journal of Applied Psychology. *Journal of applied psychology* 102, 3 (2017), 452.

[34] John E Mathieu and Tammy L Rapp. 2009. Laying the foundation for successful team performance trajectories: The roles of team charters and performance strategies. *Journal of Applied Psychology* 94, 1 (2009), 90.

[35] Alejandro R Mosteo and Luis Montano. 2010. A survey of multi-robot task allocation. *Instituto de Investigacin en Ingenierła de Aragn (I3A), Tech. Rep* (2010).

[36] Ernesto Nunes, Marie Manner, Hakim Mitiche, and Maria Gini. 2017. A taxonomy for task allocation problems with temporal and ordering constraints. *Robotics and Autonomous Systems* 90 (2017), 55–70.

[37] Laurent Perron and Vincent Furnon. [n.d.]. *OR-Tools.* Google. https://developers.google.com/optimization/

[38] Pooja Prajod, Mohammed Al Owayyed, Tim Rietveld, Jaap-Jan van der Steeg, and Joost Broekens. 2019. The effect of virtual agent warmth on human-agent negotiation. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems.* 71–76.

[39] Phanish Puranam, Oliver Alexy, and Markus Reitzig. 2014. What's "new" about new forms of organizing? *Academy of Management Review* 39, 2 (2014), 162–180.

[40] Sarvapali D. Ramchurn, Trung Dong Huynh, Yuki Ikuno, Jack Flann, Feng Wu, Luc Moreau, Nicholas R. Jennings, Joel E. Fischer, Wenchao Jiang, Tom Rodden, Edwin Simpson, Steven Reece, and Stephen J. Roberts. 2015. HAC-ER: A Disaster Response System Based on Human-Agent Collectives. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems.* 533–541.

[41] Bruce A Reinig. 2003. Toward an understanding of satisfaction with the process and outcomes of teamwork. *Journal of Management Information Systems* 19, 4 (2003), 65–83.

[42] Peter Robe and Sandeep Kaur Kuttal. 2022. Designing pairbuddy—a conversational agent for pair programming. *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 4 (2022), 1–44.

[43] Paul Robinette, Alan R. Wagner, and Ayanna M. Howard. 2013. Building and Maintaining Trust Between Humans and Guidance Robots in an Emergency. In *AAAI Spring Symposium: Trust and Autonomous Systems.* Stanford, CA, 78–83.

[44] Ariel Rosenfeld, Noa Agmon, Oleg Maksimov, and Sarit Kraus. 2017. Intelligent agent supporting human–multi-robot team collaboration. *Artificial Intelligence* 252 (2017), 211–231.

[45] Chen Rozenshtein and David Sarne. 2020. Incorporating Failure Events in Agents' Decision Making to Improve User Satisfaction. *International Joint Conferences on Artificial Intelligence Organization.*

[46] Nathan Schurr, Janusz Marecki, Milind Tambe, Paul Scerri, and J.P. Lewis. 2009. *Published Articles & Papers.* Vol. 42. Kluwer, Chapter The Defacto System: Coordinating Human-Agent Teams for The Future of Disaster Response.

[47] Yoav Shoham and Kevin Leyton-Brown. 2008. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations.* Cambridge University Press.

[48] Eric M Stark and Paul E Bierly III. 2009. An analysis of predictors of team satisfaction in product development teams with differing levels of virtualness. *R&d Management* 39, 5 (2009), 461–472.

[49] Ruth Wageman, J Richard Hackman, and Erin Lehman. 2005. Team diagnostic survey: Development of an instrument. *The journal of applied behavioral science* 41, 4 (2005), 373–398.

[50] Michael Wooldridge, Nicholas R Jennings, and David Kinny. 1999. A methodology for agent-oriented analysis and design. In *Proceedings of the third annual conference on Autonomous Agents.* 69–76.

[51] Wanqing Zhao, Qinggang Meng, and Paul WH Chung. 2015. A heuristic distributed task allocation method for multivehicle multitask problems and its application to search and rescue scenario. *IEEE transactions on cybernetics* 46, 4 (2015), 902–915.

[52] Shuo Zhou, Timothy Bickmore, Michael Paasche-Orlow, and Brian Jack. 2014. Agent-user concordance and satisfaction with a virtual hospital discharge nurse. In *International conference on intelligent virtual agents.* Springer, 528–541.