Multimodal Bio-Behavioral Approaches to Study Trust in Human-Robot Collaboration

Aakash Yadav
Industrial and Systems Engineering
Texas A&M University
College Station, USA
aakash@tamu.edu

Sarah K. Hopko
Industrial and Systems Engineering
Texas A&M University
College Station, USA
s.hopko@tamu.edu

Yinsu Zhang
Industrial and Systems Engineering
Texas A&M University
College Station, USA
yinsu.zhang@tamu.edu

Ranjana K. Mehta
Industrial and Systems Engineering
Texas A&M University
College Station, USA
rmehta@tamu.edu

***Abstract*—Trust is a central factor for long-term success of human-robot collaborations (HRC). Extending our understanding of trust dynamics and transitivity in HRC requires systematic efforts towards developing novel and continuous metrics to capture trust and trusting behaviors, which are currently limited to self-reports. Survey responses provide minimal information about the changing states of trust and can interfere with the experiment itself. However, when coupled with neurophysiological and behavioral measures, i.e., brain imaging, heart rate data, eye-tracking, and human performances, these multimodal metrics can provide a more accurate and comprehensive picture of the human's cognitive and affective states and behavior. Such bio-behavioral metrics enable an understanding of human behavior in HRC at a much granular and more profound level, and draw in newer perspectives on trust in HRC.**

***Index Terms*—human-robot collaboration, trust in automation, human-in-the-loop, industrial robots**

I. Introduction

Initially designed to perform repetitive tasks in an industrial setting, robots can now be found almost everywhere due to the advancements in the field of robotics. This has led to increased collaboration between humans and robots [1]. Although robotics has evolved rapidly, the development of human factors has largely failed to keep up the pace. A significant amount of research has focused on designing hardcoded robots and automation behaviors that can work with an average human; however, these implementations fail to account for inherent differences between people. Most human factors literature fails to provide details about the internal cognitive state of the operator, which is necessary to learn about trust, cognitive fatigue, and situation awareness [2], [3]. Where studied, behavioral understanding of the human component has been mostly driven by the subjective response, which has several limitations due to its subjective and discrete nature.

Trust is an important design factor for automation; under-trust can lead to suboptimal performance and technology rejection, and over-trust can lead to misuse and accidents [4]. It has been shown that people behave differently and have a varying amount of trust in robots based on their personality and personal dispositions [5]. This means that designing a robot inspired by a general operator behavior will not result in optimal experience from everyone. Systematic ignorance of the granular considerations of trust and its connection to human behavior has led to a lack of understanding of how trust influences behavior and predispositions between users. Hence, there is a need for new measures that quantify trust at cognitive and behavioral levels to adapt the robot. Utilization of multimodal and new perspectives (e.g., bio-instrumentation, neuroergonomics, multiple types of performance metrics, etc.) can help realize this goal.

Funded by SAI Under Grant Number: 28-235950-00001.

II. Need for new perspectives

The default method of interpersonal and human-robot trust remains subjective questionnaires and interviews [6]. They have been the foundation of most past studies investigating technology acceptance, human-robot interactions, collaborative robot designs, etc. However, the nature of self-reported subjective measures require disruption or at least distraction from the ongoing task. Specifically, a trust survey can disrupt cognitive processes associated with completing the task. While subjective measures may remain mainstream in future trust studies, the deployment of objective measures alongside subjective responses can provide a more holistic view of the human. Correlations between neurophysiological measures, such as electrocardiography (ECG), electrodermal activity (EDA), neural activity, and eye-tracking, and human-automation trust have been drawn in existing empirical studies [7]–[9]. In addition, behavioral analysis has also been implemented for trust measurement; with proper coding of the task behaviors, behavioral differences have been found between different trust levels [10], [11]. These measures can reveal insights that subjective responses cannot capture, such as the robot and human performance bottlenecks and task disengagement.

III. The brain, heart, and eye's

Physiological measures such as brain-imaging, EDA, and eye-tracking can provide unbiased insight into the physiological, neurological, and psychological state and can help identify trust change. These measures do not disrupt the task at hand and offer the opportunity to collect data in a continuous fashion. Brain imaging techniques, such as EEG and functional near-infrared spectroscopy (fNIRS), are relatively low cost and portable, allowing for studies in naturalistic settings [12], [13]. Brain imaging can help understand mechanisms that drive certain behaviors; it can help establish associated relationships

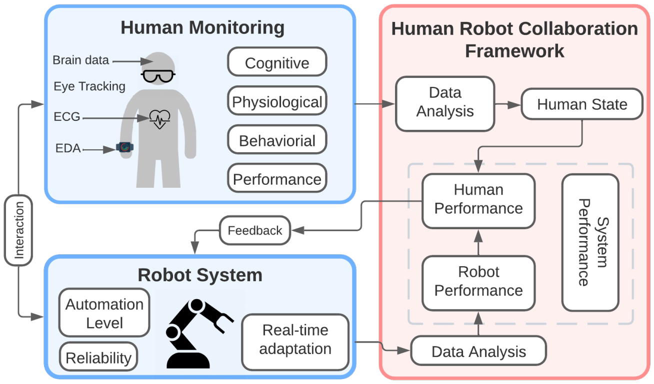

Fig. 1. HRC with bio-behavioral feedback

between different brain regions and identify engagement or disengagement of specific regions associated with cognitive state changes [8]. Eye-tracking can provide cognitive and behavioral information; more specifically, pupil and gaze data (e.g., pupil dilation, fixation features, gaze entropy) can provide information about top-down and bottom-up cognitive processing, and it is also associated with task engagement [14]. Analysis of ECG data can reveal information about both the parasympathetic and sympathetic nervous systems. On the other hand, EDA can provide insights into the affective system, and it has often been related to stress and anxiety [15], [16].

IV. Deep dive into human behavioral metrics

In HRC, performances are often optimized at the system level (e.g., minimizing time to complete a task). However, delineation of human and robot performances is imperative to truly understand the interactive effects of human-robot collaborative systems. For example, a perturbation in robot performance will directly affect human performance and behavior [17]. Utilization of a system performance metric may reveal the overall consequence of faulty robot behavior, but it cannot provide context into how or why human performance is causing an additive decrement. As collaborative robotics become smarter, robot performances are additionally likely to adapt to changes in human performance. Thus, delineating human, robot, and system performances is essential for future work [6]. In addition to delineation, multiple quantitative (e.g., efficiency, accuracy, precision, utilization) and qualitative (e.g., slips, mistakes, lapses) human performance metrics should be considered. These coupled considerations can reveal insights into the cognitive state driving performance changes. Having this level of understanding of human behavior would allow researchers to provide recommendations for robot or environmental support to mitigate slips specifically. This more profound understanding of the human, alongside the use of neurophysiology and subjective responses, can fill knowledge gaps on human factors that have been systematically overlooked in less granular metrics.

V. Implications of new perspectives in HRC

A holistic view of human behavior and performance can help design better HRC. This knowledge can help us understand why trust connects to how humans behave and interact with the robot in certain scenarios (under fatigue, reduced reliability, etc.). This new knowledge of a human's internal state can be used to design HRC such that it minimizes operator workload, fatigue, and increases overall system performance. Further, the collaboration can also be designed to have preventive measures against injuries by detecting the cognitive state. For example, it might be better for the robot to stop the ongoing task if it detects high fatigue and reduced situation awareness. Multi-modal physiological monitoring can help pinpoint the exact conditions and causes that lead to poor performance and help identify the state changes in humans that they are unable to quantify themselves.

VI. Conclusion

Understanding the inner workings of human behavior in the context of HRC is crucial. People behave inherently differently from one another, and designing for the same type of interaction across different people might not be optimal. Someone with a high propensity to trust in automation might perform better than another who does not trust automation. Multi-modal approaches can allow for flexible and fluid collaboration, tailored to each individual. Furthermore, even the behavior of a single individual is ever-evolving, thus requiring the robot to adapt continuously to accommodate for behavior changes and to optimize performance throughout (Fig. 1). This closed-loop control over the robot behavior allows for continuous adaptation, making the system performance robust to robot failure and human fatigue. Physiological measurements allow for data to be obtained in near real-time, and it can be used to adapt robot behavior in situ without disturbing the ongoing task in contrast to subjective measure. Even if adaptation is not the imminent goal, data obtained from multimodal physiological measurements can help identify the states that lead to an increase or decrease in performance. This can improve robot behavior resulting in increased performance and worker satisfaction.

References

[1] M. Knudsen and J. KaiVo-Oja, "Collaborative Robots: Frontiers of Current Literature," *Journal of Intelligent Systems: Theory and Applications*, pp. 13–20, Nov. 2020. [Online]. Available: https://dergipark.org.tr/en/doi/10.38016/jista.682479

[2] G. Matthews, J. Lin, A. R. Panganiban, and M. D. Long, "Individual Differences in Trust in Autonomous Robots: Implications for Transparency," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 3, pp. 234–244, Jun. 2020, conference Name: IEEE Transactions on Human-Machine Systems.

[3] S. Nikolaidis, Y. X. Zhu, D. Hsu, and S. Srinivasa, "Human-Robot Mutual Adaptation in Shared Autonomy," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: Association for Computing Machinery, Mar. 2017, pp. 294–302. [Online]. Available: https://doi.org/10.1145/2909824.3020252

[4] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, Mar. 2004, publisher: SAGE Publications Inc. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1518/hfes.46.1.50_30392

[5] A. Rossi, K. Dautenhahn, K. L. Koay, and M. L. Walters, "The impact of peoples' personal dispositions and personalities on their trust of robots in an emergency scenario," *Paladyn, Journal of Behavioral Robotics*, vol. 9, no. 1, pp. 137–154, Jul. 2018, publisher: De Gruyter Open Access. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/pjbr-2018-0010/html

[6] S. Hopko, J. Wang, and R. Mehta, "Human Factors Considerations and Metrics in Shared Space Human-Robot Collaboration: A Systematic Review," *Frontiers in Robotics and AI*, vol. 9, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2022.799522

[7] S. Hergeth, L. Lorenz, R. Vilimek, and J. F. Krems, "Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving," *Human Factors*, vol. 58, no. 3, pp. 509–519, May 2016, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/0018720815625744

[8] S. K. Hopko and R. K. Mehta, "Neural Correlates of Trust in Automation: Considerations and Generalizability Between Technology Domains," *Frontiers in Neuroergonomics*, vol. 2, 2021. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnrgo.2021.731327

[9] F. Walker, J. Wang, M. H. Martens, and W. B. Verwey, "Gaze behaviour and electrodermal activity: Objective measures of drivers' trust in automated vehicles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 401–412, Jul. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1369847818306703

[10] M. Dikmen and C. Burns, "Trust in autonomous vehicles: The case of Tesla Autopilot and Summon," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2017, pp. 1093–1098.

[11] J. Xu and A. Howard, "How much do you Trust your Self-Driving Car? Exploring Human-Robot Trust in High-Risk Scenarios," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2020, pp. 4273–4280, iSSN: 2577-1655.

[12] Y. Zhu, C. Rodriguez-Paras, J. Rhee, and R. K. Mehta, "Methodological Approaches and Recommendations for Functional Near-Infrared Spectroscopy Applications in HF/E Research," *Human Factors*, vol. 62, no. 4, pp. 613–642, Jun. 2020, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/0018720819845275

[13] R. Mehta and R. Parasuraman, "Neuroergonomics: a review of applications to physical and cognitive work," *Frontiers in Human Neuroscience*, vol. 7, 2013. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnhum.2013.00889

[14] B. Shiferaw, L. Downey, and D. Crewther, "A review of gaze entropy as a measure of visual scanning efficiency," *Neuroscience & Biobehavioral Reviews*, vol. 96, pp. 353–366, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149763418303075

[15] Y. Liu and S. Du, "Psychological stress level detection based on electrodermal activity," *Behavioural Brain Research*, vol. 341, pp. 50–53, Apr. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166432817314006

[16] J. Naveteur and E. Freixa I Baque, "Individual differences in electrodermal activity as a function of subjects' anxiety," *Personality and Individual Differences*, vol. 8, no. 5, pp. 615–626, Jan. 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0191886987900596

[17] S. Hopko, R. Khurana, R. K. Mehta, and P. R. Pagilla, "Effect of Cognitive Fatigue, Operator Sex, and Robot Assistance on Task Performance Metrics, Workload, and Situation Awareness in Human-Robot Collaboration," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3049–3056, Apr. 2021, conference Name: IEEE Robotics and Automation Letters.