Preliminary Interactions of Human-Robot Trust, Cognitive Load, and Robot Intelligence Levels in a Competitive Game

Extended Abstract

Michael Novitzky
Massachusetts Institute of Technology
Cambridge, Massachusetts
novitzky@mit.edu

Paul Robinette
Massachusetts Institute of Technology
Cambridge, Massachusetts
paulrobi@mit.edu

Michael R. Benjamin
Massachusetts Institute of Technology
Cambridge, Massachusetts
mikerb@mit.edu

Danielle K. Gleason
Massachusetts Institute of Technology
Cambridge, Massachusetts
dgleason@mit.edu

Caileigh Fitzgerald
Massachusetts Institute of Technology
Cambridge, Massachusetts
caileigh@mit.edu

Henrik Schmidt
Massachusetts Institute of Technology
Cambridge, Massachusetts
henrik@mit.edu

ABSTRACT

This paper presents a pilot study in which we examine the interactions between human-robot teammate trust, cognitive load, and robot intelligence levels. In particular, we attempt to assess these interactions during a competitive game of capture the flag played between a human and a robot. We present results while the human plays against robots of different intelligence levels and determines their level of trust of each robot as a potential teammate through a post experiment questionnaire. We also present our exploration of heart rate measures as approximations of cognitive load. It is our goal to determine guidelines for future autonomy and interaction designers such that their systems will reduce cognitive load and increase the level of trust in robot teammates. This is an initial experiment that uses the least amount of vehicles yet still gathers competitive data on the water. Future experiments will increase in complexity to many opponents and many teammates.

ACM Reference Format:
Michael Novitzky, Paul Robinette, Michael R. Benjamin, Danielle K. Gleason, Caileigh Fitzgerald, and Henrik Schmidt. 2018. Preliminary Interactions of Human-Robot Trust, Cognitive Load, and Robot Intelligence Levels in a Competitive Game: Extended Abstract. In *HRI '18 Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion, March 5–8, 2018, Chicago, IL, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3173386.3177000

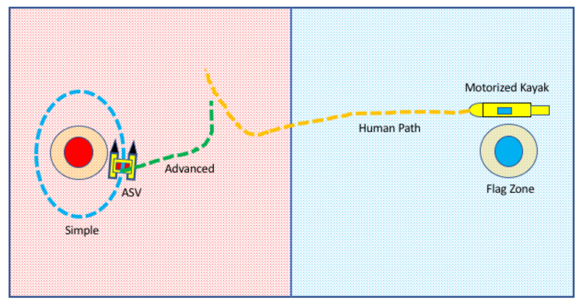

Figure 1: The capture the flag field. The human operated motorized kayak starts at the blue flag while the autonomous surface vehicle (ASV) defends the red flag. The dashed lines represent vehicle paths. The orange line depicts the human vehicle's trajectory. The ASV defense behaviors are colored blue for the *Simple* (only circles the flag – less capable) and green for the *Advanced* (intercepts and follows the opponent – more capable).

1 INTRODUCTION AND RELATED WORK

Modern systems are combining manned vehicles with autonomous vehicles to perform tasks in challenging environments. For example, the U.S. Air Force's "Loyal Wingman" project is exploring manned-unmanned teaming in which an unmanned aerial system (UAS) and a manned aircraft work directly on missions such as air interdiction, attack on integrated air defense systems, and offensive counter air [2]. In this work, we explore a similar manned-unmanned teaming concept in the marine domain. The marine domain is more accessible for deploying autonomous vessels (no approval is required from government agencies and vehicles can be easily stopped on the water) and yet still challenging given the elements in the environment. Our manned vessel is a motorized kayak and our autonomous teammate is an autonomous surface vehicle (ASV). We explore the interplay between operator load, robot autonomy, and human-robot teammate trust. Solovey et al. [3] classified driver workload using physiological measurements (heart rate and skin conductance level), cognitive tasks (n-back tests) and driving performance data (driving speed, steering wheel position, acceleration data). Similar to Solovey et al. [3], we explore measuring cognitive load through heart rate data. We use a situational definition of trust from [4]. In short, we define trust as the belief that a trustee will mitigate a trustor's risk in a particular situation. This paper describes our pilot experiments as part of Project Aquaticus in which we measure human-robot teammate trust after the human has competed against the robot in a half-court game of capture the flag. This is akin to sports teams that have their offense and defense play against each other during practice in half-court games. Our hypothesis is that a human's trust

HRI '18 Companion, March 5–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-5615-2/18/03.
https://doi.org/10.1145/3173386.3177000

for a robot teammate will vary based on the competitiveness of the robot. We vary the robot's capability and have the human compete against each variation. We also present preliminary results of heart rate and heart rate variability which we plan to use as a measure of cognitive load. These are our initial experiments in Project Aquaticus - we used the fewest number of vehicles possible while still gathering competitive data on the water. The ultimate goal of the project is to have two humans and two robots on the same team competing in a game of capture the flag against a similarly situated team. It is our goal to provide lessons learned from our platform in the marine domain to other challenging environments.

2 EXPERIMENTAL SETUP

The human vehicle is a Mokai ES-Kape motorized kayak which has been augmented with a semi-rugged laptop, compass, GPS, long range WiFi antenna, and a headset for voice interaction with the game. The ASV is a Clearpath Robotics Heron M300. The autonomy for the ASV is provided by MOOS-IvP [1]. MOOS is a robot middleware and the autonomy is provided by the IvP Helm behavior-based decision engine architecture.

The capture the flag game mechanics we have implemented are based on scoring the most points in ten minutes. In order to score a point, an opponent's virtual flag must be grabbed from their side of the field and returned to one's home flag zone. Players can defend their side of the field through the use of tags. A tagged player must untag themselves by returning to their home flag zone.

In this experiment, the task for the ASV is to defend its flag by tagging opponents. For both intelligence levels, the ASV will tag the opponent if it is within 10 meters. As seen in Figure 1, in the *Simple* defense behavior, seen in dashed blue line, the ASV traverses a loiter pattern circling around its flag without consideration of the opponent's actions - this is considered to be less capable and less competitive. The *Advanced* defense behavior, seen as a dashed green line, is much more aggressive as the ASV will intercept and chase the opponent– this is considered a much more capable and more competitive behavior.

The participant plays a half court game against both the *Simple* and *Advanced* defense behaviors but their order is randomized. The participant is told that they are being measured on how many points they can score against the robot in a ten minute round. During each trial, the performance of each vehicle is logged in terms of each vehicle's track data and game state information. Heart rate (HR) and heart rate variability (HRV) is recorded by a Zephyr Bioharness monitor. After each ten minute round, the participant fills out a questionnaire. After both rounds are completed, a final questionnaire is completed along with post experiment debriefing. The key question in this experiment is "If you are not allowed to defend your flag and you only are allowed one robot: Which defense behavior would you choose to defend your flag: A, B, Neither, No preference between A and B" Followed by asking for a rating of "How willing are you to risk your flag to behavior A defending your flag?"

3 PRELIMINARY RESULTS

The two participants in this pilot study overwhelmingly would trust the *Advanced* defense behavior over the *Simple* defense behavior.

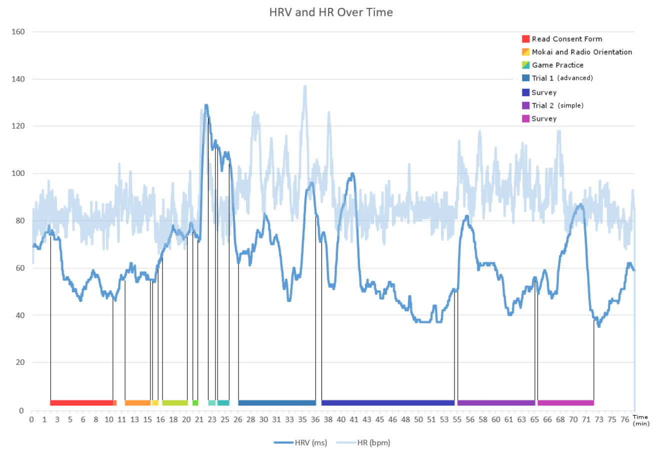

Figure 2: A participant's heart rate (HR in bpm) and heart rate variability (HRV in ms) plotted by timeline and task.

They both chose to risk the role of defending their flag to the *Advanced* defense behavior over the *Simple* behavior and rated their trust of the *Advanced* behavior higher than the *Simple* behavior (6 vs 2 and 6 vs 1 out of a 7 point scale). In Figure 2, we can see a participant's heart rate measurements plotted over the duration of the experiment. It is expected that the heart rate variability would be lower while the participant has higher cognitive load, which can be seen during the first survey. We will wait for a larger sample size before drawing conclusions regarding the cognitive load of the participants while competing against the different intelligence levels of the ASV.

Presented here are preliminary results of a pilot study in which we are measuring the interactions of human-robot teammate trust, cognitive load, and robot intelligence. This is an initial experiment that used the least amount of vehicles to gather competitive data on the water. We will continue this pilot study to recruit a larger number of participants and more complicated interactions.

ACKNOWLEDGMENTS

This work has received funding from ONR grant N00014-17-1-2570 under Dr. Behzad Kamgar-Parsi, DARPA TTO under Dr. Daniel Patt and Dr. Timothy Chung, and Battelle under Michael Mellott. We thank the MIT Sailing Pavilion staff for their professionalism and support in designing and maintaining our safety protocol. We thank Nick DePalma and Alan Wagner for their feedback and guidance.

REFERENCES

[1] Michael R. Benjamin, Henrik Schmidt, Paul Newman, and John J. Leonard. 2010. Nested autonomy for unmanned marine vehicles with MOOS-IvP. *Journal of Field Robotics* 27, 6 (2010), 834–875.

[2] Kirsten Kearns. 2015. RFI: Autonomy for Loyal Wingman. *Air Force Research Laboratory (AFRL)* (15 July 2015).

[3] Erin T Solovey, Marin Zec, Enrique Abdon Garcia Perez, Bryan Reimer, and Bruce Mehler. 2014. Classifying driver workload using physiological and driving performance data: two field studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 4057–4066.

[4] Alan R Wagner and Ronald C Arkin. 2011. Recognizing situations that demand trust. In *RO-MAN, 2011 IEEE*. IEEE, 7–14.