The Role of Psychophysiological Measures as Implicit Communication Within Mixed-Initiative Teams

Kim Drnec[1]([✉]), Greg Gremillion[2], Daniel Donavanik[2], Jonroy D. Canady[1],
Corey Atwater[3], Evan Carter[1], Ben A. Haynes[4], Amar R. Marathe[1],
and Jason S. Metcalfe[1]

[1] Army Research Laboratory, Aberdeen Proving Ground, Aberdeen, MD, USA
kim.a.drnec2.ctr@mail.mil,
{jonroy.d.canady.civ,evan.c.carter3.civ,
amar.r.marathe.civ,jason.s.metcalfe2.civ}@mail.mil
[2] Adelphi Laboratory Center, Adelphi, MD, USA
gregory.m.gremillion.civ@mail.mil
[3] DCSCorp, Warren, MI, USA
catwater@dcscorp.com
[4] U.S. Army Tank Automotive Research Development and Engineering Center,
Warren, MI, USA
ben.a.haynes2.civ@mail.mil

Abstract. There has been considerable effort, particularly in the military, at integrating automated agents into human teams. Currently, automated agents lack the ability to intelligently adapt to a dynamic operational environment, which results in them acting as tools rather than teammates. Rapidly advancing technology is enabling the development of autonomous agents that are able to actively make team-oriented decisions meaning truly intelligent autonomous agents are on the horizon. This makes the understanding of what is important to team performance a critical goal. In human teams, mission success depends on the development of a shared mental models and situation awareness. Development of these constructs requires good intra-team communication. However, establishing effective intra-team communication in a mixed-initiative team represents a current bottleneck in achieving successful teams. There has been significant research aimed at identifying modes of communication that can be used both by human and agent teammates, but often neglects a source of communication or information for the agent teammate that has been adopted by the human robot community to increase robot acceptance. Specifically, the use of psychophysiological features supplied to the agent that can then use algorithms to infer the cognitive state of the human teammate. The utility of using psychophysiological features for communication within teams has not been widely explored yet representing a knowledge gap in developing mixed-initiative teams. We designed an experimental paradigm that created an integrated human-automation team where psychophysiological data was collected and analyzed in real-time to address this knowledge gap. We briefly present a general background to human automation teaming before presenting our research and preliminary analysis.

1 Introduction

Automated agents, regardless of their instantiation, i.e., a computer or a robot, have become increasingly adopted by both the military and society in general. In the case of the military, automated agents are generally employed to carry out tasks that are too difficult or too dangerous for a Soldier [1]. For example, the MARCbot is used by the military to investigate suspicious objects, and can even look behind doors because of the placement of its camera [1]. Alternatively, automated agents in society are usually used for tasks that are repetitive or that require time-sensitive and complex computations. In both domains, however, the agent, currently, is largely used as a tool [1] to extend the human operator's functional 'reach', or to improve performance of a task, or both, as opposed to a teammate. Moreover, despite the near ubiquitous presence of automated agents, joint system (i.e., the human operator and the agent) performance has failed to reach desired expectations [2]. While many reasons for this failure have been forwarded [3–5], one of the most important identified by the Department of Defense [9] is that the human operator normally assumes a supervisory role. This role often places the human operator 'out of the loop' [6–8] resulting in reduced operator situation awareness of agent function and degree of task completion [9]. This supervisory role, which might include tasks such as monitoring a display has been necessary due to the inherent limitations of automated agents; they have not been capable of autonomous behavior, but rather they have often replaced a human activity in a brittle manner.

Recently, two key developments have occurred, largely in the field of robotics, that are likely to change the way that humans and automated agents interact to achieve a shared goal in fully collaborative situations. First, advances in artificial intelligence (AI) are rapidly providing automated agents with the ability to flexibly adapt behavior based on operator needs, or changing task demands. Although these abilities such as learning about a human partner or an environment are yet still relatively primitive, this technical achievement in AI enables truly autonomous agents, capable of making team-oriented decisions [10]. Secondly, there is emphasis being placed on fully integrating autonomous agents and humans, particularly in the military domain [9]. This full integration of agent and human results in mixed-initiative teams that will require genuine human-agent collaboration. These developments will essentially move automated agents from tools to teammates [10–12]. The purpose of this paper is twofold. First we broadly discuss the advances in theory and technology enabling the formation of successful mixed-initiative teams, focusing on intra-team communication. We then highlight a critical gap that is, as yet to be fully-addressed. That is, we focus both our literature review and research on how to improve intra-team communication through the use of implicit human state information derived from psychophysiological signals. Second, we describe our research that is aimed at filling that gap, and finally, we present very preliminary results from several individuals. In the next section, we enter our discussion

by carefully delineating a line of thought that arrives at the concept of better human-agent teaming through shared mental model development as augmented by the application of physiology-based monitoring and prediction of human decisions during a joint human-automation driving task.

2 Moving Agents from Tool to Teammate Within Mixed-Initiative Teams

Within the human agent teaming (HAT) literature, the terms robot, autonomy, and automation are seldom defined and are occasionally conflated. For example, the Defense Science Institute in 2016 [9] held a workshop on bi-directional communication for human robot teaming, but within the report, autonomy is used both as an adjective to describe robot capabilities, and as a synonym for robots [9]. To avoid confusion in the following discussion we consider the space of (HAT) as being represented by the Venn diagram in Fig. 1. Further, we define the term automation to be any machine or software that performs an automated function that generally lacks the capability of independent action. We define an autonomy as an intelligent agent that may be capable of independent action, i.e., performing actions not previously hard coded. Therefore, a robot might be either an automation or an autonomy. However, as technology improves and increasingly sophisticated autonomies appear on the horizon, there will inevitably be some grey area between automation and autonomy. We use the term 'agent' to designate a robot, an autonomy, or an automation. Finally, we define a team as 'a group of two or more people who interact dynamically, interdependently, and adaptively towards a common value or goal' [13]. We use the term mixed-initiative teams as used in the literature [14], to designate a team comprising at least one human and one agent.
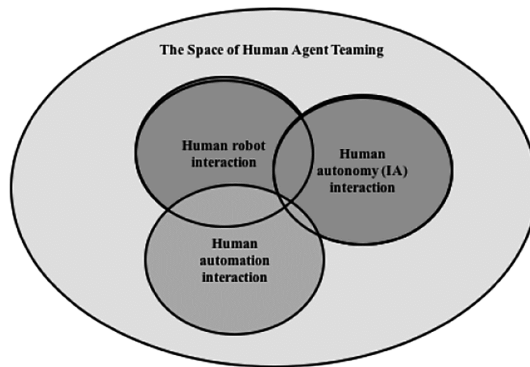

Fig. 1. The conceptual space of human agent teaming as discussed herein

Literature involving human only teams posits that successful mission completion heavily relies on the development of two main constructs. Teams need to have a shared mental model of the problem space, and shared situation awareness of evolving environments [10, 13, 15]. Mental models held by team members reflect an understanding

of teammate intent during problem solving such that each team member can infer the causes of a teammate's behavior [16], and act symbiotically to achieve a goal. Shared situation awareness includes a common representation of the dynamically changing environment or problem. In human-human teams, these constructs are continuously being updated [12], and therefore it stands to reason that the ability for an agent teammate to update its mental model and situation awareness will be critical to successful mixed-initiative teams. Research efforts entailing mixed-initiative teams must focus on engendering these constructs effectively within the team. Therefore, team dynamics as a whole need to be directly addressed as opposed to the traditional model, where focus is usually directed toward improving human *or* agent performance [17], essentially treating teammates as independent actors rather than interdependent entities.

Effective communication between teammates, whether explicit or implicit, is the cornerstone of establishing accurate mental models and shared situation awareness within a team [10, 15, 17, 18]. Within this team construct, it is implied that the communication be bi-directional between the team members in order to be effective. For mixed-initiative teams, the requirement for bi-directional communication is even more important if an agent is to move from being a tool to a teammate [19]. However, this ability to have bi-directional communication, a key for team success, represents a bottleneck in furthering mixed-initiative team performance [15]. There are, therefore, considerable research efforts aimed at identifying effective methods of bi-directional communication for use in mixed initiative teams [11, 19–22]. This research must explore modes of communication that consider the nature and characteristics of each teammate, as well as any constraints imposed by the current operational environments.

Multimodal communication (MMC) has been identified as one method to facilitate successful bi-directional communication [18, 20]. MMC is communication using more than one modality (voice, gestures, etc.) [20]. The reported flexibility of MMC means that communication can be accomplished intuitively and efficiently to best suit the individual situation [20]. For example, gesturing allows for silent communication, and haptic communication [20] allows team members to communicate even when they are out of the line of sight of each other. MMC, however, while flexible, and intuitive to human teammates, explicitly centers on the five human senses [20]. This five-sense-centric model of bi-directional team communication ignores signals that could be readily interpreted by most agents, namely, psychophysiological signals, which can be acquired and processed online using contemporary signal processing tools [23].

Psychophysiological signals have been successfully used to infer affective state during human robot interactions. For example, it is known that rapid movements by robots tend to induce a degree of anxiety in humans [24, 25]. However, if a robot teammate identifies that the human is in a state of anxiety, and adjusts its movements appropriately, anxiety is reduced [24]. Similarly, inferences based on psychophysiological signals can allow the robot to dynamically alter its behavior. For instance, it can change its level of autonomy, reallocate tasks, and infer when, and how, to query the human teammate [21, 26]. A predominant of work in this domain, with notable exceptions [12, 15, 21], appears to be largely focused on two goals; either a robot *replacing* a human activity or towards making the robot more acceptable to the user, which engenders trust, long considered necessary for successful human agent teaming [27, 28]. However, little

research has been directed at explicitly exploring the utility of incorporating implicit psychophysiological signals as a form of communication into mixed-initiative teams. Our research aims to address this gap.

For our experiment, we developed a research paradigm that has some of what we believe to be critical elements of a mixed-initiative team, even if not at the full scale of emerging mixed initiative teams that are of ultimate interest. The elements included in our research paradigm consisted of a participant, a simulated vehicle, an optional driving automation and a controller. All these elements were fully integrated into a control theoretic framework that approximated a team. We provided psychophysiological communication to the controller in real time with the intent that the controller use this implicit communication as data to predict driver behavior. If participant driver behavior was predicted to be sub-optimal the control system could suggest, through a visual display, what behavior might be more beneficial for successful task completion. We had three expectations that would approximate metrics of the utility of including psycho-physiological communication to the agent. First, that in conditions with integrated psychophysiological communication, joint system task performance would be better than without this implicit form of communication. Second, that participant trust would be highest in conditions including psychophysiological communication to the controller. Finally, subjective participant workload measures would be reduced in conditions with the psychophysiological communication as compared to conditions without that communication.

3 Materials and Methods

3.1 Overview

We based this research on a leader-follower simulated driving paradigm, well explored by our lab [29], involving a semi-automated driving automation capable of lane and speed maintenance, but not collision avoidance. We contend that this is a natural context for observation of human behaviors in the presence of an automated agent, particularly as driving is a common task familiar to most adults. In addition to providing the participant with a static automation, i.e., non-adaptable automation *vis a vis* capabilities, we added a control system designed to help the participant make appropriate decisions about what agent (human or automation) should be in control of the simulated vehicle at a point in time given relative agent capabilities. Individual participant capabilities were sampled by the control system at the outset of the experiment, essentially establishing a type of 'ground truth' for participant ability, while the control system had *a priori* knowledge of the automation's performance. In our paradigm, the control system functioned as an advanced automation, capable of assessing participant ability and identifying relevant environmental features such as a tight turn. The control system could then integrate that information with psychophysiological signals to produce team-oriented probabilistic decisions about what agent should be in control of the simulated vehicle. Further, the control system, was designed such that it could 'act' on the outcome of the calculation. It accomplished this by adjusting a visual display (described later), from herein termed the actuator, where the term actuator is defined as the mechanism by which

the controller acted on its environment. The actuator indicated to the participant what agent the system believed was likely to perform better given the previously calculated 'ground truth', and the current environment. The controller utilized one of three different algorithms, depending on experimental condition, to control the behavior of the actuator. One of these conditions integrated psychophysiological features.

3.2   Facilities and Equipment

The equipment and methods are nearly identical to that of previous research and a detailed description of the paradigm is given by Metcalfe et al. (2017). The current research, described below, was conducted in a sound proofed experimental chamber at the Cognitive Assessment Simulation and Engineering Laboratory, Aberdeen Proving Ground (APG). The simulation was realized by three large monitors on a desktop that displayed the simulated environment provided by SimCreator (Realtime Technologies, Inc; Royal Oak, MI). Participants sat in a chair that resembled an actual driving seat. During the experimental tasks participants were outfitted with a suite of sensors for recording psychophysiological activity. Specifically, participants donned a Biosemi ActiveTwo (BioSemi BV; Amsterdam, Holland) system to enable recording electroencephalographic brain activity (EEG), electrocardiographic activity (ECG), electrooculargraph activity reflecting eye motion (EOG), and skin electrical conductance (electrodermal activity; EDA).

Data collected, processed and analyzed in real-time included psychophysiological (EEG, EDA, and vertical and horizontal eye movements (vEOG, hEOG)), behavioral (e.g. automation use, brake and throttle inputs), and variables related to the participant's vehicle (e.g. heading error, speed etc.). In addition to the psychophysiological data collected, participants completed a set of surveys including the Big Five Inventory for assessing personality traits [31], a demographic questionnaire, and after each condition surveys appropriate to the condition. After all conditions participants completed a NASA-TLX [32] to assess the level of subjective workload, and a simulator sickness questionnaire to assess motion sickness. Further, in conditions where the automation was available participants completed a system trustworthiness and display trustworthiness surveys. Behavioral and participant vehicle data were collected at 60 Hz and recorded by SimCreator.

3.3   Experimental Task and Design

In this leader follower task participants drove one and one-half laps around a two-lane simulated course with ambient traffic (Fig. 2 bottom). Task objectives included maintaining lane position, a "safe" distance from the lead vehicle, and avoiding collisions with ambient traffic and frequently appearing pedestrians. The automation could control speed and lane position, but had no collision avoidance capabilities. Pedestrians appeared approximately every 7 s, distributed randomly on either side of the road, and 15% stepped into the vehicle path. Participants were asked to respond to pedestrians using buttons on a game controller, which removed the pedestrian from the road, thereby avoiding potential collisions. For conditions in which the driving automation was

available, participants had the option to enable or disable it at any moment. It could be disabled through application of the brake, throttle inputs, or an accelerator-pedal- adjacent toggle foot switch, and enabled by depressing the same toggle foot switch. Lateral and longitudinal perturbations were introduced to add additional challenge to the driving task. The lateral perturbations simulated gusts of wind that tended to push the participant vehicle out of the lane increasing the risk of collisions with ambient traffic and pedestrians. Longitudinal perturbations were implemented by increasing or decreasing the speed of the lead vehicle. Participants were provided with a visual display screen (Fig. 2 top) denoting current score (described later), consequence zones (described later), control agent (human or automation) indicated with a green chevron (Fig. 2 top image), and in three of the five conditions, a functioning actuator display Fig. 2, top image, black bar at the top of screen; described in detail later).
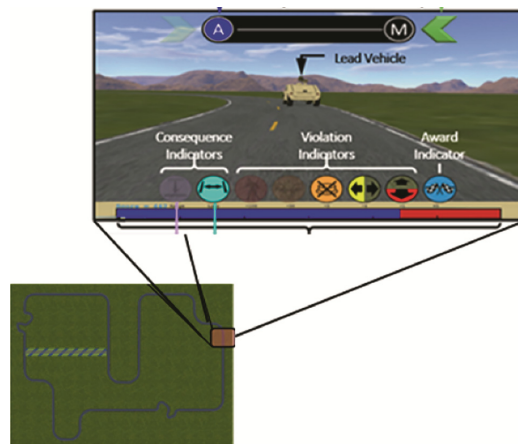

Fig. 2. At the top is a depiction of the visual display provided for the participants. At the top of the display is an indicator of what agent is in control. Here, the chevron points to 'M' (manual) so the participant is in control. At the bottom of the display is a sliding bar depicting point losses. The bottom Fig. illustrates the course the participants were to navigate.

The average drive time around the course (Fig. 2 bottom image) for each condition lasted approximately 18 min and consisted of 19 zones defined by environmental feature. For example, a section of course with an s-curve would be one zone, whereas another section of the course with a straight-away would constitute another zone. All zones were approximately equal in length. The purpose of the zones is to create circumstances wherein one agent would be preferred over the other in terms of performance. For instance, humans are generally superior at handling s-curves, whereas the driving automation was superior at handling straight-aways. In different conditions, zones would be assigned changing consequence designations for either lane or range violations and the task with an increased consequence was indicated by the visual display. For example, a zone might be designated high consequence for lane deviation meaning that, in that zone, penalties for lane deviations increased. There were four different sets of zone consequences; one for the first and last conditions, and three counterbalanced sequences for

the other three conditions. For each experimental condition participants started with 500 points and lost points for each performance decrement. The point loss was proportional to real-life consequences, meaning that the most points were deducted for collision with a pedestrian, and the fewest points were deducted for lane and range violations. For zones denoted by increased consequence, point loss increased for that zone for the indicated performance decrement. Points remaining at the end of each condition were translated into monetary compensation with a simple algorithm. Point losses and current score were indicated by a sliding color bar (Fig. 2) at the bottom of the display screen.

Fig. 3. Informational icons provided in the display to the participant. At the top (A) are icons that represent possible performance decrements. These are greyed out (B) when there is no error, and are lit when the error occurs. In Fig. 3 B, illuminated icons indicate that there has been a collision with a pedestrian and that the wrong button has been pressed to clear the pedestrian from the road. At the far right are the consequence indicators. Here, in the bottom row (B) the lane deviation icon is illuminated signifying that for the current zone there is an increase in point losses for this error. Only one row of icons is displayed during the experiment; Fig. 3 A where all icons are illuminated is for illustrative purposes only.

Each experimental session consisted of five unique conditions; three conditions determined by the algorithm the controller used to control the behavior of the actuator, one manual condition, and one condition with no actuator present, but with the automation available. The first condition was always a 'manual' condition where the driving automation was not available, and the actuator did not function, although the actuator display indicated with a green chevron that the participant was in control of the vehicle. This condition served two important purposes. First, it is important for a participant to be able to gauge their ability to perform the task without aid from an automation. This is because experience indicates that whether a person uses an automation is highly dependent on their perceived ability to succeed at a task themselves versus their perception of an automation's ability to perform the same task [33, 34]. Second, the control system, to be able to generate relative probabilities of success for each agent for each zone, needed to have an accurate assessment of the participant's innate ability to perform the task relative to the known ability of the automation.

The other four conditions were as follows; no actuator (NA), unscaled actuator (UA), scaled actuator (CA), and psychophysiologically scaled actuator (PA). These actuator displays are shown in Fig. 4, and the method of counterbalancing conditions and consequence sequences is described below. The NA condition was a condition where the automation was available, but no indication was provided to the participant regarding probable success of either agent. The UA actuator algorithm determined the objective probability of each agent to succeed at the task in a zone. The CA actuator algorithm

considered not only each agent's abilities, but the consequence designation of the zone. For example, if a zone with a tight s-curve was the current zone and it was designated as an increased consequence for range deviation the control system needed to consider that fact. The automation is generally better at range maintenance, and the human driver is generally superior at maneuvering tight curves. Therefore, if an increase consequence for range deviation in a tight curve was indicated the system might still suggest that the automation be in control instead of the participant. When active, the actuator indicated the preferred agent by presenting a colored ball that moved smoothly from 'A' to 'M' (Fig. 4) as the outputs of the algorithm were derived. The ball changed colors depending on the recommended agent. If the automation was recommended the ball turned blue as it moved towards the 'A'. Conversely, as the ball moved toward the 'M', suggesting that manual control was preferred, it turned red. It should be noted, however, that the actuator display in the UA, CA, and PA conditions appeared identical to the participant.

Fig. 4. Examples of the actuator display for CA, UA and PA conditions which were visually identical, differing only in the algorithm used to control their behavior. In the top left the automation is in control and the output of the control algorithm is leaning towards automatic control. In the top, right the automation is also in control, but the controller is leaning towards manual control. In the bottom left the automation is in control, and the controller is suggesting that this state be continued. In the bottom, right, the participant is in control and the controller is suggesting that they remain in control. (Color figure online)

The PA condition was introduced as an explicit test of the utility of providing implicit communication to an agent in the form of psychophysiological signals from the participant. The control algorithm that drove the actuator behavior in the PA condition considered the objective probability of the success for both agents, as in the UA condition. However, in the PA condition, the control algorithm also factored in inferences made from psychophysiological features about the likelihood of the participant activating or deactivating the automation at any given time. This ability was made feasible by previous work by our lab which demonstrated that a machine classifier, taking specific environmental and psychophysiological features as inputs, could predict when a participant was about to make a change in control authority with almost 70% accuracy. This classifier was implemented in real time in the control system, and the output was included in the algorithm driving the behavior of the actuator. In the PA condition the probability of one agent being in control versus the other was added to the objective probability of success for each agent in a principled way based on probability theory.

The purpose of the experiment was twofold, and one of the aims partially dictated experimental design. The primary purpose was to assess the utility of adding implicit communication through psychophysiological signals to the controller as measured by overall performance of the task. The second purpose was to act as a continuation of a

previous experiment and essentially increase overall subject numbers. The earlier experiment which included the manual, CA, UA and NA conditions was conducted at the U.S. Army's Tank and Automotive Research Development and Engineering Center (TARDEC) in Michigan. Therefore, the CA, UA, and NA conditions that occurred in both experiments were always counterbalanced along with their three possible consequence sequences using a Latin Square, creating a 3×3 factorial design. For execution of the experiment described in this paper, the manual condition was always first for reasons described earlier. To allow for inferences regarding the effects of integrating psychophysiological signals into the control system, half of the participants (n=4) experienced the PA condition immediately after the manual condition and half (n=5) experienced it as the last of the five conditions.

4   Preliminary Results

Preliminary results are based on data available from an initial sample of 5 participants, although we have plans to collect data from a total of 18 participants. We originally considered that performance and trust (trust of the actuator display) would be highest in the PA condition. The inclusion of performance based metrics in our analysis has obvious motivation. If the metrics reflected high levels of performance, it might be inferred that the team in that condition was successful as compared to other conditions. We used points lost per second as a metric of performance (Fig. 5A) as a way of accounting for varying lengths of time on the course. Trust was considered as a relevant variable because of its presumed importance as to how an operator decides to use an automation. It is widely considered that if there is an appropriate level of trust (not too high and not too low) the automation will also be used appropriately. Figure 5B shows the mean trustworthiness of the display scores in the three conditions where an actuator display was available. We also expected that workload (Fig. 5C) would be reduced in the PA condition as compared to all five conditions. There was no statistical difference in performance, trust, or workload across the conditions, but there are tendencies seen on visual inspection of Fig. 5 that if they remain when a larger data set is analyzed, would make interesting results. For example, it appears that performance was improved overall by having the automation available (Fig. 5A) as compared with the manual condition. Further, trust (Fig. 5B) appears to be highest in the PA condition with a very narrow distribution, indicating that most participants felt they could trust the actuator display. Interestingly, workload also appears to be lowest in conditions where an actuator was available (UA, CA, PA), which might suggest that participants found the task easier when they were given cues as to which agent should be in control.
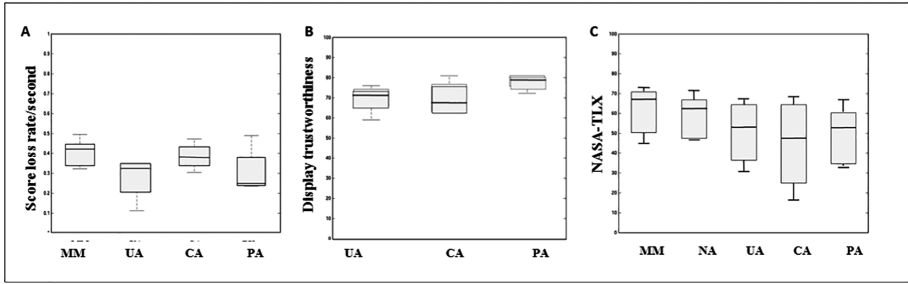
Fig. 5. (A) Score loss per second, (B) actuator display trustworthiness, (C) workload (unscaled NASA-TLX). MM (manual condition), UA (unscaled actuator), CA (scaled actuator), NA (no actuator), PA (psychophysiological actuator). The black line in the bars represents the mean value.

5   Discussion

We designed this experiment to determine if integrating implicit communication in the form of psychophysiological data would increase overall team performance, presumably by facilitating development of shared mental models and situation awareness. As noted, this is an important question that could significantly affect the design and future success of mixed-initiative teams. We formulated three general expectations that if confirmed, would support the utility of this form of intra-team communication. The first was that in conditions with integrated psychophysiological communication, joint system task performance would be better than without this implicit form of communication. The second expectation was that participant trust would be highest in conditions including psychophysiological communication to the controller. Finally, the third was that subjective participant workload measures would be reduced in conditions with the psychophysiological communication as compared to conditions without that communication.

Although there were no statistically significant results to report, we nevertheless can speculate about interesting observations regarding the behavior of performance, trust and workload across different conditions. The results seem to indicate that the use of physiological information within the controller has the potential to produce performance that is at least on par with other actuated conditions; according to our preliminary observations, display trustworthiness may tend to be improved and workload reduced when physiology-based state prediction was incorporated. If these results hold up with the addition of more data, it could be argued that there is a compelling reason to include psychophysiological signals. A reduction in subjective workload can affect how a human operator processes environmental stimuli that directly relate to the task [35]. If a human operator is experiencing a high level of workload, the operator is likely to not only have fewer cognitive resources to process important task-relevant information, but also she is likely to have a reduced situation awareness [36].

As noted, the small sample size makes achieving significant statistical results difficult, but other potential design related causes for negative results should be considered. For example, the controller, while able to operate in real time to provide suggestions to the participant as to which agent (automation or human) would be likely to better perform

the task at a given time, was dependent on relatively brittle algorithms that used individual level data regarding expected driving behavior, but could not adapt to variability of individual participants. This inability to adapt the controller meant that the weight given to each of the components (behavioral data and the psychophysiological data) was fixed in the sensor fusion algorithm. In future experiments, we hope to develop approaches to dynamically assess the appropriate weight for each component in order to maximize overall performance.

Another potential challenge, and area of improvement, is that the algorithms used to predict whether an individual is likely to change control (switch from manual drive to autonomy drive or from autonomy drive to manual drive) were developed using group level data from a previous experiment [29] using a similar experimental setup. In that previous study, control toggles between manual drive and autonomy drives were relatively rare. Thus in this study, we were unable to collect enough individual specific training data to train our algorithms. We leveraged data from 15 participants in that previous study to build a group level model based on the recorded physiological, behavioral and environmental data. While this model has been shown effective at predicting control toggles in a manuscript currently under review elsewhere, it does not capture the individual differences that naturally occur between participants.

Additionally, it should be noted that we chose a visual actuator as opposed to one that uses a different modality. It is accepted that humans tend to ignore visual information over time as compared to when information or communication is presented using a different modality. This point reflects the importance of the choice of methods of communication within mixed-initiative teams in future studies. In this case, a visual actuator was chosen because of time and engineering constraints. Future studies may explore different option for multimodal interfaces to communicate with the human are likely to demonstrate improvements in performance.

In summary, given the limited data available, the inclusion of physiological information as a form of implicit communication seems to improve joint human-autonomy performance. Although there were no explicit measures of shared mental models or situation awareness, the literature suggests that communication between team members is critical to the development of these constructs within teams. We therefore suspect that improved team performance is reflective of successful development of shared mental models and situation awareness. Further analysis of this ongoing data collection will enable a more detailed interpretation of these data. Additionally, future work in this area will seek to expand these findings into other task and environment domains, as well as potentially exploring the mechanisms involved in improved team performance.

References

1. Teo, G., Reinerman-Jones, L.: Robot behavior for enhanced human performance and workload. In: Shumaker, R., Lackey, S. (eds.) VAMR 2014. LNCS, vol. 8525, pp. 117–128. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07458-0_12
2. Sheridan, T.B., Parasuraman, R.: Human-automation interaction. Rev. Hum. Factors Ergon. **1**(1), 89–129 (2005)
3. Parasuraman, R., Riley, V.: Humans and automation; use misuse, disuse, abuse. Hum. Factors **39**(2), 23 (1997)
4. Manzey, D., Bahner, J.E., Hueper, A.-D.: Misuse of automated aids in process control: complacency, automation bias and possible training interventions. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, no. 3, pp. 220–224. Sage Publications (2006)
5. McGuirl, J.M., Sarter, N.B.: Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. Hum. Factors **48**(4), 656–665 (2006)
6. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. Hum. Factors: J. Hum. Factors Ergon. Soc. **37**(2), 381–394 (1995)
7. Kaber, D.B., Endsley, M.R.: Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. Process Saf. Prog. **16**(3), 126–131 (1997)
8. Cummings, M.L., Mastracchio, C., Thornburg, K.M., Mkrtchyan, A.: Boredom and distraction in multiple unmanned vehicle supervisory control. Interact. Comput. **25**(1), 34–47 (2013)
9. David, R.A., Nielsen, P.: Defense science board summer study on autonomy. Defense Science Board Washington United States (2016)
10. Schaefer, K.E., Straub, E.R., Chen, J.Y., Putney, J., Evans, A.: Communicating intent to develop shared situation awareness and engender trust in human-agent teams. Cogn. Syst. Res. **46**, 26–39 (2017)
11. Lackey, S., Barber, D., Reinerman, L., Badler, N.I., Hudson, I.: Defining next-generation multi-modal communication in human robot interaction. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 55, no. 1, pp. 461–464. SAGE Publications Sage CA, Los Angeles (2011)
12. Phillips, E., Ososky, S., Grove, J., Jentsch, F.: From tools to teammates: toward the development of appropriate mental models for intelligent robots. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 55, no. 1, pp. 1491–1495. SAGE Publications Sage CA, Los Angeles (2011)
13. Ososky, S., et al.: The importance of shared mental models and shared situation awareness for transforming robots from tools to teammates. In: Proceedings of SPIE, vol. 8387, p. 838710 (2012)
14. Hill, S.G., Barber, D., Evans III, A.W.: Achieving the vision of effective Soldier-robot teaming: Recent work in multimodal communication. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, pp. 177–178. ACM (2015)

15. Kruijff, G.J.M., Janíček, M., Keshavdas, S., Larochelle, B., Zender, H., Smets, N.J.J.M., Mioch, T., Neerincx, M.A., Diggelen, J.V., Colas, F., Liu, M., Pomerleau, F., Siegwart, R., Hlaváč, V., Svoboda, T., Petříček, T., Reinstein, M., Zimmermann, K., Pirri, F., Gianni, M., Papadakis, P., Sinha, A., Balmer, P., Tomatis, N., Worst, R., Linder, T., Surmann, H., Tretyakov, V., Corrao, S., Pratzler-Wanczura, S., Sulk, M.: Experience in System Design for Human-Robot Teaming in Urban Search and Rescue. In: Yoshida, K., Tadokoro, S. (eds.) Field and Service Robotics. STAR, vol. 92, pp. 111–125. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-40686-7_8

16. Klien, G., Woods, D., Bradshaw, J., Hoffman, R., Feltovich, P.: Ten challenges for making automation a team player in joint human agent activity. IEEE Intell. Syst. **19**(6), 4 (2004)

17. Christoffersen, K., Woods, D.D.: How to make automated systems team players. In: Advances in human performance and cognitive engineering research. Emerald Group Publishing Limited, pp. 1–12 (2002)

18. Barber, D., Lackey, S., Reinerman-Jones, L., Hudson, I.: Visual and tactile interfaces for bi-directional human robot communication. In: SPIE Defense, Security, and Sensing. International Society for Optics and Photonics, pp. 87410U (2013)

19. Oh, J., Howard, T.M., Walter, M.R., Barber, D., Zhu, M., Park, S., Suppe, A., Navarro-Serment, L., Duvallet, F., Boularias, A., Romero, O., Vinokurov, J., Keegan, T., Dean, R., Lennon, C., Bodt, B., Childers, M., Shi, J., Daniilidis, K., Roy, N., Lebiere, C., Hebert, M., Stentz, A.: Integrated intelligence for human-robot teams. In: Kulić, D., Nakamura, Y., Khatib, O., Venture, G. (eds.) ISER 2016. SPAR, vol. 1, pp. 309–322. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-50115-4_28

20. Barber, D.J., Abich IV, J., Phillips, E., Talone, A.B., Jentsch, F., Hill, S.G.: Field assessment of multimodal communication for dismounted human-robot teams. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 59, no. 1, pp. 921–925. SAGE Publications Sage CA, Los Angeles (2015)

21. Kaupp, T., Makarenko, A., Durrant-Whyte, H.: Human–robot communication for collaborative decision making—A probabilistic approach. Robot. Auton. Syst. **58**(5), 444–456 (2010)

22. Schaefer, K.E., Evans, A., Hill, S.G.: Command and control in network-centric operations: trust and robot autonomy. In: 20th International Command and Control Research and Technology Symposium, Annapolis, MD (2015)

23. Kulic, D., Croft, E.A.: Affective state estimation for human–robot interaction. IEEE Trans. Robot. **23**(5), 991–1000 (2007)

24. Kulic, D., Croft, E.: Anxiety detection during human-robot interaction. In: 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2005, pp. 616–621. IEEE (2005)

25. Kulic, D., Croft, E.: Estimating robot induced affective state using hidden markov models. In: 2006 The 15th IEEE International Symposium on Robot and Human Interactive Communication, ROMAN 2006, pp. 257–262. IEEE (2006)

26. Rani, P., Sarkar, N., Smith, C.A., Kirby, L.D.: Anxiety detecting robotic system–towards implicit human-robot collaboration. Robotica **22**(1), 85–95 (2004)

27. Dzindolet, M., Peterson, S., Pomranky, T., Beck, P.: The role of trust in automation reliance. Int. J. Hum.-Comput. Stud. **58**(6), 21 (2003)

28. Muir, B.M., Moray, N.: Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics **39**(3), 31 (1996)

29. Drnec, K., Metcalfe, J.S.: Paradigm development for identifying and validating indicators of trust in automation in the operational environment of human automation integration. In: Schmorrow, D.D.D., Fidopiastis, C.M.M. (eds.) AC 2016. LNCS (LNAI), vol. 9744, pp. 157–167. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39952-2_16
30. Metcalfe, J.S., et al.: Building a framework to manage trust in automation. In: Micro-and Nanotechnology Sensors, Systems, and Applications IX. Vol. 10194. International Society for Optics and Photonics (2017)
31. John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory—versions 4a and 54. Berkeley, CA. University of California, Berkeley, Institute of Personality and Social Research (1991)
32. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv. Psychol. **52**, 139–183 (1988)
33. Lee, J.D., Moray, N.: Trust, self-confidence, and operators' adaptation to automation. Int. J. Hum.-Comput. Stud. **40**(1), 153–184 (1994)
34. Lee, J., Moray, N.: Trust, control strategies and allocation of function in human-machine systems. Ergonomics **35**(10), 1243–1270 (1992)
35. Carmody, M.A., Gluckman, J.P.: Task specific effects of automation and automation failure on performance, workload and situational awareness. In: Proceedings of the Seventh International Symposium on Aviation Psychology, vol. 1, pp. 167–171 (1993)
36. Ma, R., Kaber, D.B.: Situation awareness and workload in driving while using adaptive cruise control and a cell phone. Int. J. Ind. Ergon. **35**(10), 939–953 (2005)