Predicting Individual Human Performance in Human-Robot Teaming

Jack Kolb
kolb@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

Mayank Kishore
mkishore3@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

Kenneth Shaw
kshaw2@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, PA, USA

Harish Ravichandar
harish.ravichandar@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

Sonia Chernova
chernova@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

ABSTRACT

Humans differ significantly in cognitive traits associated with human-robot teaming. Not utilizing these inherent differences in assigning roles can be detrimental to the team's performance. We developed cognitive tests to quantify two human traits – situational awareness and network conductivity – and are experimenting whether scores from these tests correlate to performance in three typical interactive human-robot tasks. This work is the first to explore linking human cognitive traits to human-robot task performance.

CCS CONCEPTS

• **Human-centered computing** → *HCI theory, concepts and models*; • **Computer systems organization** → Robotics.

KEYWORDS

human-robot interaction, human-robot teaming

ACM Reference Format:
Jack Kolb, Mayank Kishore, Kenneth Shaw, Harish Ravichandar, and Sonia Chernova. 2021. Predicting Individual Human Performance in Human-Robot Teaming. In *Proceedings of HRI '21: ACM/IEEE Int. Conf. Human Robot Interaction (HRI).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Human-robot teaming (HRT) enables groups of humans and autonomous robots to communicate, coordinate, and collaborate together to perform a joint activity. HRT has been studied across a wide range of domains, including search and rescue [5], defense [9], and space exploration [3]. In the context of HRT, the problem of *task assignment* is that of determining which task or role each agent (human or robot) should perform.

HRI, March 8–11, 2021, Virtual

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Prior work on task allocation involving a mix of human and autonomous agents has largely assumed that all human agents within a given category (e.g., soldier, firefighter, rescuer) are interchangeable and can be assigned arbitrarily. However, treating all human operators as identical fails to account for individualized differences in capabilities, skills, or cognitive abilities between operators. For example, prior work has shown that humans varied up to 87.5% in two traits associated with active robot path planning [12]. The resulting task assignment fails to take advantage of the full potential of certain individuals, harming team performance. We hypothesize that developing explicit models of individual human strengths and weaknesses can one day help improve task allocation in complex human-robot teams.

In this work, we seek to develop a set of simple pretests that enable us to model the variations in human cognitive abilities that are pertinent to human-robot interaction, and how such models can help predict an operator's ability to control a team of agents. In particular, we seek to identify a correlation between an operator's performance on simple cognitive tests, and their performance in complex swarm coordination tasks. In the future, we envision this type of predictive model will enable team-level coordination algorithms, such as team composition and task assignment, to optimize human-robot teaming performance.

2 STUDY OVERVIEW

Our objective is to develop a set of cognitive tests that evaluate innate human abilities relevant to human-robot teaming, and to demonstrate that an individual's performance on these pretests has a correlation with their performance on certain human-robot teaming tasks. We hypothesize that some cognitive pretest scores will correlate with the participants' performances on some teaming roles, but not others.

To test our hypothesis we will conduct a within-subject study in which participants first take two cognitive pretests, and then complete three scenarios in a simulated "search and retrieval" mission. We will then evaluate the results to determine whether correlations exist between each pretest and each scenario of the simulation. Correlations would indicate that the pretests are useful in predicting participant performance and therefore in assigning humans to roles in human-robot teams. The simulation roles are designed to mimic commonplace real-world robot control scenarios.

Forty participants will be recruited using Amazon Mechanical Turk and will be fully informed of the study. The study will be carried out remotely using a web browser and an internet connection.

3 COGNITIVE PRETESTS

Prior work has identified a number of cognitive traits that affect a human's ability to control robots. Among the most prominently cited traits are situational awareness [1, 4, 10], prior experience with related tasks [10], understanding of the robots' autonomy [1, 4], and ability to context switch between tasks [1]. However, no prior studies have attempted to find correlations between these cognitive traits and a human's performance in human-robot teaming tasks.

Below, we introduce two cognitive pretests that each seek to estimate mental capabilities related to HRT. The pretests have the following characteristics, which we deemed important to keep pretests generalized and applicable to a variety of human-robot tasks:

- Each pretest is abstract and does not directly mimic a specific HRT task.
- Each pretest seeks to estimate a single human trait or ability.
- There is significant variance in participant performance on a given pretest.

Our first pretest focuses on quantifying situational awareness, and the second on quantifying a user's ability to mentally model networks.

3.1 Situational Awareness Pretest

Situational awareness (SA) is a person's mental model of an environment. In prior work, a number of metrics have been developed to estimate a user's situational awareness [2, 8, 11]. Most widely used is the Situation Awareness Global Assessment Technique (SAGAT), a format where the user is periodically interrupted from a task and asked questions about the task's environment [2].

In this pretest, we have developed an abstract task in which the SAGAT metric is used to quantify the strength of a participant's mental models. In the simulation shown in Figure 1, the user watches 'packages' (represented by shapes) be distributed through an abstracted warehouse network. The participant must keep track of the capacity levels of each warehouse. To evaluate the participant's situational awareness, the warehouse simulation is run for 30 seconds, then is paused and hidden. The participant is asked to identify the capacity level of each warehouse as accurately as possible. The simulation then resumes and this process is repeated five times. The participant is scored by the accuracy of their labeling.

3.2 Network Connectivity Pretest

Prior work in cognitive science has demonstrated that it is possible to effectively model and predict a human's ability to learn abstract structures and relationships between a stochastic sequence of events [6]. Most interestingly, a simple simulated test was shown to be an effective tool to measure this ability in a person. Effective interaction with a team of autonomous robots similarly requires a user to mentally model complex structural information, such as the underlying communication topology, sensing capabilities, and relative influence of each agent on the swarm. We hypothesize that simple cognitive tests, such as the one in [6], will help model and predict an individual's innate ability to model network topologies, and therefore their performance in HRT tasks.

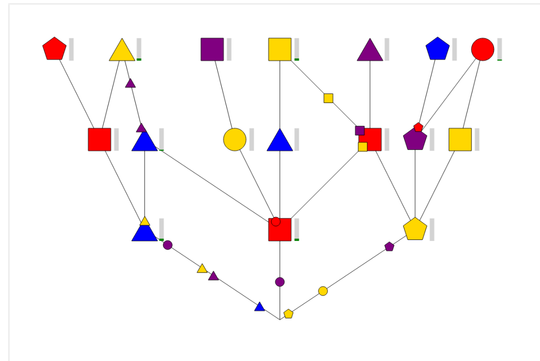

Figure 1: Screenshot of the situational awareness pretest, showing the packages (small shapes) being distributed through the warehouse network (large shapes).

Toward this goal, in our preliminary work we implemented the cognitive test from [6]. However, we found it challenging to apply in practice because it took up to 40 minutes of the participant's time. As a result, we created a different pretest variant, loosely inspired by the prior work. Our pretest, shown in Figure 2, evaluates an individual's ability to efficiently propagate information across a given network, much like a swarm's communication network. The test consists of two phases. In the first phase, the participant observes exemplar runs illustrating how information originating at various nodes propagates to the rest of the network, illustrated by flashing nodes (not pictured). In the second phase, the participant is asked to select the origin node such that information propagates to the rest of the network in the shortest amount of time. This process is repeated for seven networks of differing complexity (Figure 2(a) and (b) are two examples). The underlying connectivity structure of the nodes (dotted lines in the figure) is not shown to the user, thus the operator must learn the underlying structure of the graph, just as in a swarm interaction scenario the operator must maintain a mental model of robot connectivity, line of sight, or other latent features. The participant is scored by the number of hops it takes for information to be propagated to all nodes in the graph.

4 HUMAN-ROBOT TEAMING DOMAIN

To validate participant performance in a human-robot teaming task, we developed a WeBots [7] simulation of a "search and retrieval" operation (Figure 3). In the simulation, participants control unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs) to locate and retrieve several supply caches hidden throughout a 3D environment. To independently validate different human skill sets, we split the task into three scenarios. The scenarios loosely represent three different operator roles that would typically be performed in parallel:

Scenario 1 Search the environment to locate five hidden caches.
Scenario 2 Construct a communications relay network that extends to the caches.
Scenario 3 Leverage the relay network to retrieve the caches.

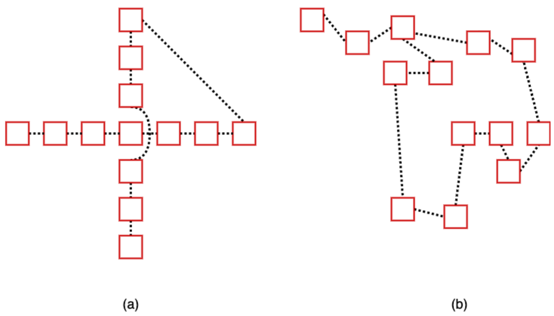

Figure 2: Screenshot of the network connectivity pretest, showing nodes connected by edges (not shown to the user) that determine the order in which information is passed across the hidden network.

Each scenario requires a different skill set and is subject to different constraints. We expect participants to score higher on scenarios that emphasize their inherent traits.

While the scenarios build on each other sequentially (e.g., cache locations from Scenario 1 are used in Scenarios 2 and 3, and the relay network in Scenario 2 is used in Scenario 3), our experimental procedure has been designed to keep them independent. Specifically, each participant will begin each scenario with the same configuration, regardless of their performance or outcome at a previous scenario. Thus, for example, a poorly constructed relay network from Scenario 2 will not affect a participant's performance on Scenario 3.

To reduce score variation due to learning effects, prior to starting the trial simulation participants will given a visual tutorial covering the contents of each scenario, and given a simple demo world in which they can control a UAV, a UGV, and can see the live camera feeds from each robot.

Figure 3: Screenshot of the simulated environment.

4.1 Scenario 1: Target Search and Identification

Participant Capability Being Assessed: Multitasking; context switching between multiple largely independent tasks.

Overview: This scenario focuses on surveillance, a common task in human-robot teaming. The participant has control of four UAVs, each equipped with a downward facing camera. The scenario begins with a blank map, and all UAVs co-located at some pre-determined

	Map Visibility	Network Type	Robot Control	Multi-Robot Coordination
Sc. 1	partial	fully connected	high level	none
Sc. 2	full	ad-hoc relay	high level	high
Sc. 3	full	ad-hoc relay	low level	low

Table 1: Summary of differences between scenarios.

base location. The participant's objective is to use the UAVs to explore the map, to locate each of the five caches in the environment, and to mark the location of each cache on the map as accurately as possible (Figure 4). The participant controls the UAVs by setting waypoints on a 2D overhead map of the area, with the UAVS autonomously navigating between waypoints. In this scenario, we assume unlimited and unrestricted communication between the UAVs and the operator, regardless of distance. To assist the participant's search, circular regions of interest are marked on the map, with one cache guaranteed to be in each region. The scenario completes when all five caches are located.

Metrics: The participant is scored by the time it takes to locate all five caches, and by the accuracy of their marked coordinates.

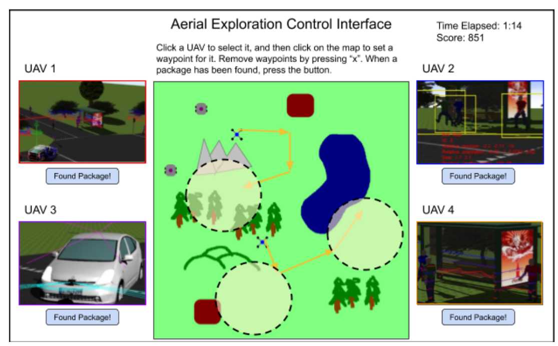

Figure 4: Mockup of the Scenario 1 user interface.

4.2 Scenario 2: Relay Network

User Capability Being Assessed: Situational awareness in the context of high level robot control; network modeling.

Overview: The second scenario focuses on constructing a communication relay, or ad-hoc network, made up of a series of robots – a common task in search and rescue missions. We assume that the operator is given a complete map of the environment, including location of the caches. Additionally, we implement a restriction that both ground and aerial vehicles have a limited communication range. To maintain communication with and control of a robot, the operator must keep the robot within the communication limit of the base station, or another robot connected to the base station. The operator's objective is to arrange available robots in a spatial configuration that forms a relay network reaching all of the caches. Seven UAVs and seven UGVs are used for this task.

As with Scenario 1, the participant controls the robots by setting waypoints on a 2D overhead map. The vehicles autonomously travel between waypoints. If a UAV or UGV exceeds the boundaries of the relay network, it becomes disconnected and stops; the disconnected robot will remain out of contact until reconnected to the network. The operator can retrieve disconnected robots by moving other robots to reestablish the network.

Solving this scenario requires the network to be built gradually from the robot base to the caches. While the UAVs' movements are not affected by obstacles in the environment, the UGVs' are. However, as a tradeoff, the UGVs have a larger relay range than the UAVs due to their greater payload capacity. The user must therefore stay aware of the movements of the robots and the overall efficiency of the relay network. The scenario completes when the relay network extends to cover all five caches and the robot base.

Metrics: The participant is scored by the time it takes to complete the relay network, and the number of "hops" it takes from each relay to the home base.

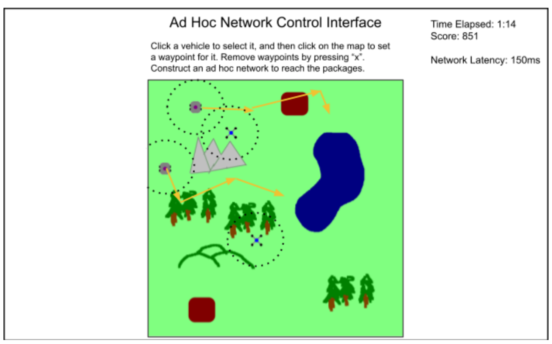

Figure 5: Mockup of the Scenario 2 user interface.

4.3 Scenario 3: Cache Retrieval

Participant Capability Being Assessed: Situational awareness in the context of low-level robot control; context switching between multiple largely independent tasks.

Overview: The final scenario focuses on low level robot control with the objective of cache retrieval. The operator's objective is to retrieve each of the five caches and return them to base. To begin the scenario, the participant is given a full map of the environment labeled with cache locations, and additionally a series of UAVs and UGVs have already been pre-arranged to form a suitable relay network. This configuration approximates the results of Scenario 2, but each participant will be given the same ad-hoc network configuration. The participant will only control the retrieval robots and will not be able to change the network.

In this scenario, we will simulate a lower level of UGV control. While in the prior scenarios UGVs autonomously path planned around obstacles, in this scenario UGVs will always follow the shortest direct path to the next waypoint. As a result, the operator will be required to closely supervise each robot and manually arrange waypoints to avoid impassable terrain. The operator must also keep each robot within the boundaries of the relay network; a UGV that exceeds the boundaries of the relay network will become disconnected and "lost". When a UGV approaches a cache, the user will be able to retrieve it with the press of a button. The scenario ends when all caches have been collected and all non-disconnected robots return to the base. A total of seven robots will be provided to retrieve five caches.

Metrics: The participant is evaluated by whether they successfully completed the scenario, how many robots were lost, and the time taken to retrieve the caches.

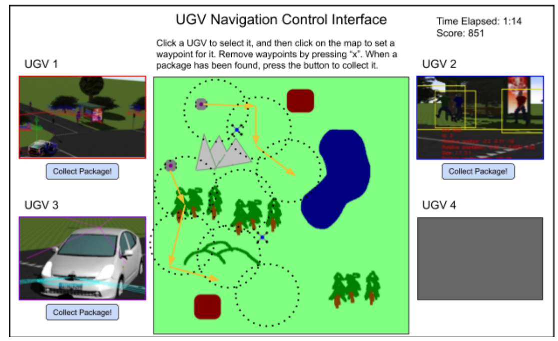

Figure 6: Mockup of the Scenario 3 user interface.

5 DATA ANALYSIS

Once user study data is collected, we will plot each of the six pretest and simulation scenario pairings and calculate the Pearson correlation coefficient for each pairing. A pairing with a high correlations will indicate a possible link between performance in that pretest and performance in that general task domain. Correlated pairings could then potentially be used to assign humans to roles in similar environments to maximize mission performance.

ACKNOWLEDGMENTS

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

REFERENCES

[1] 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.
[2] Mica R Endsley. 1988. Situation awareness global assessment technique (SAGAT). In *Proc. of IEEE national aerospace and electronics conference*. IEEE, 789–795.
[3] Terrence Fong and Illah Nourbakhsh. 2005. Interaction challenges in human-robot space exploration. *Interactions* 12, 2 (2005), 42–45.
[4] Caroline E Harriott, Adriane E Seiffert, Sean T Hayes, and Julie A Adams. 2014. Biologically-inspired human-swarm interaction metrics. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 58. SAGE Publications Sage CA: Los Angeles, CA, 1471–1475.
[5] Stefan Kohlbrecher, Alberto Romay, Alexander Stumpf, Anant Gupta, Oskar Von Stryk, Felipe Bacim, Doug A Bowman, Alex Goins, Ravi Balasubramanian, and David C Conner. 2015. Human-robot teaming for rescue missions: Team ViGIR's approach to the 2013 DARPA Robotics Challenge Trials. *Journal of Field Robotics* 32, 3 (2015), 352–377.
[6] Christopher W Lynn, Ari E Kahn, Nathaniel Nyema, and Danielle S Bassett. 2020. Abstract representations of events arise from mental errors in learning and memory. *Nature communications* 11, 1 (2020), 1–12.
[7] Olivier Michel. 2004. Cyberbotics Ltd. Webots™: professional mobile robot simulation. *International Journal of Advanced Robotic Systems* 1, 1 (2004), 5.
[8] Lucas Paletta, Amir Dini, Cornelia Murko, Saeed Yahyanejad, Michael Schwarz, Gerald Lodron, Stefan Ladstätter, Gerhard Paar, and Rosemarie Velik. 2017. Towards real-time probabilistic evaluation of situation awareness from human gaze in human-robot interaction. In *Proc. of Human-Robot Interaction*. 247–248.
[9] Raja Parasuraman, Michael Barnes, Keryl Cosenzo, and Sandeep Mulgund. 2007. Adaptive automation for human-robot teaming in future command and control systems.

[10] Sameera Ponda, Han-Lim Choi, and Jonathan How. 2010. Predictive planning for heterogeneous human-robot teams. In *AIAA Infotech@ Aerospace 2010*. 3349.

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

[12] Christopher J Shannon, David C Horney, Kimberly F Jackson, and Jonathan P How. 2017. Human-autonomy teaming using flexible human performance models: An initial pilot study. (2017), 211–224.