Leveraging Cognitive States in Human-Robot Teaming

Jack Kolb, Harish Ravichandar, Sonia Chernova

***Abstract*— Mixed human-robot teams (HRTs) have the potential to perform complex tasks by leveraging diverse and complementary capabilities within the team. However, assigning humans to operator roles in HRTs is challenging due to the significant variation in user capabilities. While much of prior work in role assignment treats humans as interchangeable (either generally or within a category), we investigate the utility of *personalized* models of operator capabilities based in relevant human factors in an effort to improve overall team performance. We call this approach *individualized role assignment (IRA)* and provide a formal definition. A key challenge for IRA is associated with the fact that factors that affect human performance are not static (e.g., one's ability to track multiple objects can change during or between tasks). Instead of relying on time-consuming and highly-intrusive measurements taken *during* the execution of tasks, we propose the use of short cognitive tests, taken *before* engaging in human-robot tasks, and predictive models of individual performance to perform IRA. Results from a comprehensive user study conclusively demonstrate that IRA leads to significantly better team performance than a baseline method that assumes human operators are interchangeable, even when we control for the influence of the robots' performance. Further, our results point to the possibility that such relative benefits of IRA will increase as the number of operators (i.e., choices) increase for a fixed number of tasks.**

I. Introduction

Mixed human-robot teams with close collaboration between humans and robots are currently being explored across a wide range of domains, including manufacturing [9], [18], [36], defense [8], and search and rescue [23]. To facilitate coordination across team members, each agent must be assigned a job to perform, which is a challenge known as the *role assignment* problem. Prior work on role assignment has demonstrated that we can take advantage of heterogeneous capabilities within a team by allocating agents to tasks or roles such that team performance is maximized [16], [19], [28], [35]. Such methods have typically been applied to robotic agents; however, they can equally be applied to *human* agents.

Existing role allocation methods that have considered human agents either model all humans as interchangeable [11], [23], [36], or distinguish high level categories of humans (e.g., medic vs. firefighter) while still treating every member within that category as interchangeable [28]. However, recent work has demonstrated that people vary greatly in their performances in human-robot task domains [20]. Extensive investigation in human factors has revealed that human-robot team performance is significantly impacted by individual cognitive differences between humans. Specifically, individual differences between operators with respect to situational awareness, visual attention, and spatial reasoning capabilities all have been shown to considerably impact human-robot teaming performance [4], [6], [8], [17].

In this work, we first attempt to bridge the gap between role assignment algorithms that do not respect individual differences in humans and the insights from human factors literature that point to the importance of such differences. We do so by considering *personalized* models of operator capabilities based in relevant human factors when assigning roles to human operators. We hypothesize that considering such personalized models will improve role assignment and overall team performance by assigning each operator to roles that more closely match their strengths.

A critical challenge in modeling individual human capabilities is that relevant human factors that affect team performance are not constant; they are influenced by fatigue, cognitive workload, behavior of the robot team, and other characteristics that can change over time [3], [16]. Typical methods that measure such factors, such as the well-studied SAGAT questionnaire [13], are time-consuming and often highly intrusive as they require interrupting the operator during task execution.

To address the above challenges and enable practical role assignment based on individualized differences, we require a fast and efficient means by which to measure these human capabilities and predict performance. Our recent work has demonstrated that short cognitive tests, taken *before* engaging in human-robot tasks, can help predict performance [20]. Leveraging such recent advances, we hypothesize that using abstracted cognitive tests that are easy to administer before engagement in robot operation tasks can effectively inform role assignment algorithms that wish to account for individual differences.

Concretely, our work explores the research question: *can the performance of human-robot teaming be improved by leveraging cognitive states for role assignment?* We are specifically interested in multi-human multi-robot teaming, where each team member is to be assigned to a unique robot operation role.

We contribute a role assignment pipeline that leverages easy-to-administer cognitive tests to perform effective assignment of human operators. Our approach first constructs a predictive model that encodes the relationship between cognitive test scores and performance on robot operation tasks. This predictive model is constructed offline by measuring

This work was supported by the Army Research Lab under Grant W911NF-17-2-0181 (DCIST CRA)

Jack Kolb, Harish Ravichandar, and Sonia Chernova are with the College of Computing at the Georgia Institute of Technology, North Avenue, Atlanta, GA 30332, USA {kolb, harish.ravichandar, chernova}@gatech.edu.

both cognitive test scores and task performance scores of human subjects. Subsequently, our pipeline optimizes the role assignment such that the team's predicted cumulative performance across all tasks is maximized.

We conducted a comprehensive user study involving 29 individual participants and 3654 possible teams. As our work is the first to apply cognitive states in assigning people to robot operation roles, we compared the performance of our pipeline to a baseline that ignores such cognitive states and assigns roles uniformly randomly. Our results show that our pipeline outperforms the baseline. Notably, we are able to demonstrate that team performance is improved using solely the cognitive states in the areas we measured, without any prior knowledge of the team members' capabilities at the human-robot tasks.

Video demonstrations of the human-robot tasks and cognitive tests presented in this paper, as well as the source code, are available at `https://github.com/GT-RAIL/cognitive-states-in-human-robot-teaming`.

II. Related Work

Our work is informed by the rich literature in human factors research. Extensive studies in this area have shown that human-robot team performance is significantly impacted by individual cognitive differences between humans [6], [8].

One widely studied characteristic is *attentional control*, which is defined as a person's ability to focus and shift attention in a flexible manner [10]. Since multi-robot control tasks inherently require an operator to perform multitasking, it is not surprising that human ability to coordinate multi-robot teams has been tied to the operator's attentional control ability. Prior work has shown that there are individual differences in multitasking performance between operators [30], [34] and that poor attentional control is closely related to low operator performance [4], [17]. Furthermore, studies show that operators with better attentional control are able to allocate their attention more flexibly and effectively, and attention-switching flexibility can predict performance on complex tasks [1], [2]. In our work, we particularly focus on *visual attention control*, a specific form of attention control as it related to visual perception [12].

Additional human factors tied to differing operator abilities are *spatial ability* and *situational awareness*, both of which have also been shown to impact operator performance in multi-robot tasks [6], [7]. For example, operators with higher spatial ability have been shown to exhibit more effective visual scanning and target detection techniques during HRI-related multitasking [5]–[7].

In prior work, visual attention control, spatial ability, and situational awareness have been measured through a variety of cognitive tasks. Visual attention tests were first developed in the 1980's [27] to understanding the impacts of design and training on visual attention control [32], [33]. These tests present users with identical balls bouncing around a contained screen area, and task the user with tracking select balls for a duration. Situational Awareness tests developed in parallel [13], tasking users with observing domains for short durations and then recounting descriptive and inferential information about environment elements. Our work develops and adopts short (under 10 minutes) tests that can effectively measure visual attention control, spatial ability, and situational awareness by leveraging recent advances in neuroscience and human factors [21], [24], [31]. Further, we administer these tests on abstracted settings *before* the operators engage in robot control tasks.

Additionally, physiological signals, such as heart rate [22], [25], have been used as real-time indicators of cognitive state during human-robot teaming. In this work, we limit ourselves to *a priori* role assignment, though real-time reassignment remains an important direction for future work.

Closely related to our work are efforts in which task assignment between multiple agents is performed with account for workload. Such works include [15], in which allocation is adapted between a human and a (single) robot, and [29], in which multiple humans are considered. While individual workload has been shown to be significant contributing factor to team performance, these works do not model or leverage individual differences in other relevant cognitive human factors. Further, existing methods either assume that human operators are interchangeable, or have been only evaluated in simulation. In contrast, we embrace inherent differences across three cognitive skills and evaluate our approach using a comprehensive user study.

III. Problem Formulation

Our overall objective is to maximize the performance of large, heterogeneous, human-robot teams. Prior works have extensively explored algorithms for assigning robots to tasks [16], [19], [28], [35]. In this work, we instead examine the problem of assigning *human operators* to *roles* so as to maximize the performance of the joint human-robot team. To study the human operator assignment in isolation, we hold constant the assignment of robot agents and only vary human operator role assignment. Formally, we define the problem as follows.

Let $\mathcal{O}$ be a set of N human operators, and $\mathcal{R}$ be a set of M roles, such that $N \geq M$. Given $\mathcal{O}$ and $\mathcal{R}$, we define $\mathbf{s}_{n,m}$ to be the true performance score of Operator $n \in \mathcal{O}$ executing Role $m \in \mathcal{R}$. We define a role assignment as the function $\pi : \mathcal{R} \to \mathcal{O}$ between operators and roles. We say that Operator n is assigned to Role m if $\pi(m) = n$. Our goal is to find the role assignment π that maximizes the cumulative performance score of all operators:

$$\mathbf{S} = \sum_{m=1}^{M} \mathbf{s}_{\pi(m),m}$$

We assume that tasks are independent, so a given assignment does not affect the performance of other operators.

The above formulation represents a domain-independent definition of role assignment that is widely used in robotics research, as surveyed in [19]. However, in the context of human task assignment, we often cannot assume exact knowledge of the true cumulative performance $\mathbf{S}$ prior to

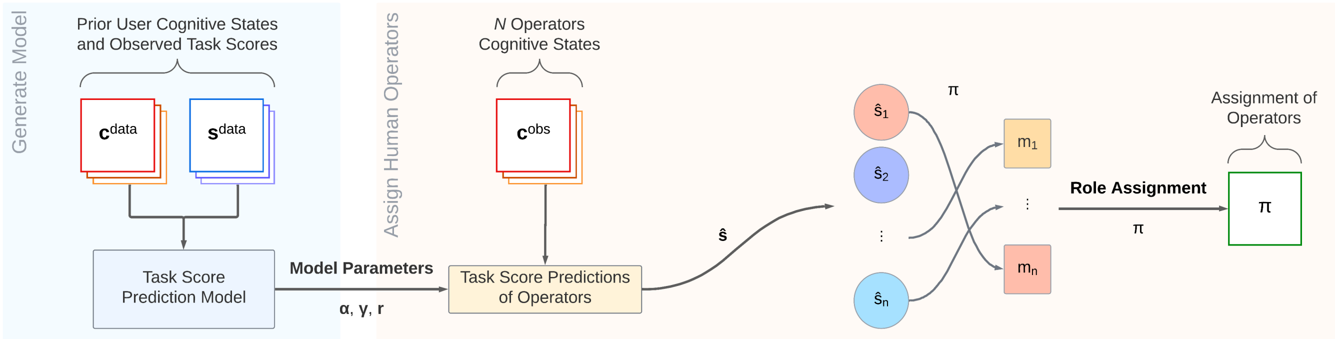


Fig. 1: Diagram overview of the assignment pipeline. A prediction model of task performance is fit using prior observations Z of users completing the cognitive skill tests and human-robot tasks. The model is then used to find the team's predicted task scores $\hat{\mathbf{s}}$, and role assignment is conducted to maximize $\hat{\mathbf{s}}$.

task execution. This may occur because Operator n has never performed Role m in the past, or because performance naturally varies based on the operator's current cognitive state (e.g., fatigue or stress affecting cognitive skills). We term this problem **individualized role assignment (IRA)**, as individual task performance, and thus the performance of the entire team as a whole, depends on the cognitive state or capabilities of the operator. In the absence of the true $\mathbf{S}$, we formulate the problem as follows:

- We model role assignment with respect to the *estimated* cumulative performance $\hat{\mathbf{S}}$, and perform role assignment to maximize this estimate, and
- We assume that estimating $\hat{\mathbf{S}}$ by exhaustively testing $N \times M$ role assignments (i.e., exhaustively measuring each operator's performance on each task) is not a viable option given the resource and time constraints of realistic scenarios.

To address the above problem, we propose that an inexpensive (w.r.t. resources and time) evaluation of *operator* cognitive skills be conducted in order to obtain observations Z consisting of scores of cognitive skill tests. We then aim to construct a predictive model $f_{\mathbf{s}}(\cdot)$ that can estimate cumulative *performance* score $\hat{\mathbf{S}}$ from observations Z and a specified assignment function π. Formally, this is given by

$$\hat{\mathbf{S}} = f_{\mathbf{S}}(Z, \pi) \tag{1}$$

As such, we can compute a role assignment π that maximizes $\hat{\mathbf{S}}$. Individualized role assignment succeeds if the assignment based on $\hat{\mathbf{S}}$ closely approximates true optimal role assignment based on the true (but unknown) performance score $\mathbf{S}$. Below, we describe a generalizable solution framework.

IV. Individualized Role Assignment

We illustrate the overview of the proposed framework to solve IRA problems in Fig. 1. First, we construct the model $f_{\mathbf{s}}(\cdot)$ that predicts task performance scores based on observations of operator capabilities Z (Section IV-A). This model is constructed based on observations of a preliminary set of users in order to capture general performance trends. Second, we use $f_{\mathbf{s}}(\cdot)$ to predict the performance scores of a new, previously unobserved, set of operators for each of the roles (Section IV-B). Third, we optimize the assignment of operators to roles such that the estimated cumulative performance score is maximized (Section IV-C). Below, we describe each of these three modules in detail.

A. Modeling capability-performance relationships

We first model the relationships between performance scores and operator capabilities, as measured by cognitive tests. To this end, we collect a dataset by having N users i) complete a battery of U cognitive tests, and ii) perform M tasks of interest. We denote this dataset by $\mathcal{D} = \{\mathbf{c}^{\text{data}}_{u,n}, \mathbf{s}^{\text{data}}_{m,n}\}$, $\forall u \in 1, ..., U$, $n \in 1, ..., N$, and $m \in 1, ..., M$, where $\mathbf{c}^{\text{data}}_{u,n}$ represents an Operator n's cognitive skill as measured by the uth cognitive test, and $\mathbf{s}^{\text{data}}_{m,n}$ represents Operator n's score on the mth task.

Given the dataset $\mathcal{D}$, we train $U \times M$ linear regression models to identify the relationship between each skill-task pairing. In contrast to requiring considerable historic performance data, these models enable us to predict a new user's performance on a task purely based on their cognitive test scores. In addition to the linear model parameters, the correlation coefficient of each skill-task pairing is calculated to measure the strength of the pairing's relationship. For the pairing between the u^{th} cognitive state and the m^{th} task, we compute both model parameters (the linear regression slopes $\alpha_{m,u}$ and y-intercepts $\beta_{m,u}$) as well as the correlation strength (the Spearman's correlation coefficients $r_{m,u}$).

We choose to encode skill-task relationships using linear models as we did not encounter non-linear trends in our experiments. However, note that our framework is not restricted to linear models and can extended to nonlinear models in situations that require them.

B. Performance Score Prediction

Given the regression models representing the relationships between every skill-task relationship, we now turn to the problem of predicting task performance for new users. Concretely, we begin by measuring the score $\mathbf{c}^{\text{obs}}_{u,n'}$ for each new operator n' on each cognitive skill test u. Given these test scores and the model parameters α, β, and r, we compute the predicted score $\hat{\mathbf{s}}_{n',m}$ of Operator n' on Task m as follows

$$\hat{\mathbf{s}}_{n',m} = \sum_{u=0}^{U} \gamma_{m,u}(\alpha_{m,u}\ \mathbf{c}^{\text{obs}}_{u,n'} + \beta_{m,u}) \tag{2}$$

where $\gamma_{m,u} = \frac{|r_{m,u}|}{\sum_{u'=0}^{U}|r_{m,u'}|}$ is the adaptive weight representing the influence of the uth skill on the mth task, and is computed as the corresponding correlation strength normalized to the sum of all correlations for the task. The predicted score for each task can be viewed as a weighted sum of linear predictions contributed by each cognitive skill.

C. Role Assignment

Finally, our framework assigns the N' new operators to our M tasks, such that each task has one operator assigned to it. To this end, we begin by computing the predicted cumulative score that would have been achieved by assigning the N' operators according to an arbitrary assignment function π as

$$\hat{\mathbf{S}} = f_{\mathbf{S}}(Z, \pi) = \sum_{m=1}^{M} \hat{\mathbf{s}}_{\pi(m),m} \tag{3}$$

where Z represents the collection of observations $\mathbf{c}^{\text{obs}}_{n',m} \forall n', m$, and $\hat{\mathbf{s}}_{\pi(m),m}$ is computed using the predictive model described in (2).

As our environment is designed such that each human can perform one task at a time, each task requires one human, and assignments are fixed, our assignment problem falls under the well-established *ST-SR-IA* classification of Gerkey's task allocation taxonomy [14]. The Hungarian Algorithm [26] is a known solution to ST-SR-IA problems, capable of identifying the role assignment π that maximizes the predicted cumulative score $\hat{S}$, as defined in (3).

V. Multi-Robot Evaluation Domains

Building upon our recent work [20], we explore individualized role assignment in the context of three human-robot tasks: target search, ad-hoc network construction, and sample return (Fig. 3). We selected these three scenarios because they represent widely encountered multi-robot scenarios in search-and-rescue and exploration domains.

Target Search: The operator controls four virtual aerial robots to search for five targets (caches) in a desert environment filled with dead trees and abandoned buildings. The robots are controlled via a top-down waypoint navigation interface shown in the middle window in Fig. 3(a). Four side windows show the view of the downward-facing camera of each robot, which the operator utilizes to locate the targets. Operators are given 10 minutes to complete the task, and their search is guided by a grey area of interest for each target. The task is designed to challenge an operators ability to monitor multiple areas simultaneously. The task performance metric is:

$$s_{n,\text{-}} = \frac{\sum_{i=0}^{4} \min_j(D_{ij}(0)) - \min_{t,j}(D_{ij}(t))}{T} \tag{4}$$

where $D_{ij}(t)$ is the distance from cache i to robot j at time t, and T is the participant's total time to complete the task.

A participant is scored by their progress towards completing the task (see Eq. 4), represented by how close they are to locating each cache, with the participant's time to complete the stage used to distinguish between participants who complete the stage in under 10 minutes.

Ad-Hoc Network Construction: The operator uses four aerial robots and four ground robots to extend a communication network from a robot base to five supply caches. Similar to Task 1, robots are controlled via a top-down waypoint navigation interface (Fig. 3(b)). The movement of aerial robots is not limited by obstacles, whereas ground vehicles will be obstructed by obstacles. Robot control is constrained to the boundaries of the communication network, with participants being unable to move robots that are not connected to the primary network, requiring participants to extend the network gradually from the base. This task is designed for participants to benefit from their ability to visualize how the network topology will change as robots are relocated. The task performance metric is:

$$s_{n,\text{-}} = \frac{T}{\max_t(C(t))} \tag{5}$$

where T is the participant n's total time to complete the task, and $C(t)$ is the number of caches connected to the network at a time t.

Sample Return: The operator controls four ground robots to return five supply caches to the base. Similar to the other tasks, robots are controlled via a top-down waypoint navigation interface (Fig. 3(c)). Operators benefit by their ability to navigate the ground robots to reach and return the caches, while avoiding obstacles and rerouting robots as necessary. The task performance metric is:

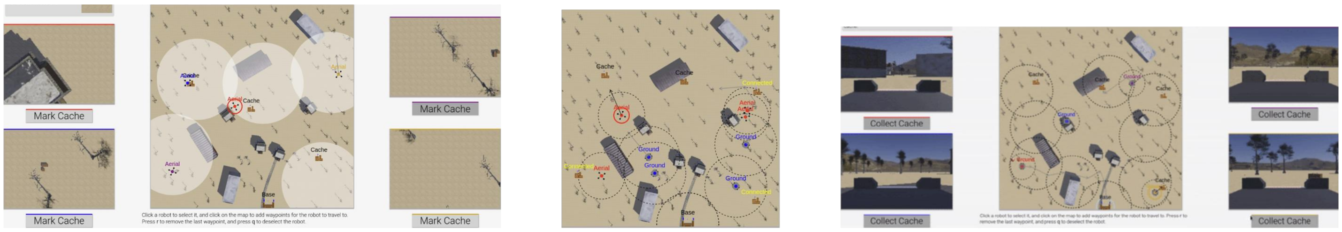
$$s_{n,\text{-}} = \frac{\sum_{i=0}^{4} D_{ij_{\text{collect}}}(0) - \min_{t,j}(D_{ij_{\text{collect}}}(t))}{T} + \frac{\sum_{i=0}^{4} D_{ij_{\text{return}}}(0) - \min_{t,j}(D_{ij_{\text{return}}}(t))}{T} \tag{6}$$

where $D_{ij_{\text{collect}}}(t)$ is robot j's distance to collecting cache i at time t, $D_{ij_{return}}(t)$ is robot j's distance to returning a collected cache i at time t, and T is the participant n's total time to complete the task.

Each of the above scenarios is of similar complexity and can be carried out by a human operator with minimal training. However, we hypothesize that each of the three roles require different types of cognitive reasoning. We utilize the cognitive ability tests presented in the following section to obtain observations Z and train $f_{\mathbf{s}}(\cdot)$.

VI. Cognitive Skill Tests

To represent a participant's cognitive state we apply three cognitive skill tests. We choose cognitive abilities that match the cognitive requirements of our human-robot interactive tasks, and design a short browser-based test for each cognitive skill. The tests are grounded in cognitive science



(a) *Target Search* (b) *Network Construction* (c) *Sample Return*

Fig. 2: Screenshots of each human-robot teaming task. In *Target Search*, users navigate aerial robots to find caches hidden in a map. In *Ad-Hoc Network Construction*, users position aerial and ground robots to extend a communications network. In *Sample Return*, users navigate ground robots to collect and return caches to a central base. Tasks are completed virtually.

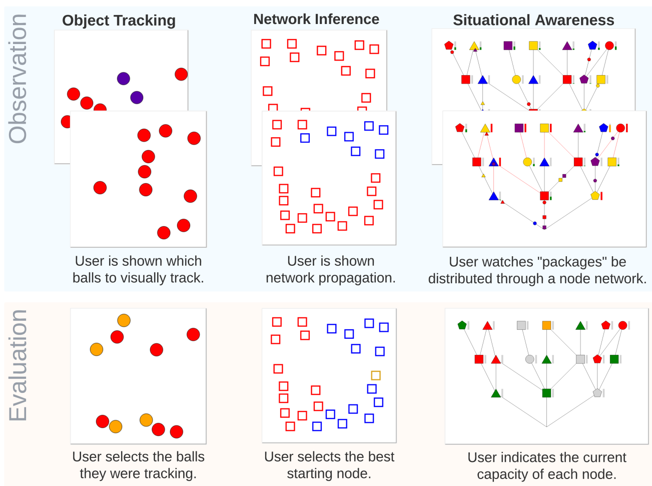


Fig. 3: Overview of each cognitive skill test: Object Tracking (Visual Attention), Network Inference, and Situational Awareness. In each test users observe an environment, and then are asked questions pertaining to the target cognitive skill. Each test has multiple increasingly difficult stages.

literature, take around 10 minutes to complete, and are difficult enough to capture variation between users.

Visual Attention: Visual attention is a user's ability to track environment stimuli. As all three of our human-robot interactive tasks involve managing multiple robots, visual attention may be important to user performance at our tasks. An established test for visual attention, multi-object tracking, has a participant keep track of one or more of a set of identical balls for a period of time as the balls collide and deflect, and then has the participant identify the tracked balls [24], [27]. We apply this test using 15 balls total, and have participants incrementally track from 1 to 7 balls (Fig. 3 left). Participants are scored by their total number of correctly identified balls from the 2nd through 7th rounds.

Network Inference: Network inference is a user's ability to learn and infer a hidden graph structure between nodes. As two of our human-robot interactive tasks involve leveraging network connections between robots, participants who are able to effectively identify graph connections may perform better on those tasks. Our network inference test is based on work in [21] and validated by our prior work [20]. Participants are presented with a set of nodes, and watch examples of color propagating through the nodes at interval timesteps. Participants then identify the starting node that enables color propagation in the least number of timesteps (Fig. 3 middle). Participants have up to three guesses per network, and are scored by their guesses cumulative timesteps from the optimal node.

Situational Awareness Situational awareness is a user's ability to maintain an accurate mental model of a changing environment. This is pertinent to managerial human-robot interactive tasks where operators must maintain situational awareness of robot locations. Related work has shown situational awareness to be correlated to task performance [31], however situational awareness is variable within a task. We instead measure a user's ability to maintain situational awareness given increasing environment complexity. Our test is based on the SAGAT format [13] and was used in our prior work [20]. In the test, participants watch a stream of packages be distributed throughout a network of warehouses. Warehouses distribute packages at a lower rate than the package influx, resulting in network breakdowns as warehouses become full. The participant is periodically asked to identify the storage level of each warehouse (Fig. 3 right). As more warehouses overfill, the network breaks down, increasing the situational awareness required to keep track of the warehouse states.

VII. User Study Design

We verified our approach by conducting a user study based on the tasks and cognitive tests described in the previous section. We recruited 29 participants for the study from the Georgia Tech community (Ages 18 to 30; 38% female; 69% novice robot operators).

Each participant completed the $U = 3$ cognitive skill tests from Section VI followed by the $M = 3$ simulated human-robot teaming tasks from Section V. Participants completed all six study components virtually while on a video call with a researcher. We applied counterbalancing to both the order of cognitive skill tests, and the order in which human-robot teaming tasks were presented.

To model $f_{\mathbf{s}}(\cdot)$, we split the data from all participants into a training set and test set. Data from participants in the training set was used to fit the predictive model $f_{\mathbf{s}}(\cdot)$ to model correlations between the cognitive test scores and task performance. Data from the test set was withheld and used to validate the performance of the role assignment algorithm.

For each experiment, the size of the test set was set to N, the number of operators available for the algorithm to choose from. For example, if $N = 5$, then five operators are available to be selected for $M = 3$ roles. We refer to the selected operators as a *team*, and report their cumulative task performance score[1]. In each experimental trial, we used $f_{\mathbf{s}}(\cdot)$ to predict how well each of the N operators would perform each of the three task roles (target search, ad-hoc network construction, and sample return), and assign them to roles to maximize the total team performance. We hypothesize that each of our three roles requires different types of cognitive reasoning, leading us to the following research hypotheses:

Research Hypothesis 1: Given an equal number of operators and roles ($N = M$), individualized role assignment will lead to improved total team performance over random operator assignment.

Research Hypothesis 2: The improvement resulting from individualized role assignment will be greater as N grows larger than M.

The second hypothesis stems from the fact that given multiple operators to choose from, the model will be able to identify more suitable individuals to assign to the tasks, leading to further improvements in performance.

VIII. Study Results

We conducted the following experiments to validate our two research hypotheses.

Experiment 1: To validate our first hypothesis, we evaluated all possible teams of $N = 3$ participants, resulting in $\binom{29}{3}$ = 3654 teams. For each evaluation, we used the team's unique combination of 3 participants in the test set, and the remaining 26 participants in the training set. We report role assignment results for the following four assignment conditions:

- *random (baseline)* – baseline approach most widely used in prior work, which does not take individual operator differences into account;
- *worst (ground truth)* – the worst possible performance achievable by the $N-3$ operators if they were assigned to their least optimal roles, included for comparison;
- *best (ground truth)* – the best possible performance achievable by the $N-3$ operators if they were assigned to their most optimal roles, included for comparison;
- *IRA (algorithmic approach)* – our approach for automated role assignment based on cognitive pretests.

Figure 4 presents a comparison of the above four role assignment variants. The top row (A) shows that random

[1]The performance of the operators on the team is independent from each other in this experiment.

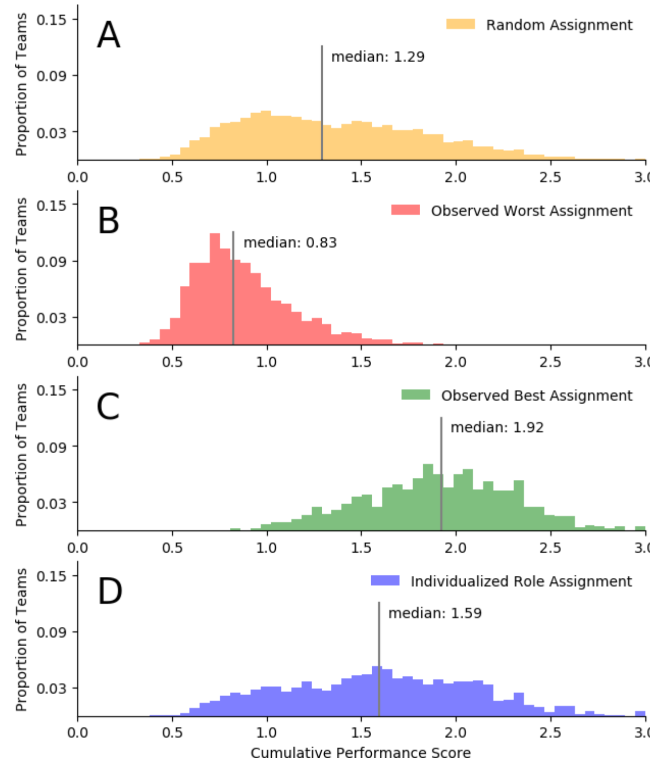


Fig. 4: Histograms of cumulative performance scores for all teams using several assignment methods, with $N = M = 3$. **A** shows the baseline *random assignment*. **B** shows the worst possible assignments using each team's observed task scores. **C** shows the best possible assignments using the observed task scores. **D** shows scores from *individualized role assignment*.

performance is widely spread, though not entirely uniform in its distribution due to fact that participant abilities themselves were not uniformly distributed. Rows (B) and (C) show the worst and best ground truth assignments, respectively; these are obtained based on each participant's true task performance in each role. We observe that even in row (C) some teams perform significantly lower than others, which occurs when three lesser-skilled participants are grouped together. Finally, in row (D) we observe the IRA performance. Comparing IRA to the random condition, we observe the median cumulative performance score increases by +0.30 (23.5%). Additionally, 30.9% of IRA teams were optimally assigned, compared to only 16.6% of random teams.

Given $N = 3$ operators and $M = 3$ roles, IRA has a total choice of 6 possible unique role assignments to make for each team. Fig. 5 shows how the IRA-chosen role assignment ranks in comparison to the other choices. If role assignment were to be performed at random, we would expect each possible role assignment to be equally likely, meaning that the best possible assignment would be just as likely as the worst. This value is shown by the horizontal line in the figure. In comparison, we observe that IRA is 4.0 times more likely to select the best assignment than the worst assignment, and overall 72.6% of IRA teams outperform random assignment.

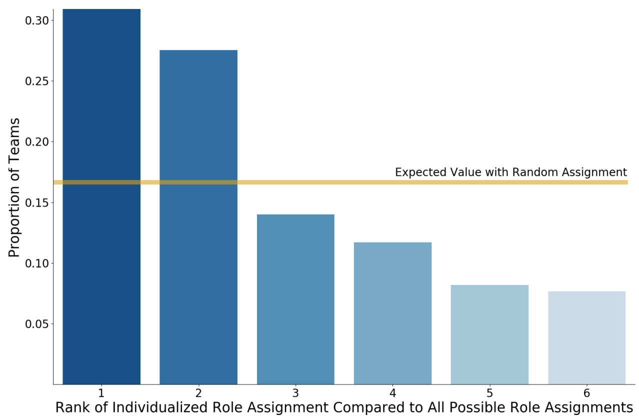


Fig. 5: Rank of each team's skill-based assignment against all six possible assignments for the team; lower rank indicates superior performance. The gold bar indicates random assignment, a uniform distribution across all ranks.

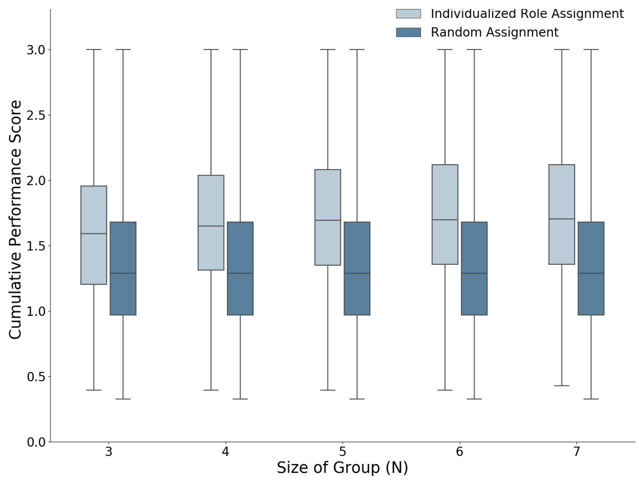


Fig. 6: Distribution of team scores for varying sample group sizes *N*. Assignment is made for *M* roles for each combination of *N* users.

In summary, we find that these results support **Hypothesis 1**, and that given an equal number of operators and roles IRA significantly outperforms random operator assignment (Mann-Whitney U test, $p < 0.001$).

Experiment 2: To investigate our second hypothesis, we vary the number of available operators from $N = 3$ to $N = 7$. Figure 6 compares the cumulative performance scores of teams formed across the random and IRA conditions. We observe that for all N trialed, IRA shows significant performance improvement over random assignment (Mann-Whitney U test, $p < 0.001$), and median cumulative performance scores increased monotonically from 1.593 ($N = 3$) to 1.703 ($N = 7$).

IX. Summary and Discussion

We find that a team's cognitive states can be leveraged to improve human-robot teaming. Through evaluating user data from three cognitive skill tests and three human-robot operator tasks, we find that role assignment using team member cognitive states result in scores significantly better than random assignment, and outperform random assignment in 72.6% of cases. Our work reinforces prior work suggesting that components of cognitive states can predict the performance of impending robot operator tasks [20], and verifies our hypothesis by applying the concept to meaningfully improve role assignment.

The assignment pipeline presented is a baseline that future work can improve upon. We explore several potential modifications to its components. In *Component 1*, using correlation coefficients as a weighting mechanism results in weak correlations still having meaningful weight. However, exponating the correlating coefficients – to force a greater disparity between meaningful and chaotic correlations – did not improve performance. Alternative weighting mechanisms can be considered to better prioritize skill-task relationships that are impactful for predicting task performance.

We also find that the metrics used to score tasks have large effects on the pipeline's performance. We trialed 10 scoring metrics for the three human-robot collaborative tasks (2 for *Task 1*, 5 for *Task 2*, 3 for *Task 3*). All metrics were monotonically decreasing measurements of task progress. Different metrics resulted in a range of average rank percentiles, from as low as 52.7% (random: 50.0%) to as high as 69.2%. Selecting an effective performance evaluation metric can be critical for finding stronger trait-task correlations.

Our work opens several avenues for future work. In the assignment pipeline, *Component 1* modeled skill-task relationships using linear regressions – alternative models (e.g., neural networks) may capture more complex relationships. Additionally, our study was situated in a virtual environment – using real-world robot operation tasks may result in different skill-task relationships. Including more cognitive skill tests and human-robot tasks can also further our understanding of how cognitive states affect HRT performance.

Another direction for this work is to integrate team dynamics. We limited our study design to prevent team dynamics from affecting task outcomes, by making tasks independent and isolating users. In reality, collaborative teamwork is an integral component of many teams. Future work can draw upon the *team effectiveness* literature to consider additional variables in role assignment for multi-user environments.

X. Ethical Considerations

Our work raises important ethical questions of whether individual human differences should be considered in role assignment, particularly with how differences can be used to justify discrimination in workplace and social environments. Importantly, our work uses cognitive states, not descriptive traits such as physical or developmental traits. Our work is reliant on the present mental state of users, using cognitive abilities that are dynamic in nature and uncoupled from descriptive traits. We are interested in the relevance of cognitive states towards robot operation performance at the time of role assignment, not long term, and thus do not see our work being applied to further political and social prejudices.

References

[1] Lisa Feldman Barrett, Michele M Tugade, and Randall W Engle. Individual differences in working memory capacity and dual-process theories of the mind. *Psychological bulletin*, 130(4):553, 2004.

[2] M Kathryn Bleckley, Francis T Durso, Jerry M Crutchfield, Randall W Engle, and Maya M Khanna. Individual differences in working memory capacity predict visual attention allocation. *Psychonomic bulletin & review*, 10(4):884–889, 2003.

[3] Udo Boehm, Dora Matzke, Matthew Gretton, Spencer Castro, Joel Cooper, Michael Skinner, David Strayer, and Andrew Heathcote. Real-time prediction of short-timescale fluctuations in cognitive workload. *Cognitive Research: Principles and Implications*, 6(1):1–29, 2021.

[4] WL Chappelle, KK McMillan, P Novy, and K McDonald. Psychological profile of usaf unmanned aerial systems predator & reaper pilots. *Aviation, Space, and Environmental Medicine*, 81:339, 2010.

[5] Jessie YC Chen. Uav-guided navigation for ground robot teleoperation in a military reconnaissance environment. *Ergonomics*, 53(8):940–950, 2010.

[6] Jessie YC Chen and Michael J Barnes. Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors*, 54(2):157–174, 2012.

[7] Jessie YC Chen and Michael J Barnes. Supervisory control of multiple robots in dynamic tasking environments. *Ergonomics*, 55(9):1043–1058, 2012.

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

[9] Kevin Chen, Nithin Shrivatsav Srikanth, David Kent, Harish Ravichandar, and Sonia Chernova. Learning hierarchical task networks with preferences from unannotated demonstrations. In *Conference on Robot Learning*, pages 1572–1581. PMLR, 2021.

[10] Douglas Derryberry and Marjorie A Reed. Anxiety-related attentional biases and their regulation by attentional control. *Journal of abnormal psychology*, 111(2):225, 2002.

[11] Hao Ding, Malte Schipper, and Björn Matthias. Collaborative behavior design of industrial robots for multiple human-robot collaboration. In *IEEE ISR 2013*, pages 1–6. IEEE, 2013.

[12] Howard E Egeth and Steven Yantis. Visual attention: Control, representation, and time course. *Annual review of psychology*, 48(1):269–297, 1997.

[13] Mica R Endsley. Situation awareness global assessment technique (sagat). In *Proceedings of the IEEE 1988 national aerospace and electronics conference*, pages 789–795. IEEE, 1988.

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

[15] Tinka RA Giele, Tina Mioch, Mark A Neerincx, and John-Jules Ch Meyer. Dynamic task allocation for human-robot teams. *ICAART (1)*, 1:117–124, 2015.

[16] Matthew Gombolay, Anna Bair, Cindy Huang, and Julie Shah. Computational design of mixed-initiative human–robot teaming that considers human factors: situational awareness, workload, and workflow preferences. *The International journal of robotics research*, 36(5-7):597–617, 2017.

[17] Michael A Goodrich, Morgan Quigley, and Keryl Cosenzo. Task switching and multi-robot teams. In *Multi-Robot Systems. From Swarms to Intelligent Automata Volume III*, pages 185–195. Springer, 2005.

[18] Lars Johannsmeier and Sami Haddadin. A hierarchical human-robot interaction-planning framework for task allocation in collaborative industrial assembly processes. *IEEE Robotics and Automation Letters*, 2(1):41–48, 2016.

[19] Alaa Khamis, Ahmed Hussein, and Ahmed Elmogy. Multi-robot task allocation: A review of the state-of-the-art. *Cooperative robots and sensor networks 2015*, pages 31–51, 2015.

[20] Jack Kolb, Mayank Kishore, Kenneth Shaw, Harish Ravichandar, and Sonia Chernova. Predicting individual human performance in human-robot teaming. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 45–50. IEEE, 2021.

[21] Christopher W Lynn, Ari E Kahn, Nathaniel Nyema, and Danielle S Bassett. Abstract representations of events arise from mental errors in learning and memory. *Nature communications*, 11(1):1–12, 2020.

[22] Steve McGuire, P Michael Furlong, Terry Fong, Christoffer Heckman, Daniel Szafir, Simon J Julier, and Nisar Ahmed. Everybody needs somebody sometimes: Validation of adaptive recovery in robotic space operations. *IEEE Robotics and Automation Letters*, 4(2):1216–1223, 2019.

[23] Brian Mekdeci and Mary L Cummings. Modeling multiple human operators in the supervisory control of heterogeneous unmanned vehicles. In *Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems*, pages 1–8, 2009.

[24] Hauke S Meyerhoff and Frank Papenmeier. Individual differences in visual attention: A short, reliable, open-source, and multilingual test of multiple object tracking in psychopy. *Behavior Research Methods*, 52(6):2556–2566, 2020.

[25] Tamzidul Mina, Shyam Sundar Kannan, Wonse Jo, and Byung-Cheol Min. Adaptive workload allocation for multi-human multi-robot teams for independent and homogeneous tasks. *IEEE Access*, 8:152697–152712, 2020.

[26] James Munkres. Algorithms for the assignment and transportation problems. *Journal of the society for industrial and applied mathematics*, 5(1):32–38, 1957.

[27] Zenon W Pylyshyn and Ron W Storm. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, 3(3):179–197, 1988.

[28] Harish Ravichandar, Kenneth Shaw, and Sonia Chernova. Strata: unified framework for task assignments in large teams of heterogeneous agents. *Autonomous Agents and Multi-Agent Systems*, 34(2):1–25, 2020.

[29] Pegah Rokhforoz, Hamed Kebriaei, and Majid Nili Ahmadabadi. A fair task and time allocation of multiple human operators using decentralized nash bargaining solution. *Computers & Industrial Engineering*, 153:107027, 2021.

[30] Joshua S Rubinstein, David E Meyer, and Jeffrey E Evans. Executive control of cognitive processes in task switching. *Journal of experimental psychology: human perception and performance*, 27(4):763, 2001.

[31] Paul M Salmon, Neville A Stanton, Guy H Walker, Daniel Jenkins, Darshna Ladva, Laura Rafferty, and Mark Young. Measuring situation awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, 39(3):490–500, 2009.

[32] Brian J Scholl and Zenon W Pylyshyn. Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive psychology*, 38(2):259–290, 1999.

[33] Benjamin Schöne, Thomas Gruber, Sebastian Graetz, Martin Bernhof, and Peter Malinowski. Mindful breath awareness meditation facilitates efficiency gains in brain networks: A steady-state visually evoked potentials study. *Scientific reports*, 8(1):1–10, 2018.

[34] Eric H Schumacher, Travis L Seymour, Jennifer M Glass, David E Fencsik, Erick J Lauber, David E Kieras, and David E Meyer. Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological science*, 12(2):101–108, 2001.

[35] Esmaeil Seraj, Letian Chen, and Matthew C Gombolay. A hierarchical coordination framework for joint perception-action tasks in composite robot teams. *IEEE Transactions on Robotics*, 2021.

[36] Panagiota Tsarouchi, George Michalos, Sotiris Makris, Thanasis Athanasatos, Konstantinos Dimoulas, and George Chryssolouris. On a human–robot workplace design and task allocation system. *International Journal of Computer Integrated Manufacturing*, 30(12):1272–1279, 2017.