Neural Signatures of Trust in Human-Robot Collaboration: A Tale of Two Use-Cases

Aakash Yadav[1], Sarah Hopko[1], Ranjana K Mehta[1,2]
Department of Industrial and Systems Engineering, Texas A&M University, Texas, USA
Department of Mechanical Engineering, Texas A&M University, Texas, USA

Abstract A42: Presented at the 2022 Neuroergonomics and NYC Neuromodulation Conferences

Synopsis

Trust in robotics is essential to its safe and effective utilization. We conducted two human-robot collaboration (HRC) studies to elicit the neurocognitive consequences of trust in robotics. Study one involved HRC for a metal surface polishing task, while study two used a gear planetary gear assembly task – these tasks differed in the human-robot interaction modes (teleoperated vs. collaborative). We manipulated robot reliability, performed brain imaging using fNIRS, and collected physiological heart rate data and subjective measures. Our analysis showed that neural signatures of trust are task-dependent, i.e., the brain responds differently to trusting and distrusting situations across the two HRC use-cases, while perceptions of trust or physiological responses did not provide such discrimination. The insights gained here can help carve newer neural metrics of trust in HRC that are informative of neurocognitive processes of downstream human behaviors.

Background

With the increase in advancements in the field of robotics, robots today are more intelligent, reliable, and robust, allowing for increased human-robot collaboration (HRC) in industrial environments [1]. Although robots are commonplace in industrial settings, they are often placed in separate workspaces where they are inaccessible to humans. Humans and robots both have different competencies and capabilities, which, when combined in the right fashion can result in improved overall performance and worker safety [2]. Humans are better at quick decision-making in unforeseen situations, while robots are resilient to fatigue and can perform complex and precise maneuvers. Despite the advancements, there is a considerable gap in understanding the impact of this collaboration on humans. One of the critical factors in HRC is the understanding of the trusting relationship between humans and robots [3]. Under trust in the robot can lead to suboptimal performance and usability, while overtrust can lead to misuse and overreliance on the robot, causing safety issues [4,5]. The majority of the work towards quantifying trust in HRC has relied on subjective responses collected from the participants. Subjective response suffers from limitations (disturb the ongoing task, discontinuous) and may not uncover underlying cognitive and affective processes. However, when combined with neurocognitive and physiological measures, they can provide deep insights on trust mechanisms essential to HRC.

Methods

Sixteen (age 25.12 ± 3.31 years) and thirty-eight (age 25.37 ± 5.7) participants, balanced by sex, were recruited for studies 1 and 2, respectively. Both the studies were approved by the IRB (IRB2020-0097DCR and IRB2020-0432F). Both studies used a collaborative UR10 robot to perform the tasks in a shared space. In study 1, the participants were asked to perform the metal surface policing task by controlling the robot's end-effector by using the joystick control. In study 2, the robot delivered parts in the shared workspace in a specific order and location, and the participants were asked to carry out a gear assembly task. In both the studies, the participants were asked to practice as many times as they wished before starting the actual experiment to reduce any learning effects. The robot’s behavior was manipulated by changing its reliability, i.e., creating unexpected and unwanted perturbations in its actions, and ten trials each in the reliable and unreliable conditions were performed. Subjective data was collected at the end of every trial in each condition. Brain activity was captured using the fNIRS device with the same probe maps in each study. The probe design consisted of 46 channels which were segmented into 11 regions of interest (anterior prefrontal cortex, left, medial, and right dorsolateral prefrontal cortex, intermediate frontal cortex, Left and right Broca, primary and secondary visual cortex) [6]. fNIRS data were processed using recommended methods published in Zhu et al. [7]. To perform the functional connectivity analysis, the 11 regions were further merged to form 3 regions, i.e., the prefrontal cortex (PFC), premotor and motor cortex (PMMC), and the visual cortex (VC). Heart rate variability data was collected using an Empatica E4 device. Statistical analysis was performed using paired t-tests (alpha=0.05) after testing for normality.

Results

Robot reliability significantly impacted trust in the robot, including both 1-pt trust collected at the end of every trial and a TRUST [8] survey collected at the end of each condition. In study 1 TRUST was higher ($p = .052$, $d = 0.51$) in reliable (5.82 ± 0.88) condition compared to unreliable condition (5.40 ± 1.05), 1-pt trust was not significantly affected. In study two 1pt trust was significantly higher ($p< .001$, $d = 0.77$) in reliable (5.79 ± 1.18) condition compared to the unreliable condition (4.87 ± 1.59). Similarly the TRUST score was significantly higher ($p = .002$, $d = 0.54$) in reliable conditions (5.39 ± 1.03) compared to unreliable conditions (4.65 ± 1.31). Unreliable behavior of the robot increased workload perception in both the studies ($p = .027$, $d = -0.59$ and $p = .042$, $d = -0.34$). Robot reliability also impacted situation awareness, SART [9] was higher ($p < .001$, $d = 0.65$) in the unreliable condition (4.57 ± 1.53) compared to the reliable condition (3.36 ± 1.62) in study 2; however, no significant difference was observed in study 1.

Functional connectivity analysis between PFC, PMMC and the VC was performed for both studies. In study 1, one connection (PFC-PMMC) was observed in both the conditions with increased connection strength in the reliable condition (0.71) as compared to the unreliable condition (0.42). In study 2, two connections (PFC-PMMC and PMMC-VC) were observed in the reliable condition with connection strengths of 0.49 and 0.54, respectively. One connection (PFC-PMMC) was observed in the unreliable condition with a connection strength of 0.62. Further, intra-region connectivity analysis in the PFC region showed that the number of connections stayed the same (i.e., 10) in both conditions in study 1; however, they increased from 6 in reliable to 9 in the unreliable condition in study two.

Brain activation analysis showed no significant difference with reliability in study 1, however, robot reliability significantly affected LBA, APFC, and M1 brain regions ($p=0.014$, $p=0.046$, and $p=0.024$ respectively) in study 2. LBA had higher activation in the reliable condition ($8.04e{-07} \pm 7.82e{-07}$) than in the unreliable condition ($6.13e{-07} \pm 6.12e{-07}$). APFC also had higher activation in reliable conditions ($6.68e{-06} \pm 2.09e{-05}$) compared to unreliable conditions ($3.75e{-06} \pm 1.42e{-05}$). M1 had higher activation in reliable condition (9.64e-07 ± 2.76e-06) than unreliable condition (8.29e-07 ± 2.71e-06).

Heart-rate variability (HRV) was significantly impacted by the reliability manipulation in study 1, however, no significant difference was observed in study 2. Low frequency (LF) ECG ($p = .006$, $d = -0.87$) were significantly higher in unreliable condition (1662.71 ± 523.17) compared to reliable condition (1274.98 ± 593.09). Similarly, high frequency ECG was significantly higher ($p < .001$, $d = -1.20$) in unreliable conditions (355.83 ± 262.28) compared to reliable conditions (264.48 ± 226.79). The mean heart rate was significantly lower ($p = .011$, $d = 0.76$) in unreliable conditions (88.44 ± 11.39) compared to reliable conditions (84.64 ± 8.97).

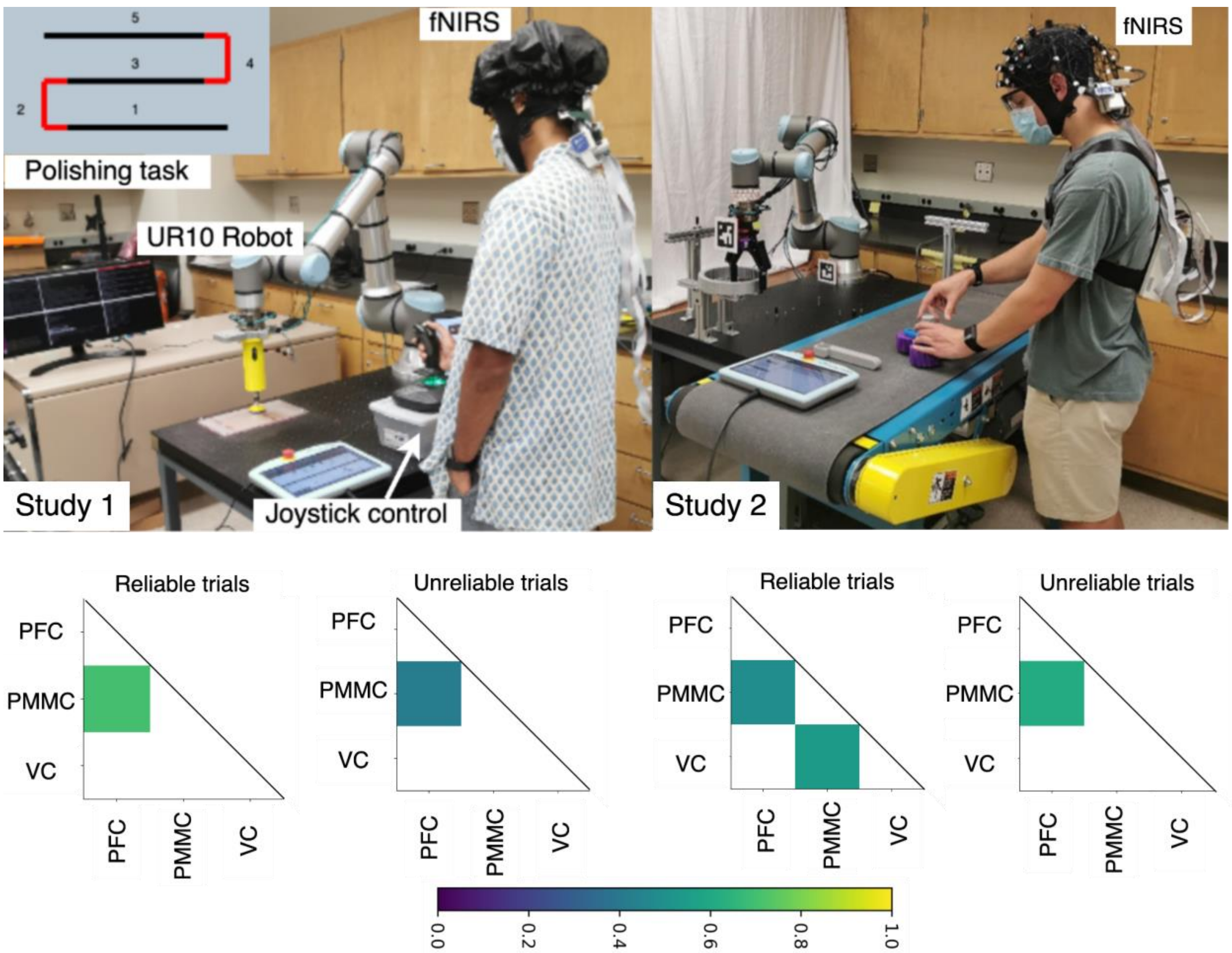

Figure 1. Top row: Experimental setup in study one and study two showing two use cases of the collaborative robot. Bottom row: Functional connectivity between the three brain regions PFC, PMMC, and VC.

Discussion

Robot reliability manipulation was carried out in a similar fashion in both studies. The subjective response shows reliability manipulation decreased trust in the unreliable condition and increased the perceived workload. Although study 2 increased situation awareness, no change was observed in study 1, and the same trend is corroborated by the functional connectivity in the PFC, where no change in connections was observed in study 1.

Although HRV has been explored as a candidate to quantify trust in HRC, we found it is very interaction mode-specific; the effect can only be parsed out only in controlled conditions. Body movements can significantly impact the ability to discern any changes between the trusting and non-trusting conditions.

Even though both studies are similar in design and reliability manipulation, both of them employ very different tasks; study one requires a precise joystick control where the subject is mostly stationary, focusing on the end-effector motion. Study two does not require fine control and the subjects are in constant motion as they try to reach for parts and assemble them. The results show that even though the subjective responses were similar in both the studies, the underlying cognitive mechanisms and the brain regions involved can be very different.

References

[1] Sheridan, T. B. (2016). Human–Robot Interaction: Status and Challenges. Human Factors, 58(4), 525–532. https://doi.org/10.1177/0018720816644364

[2] Vysocky, A., & Novak, P. (2016). Human-robot collaboration in industry. MM Science Journal, 9(2), 903–906. DOI : 10.17973/MMSJ.2016_06_201611.

[3] Hopko, S., Wang, J., & Mehta, R. (2022). Human Factors Considerations and Metrics in Shared Space Human-Robot Collaboration: A Systematic Review. Frontiers in Robotics and AI, 9.

[4] de Visser, E., Peeters, M., Jung, M., Kohn, S., Shaw, T., Pak, R., & Neerincx, M. (2020). Towards a Theory of Longitudinal Trust Calibration in Human-Robot Teams. INTERNATIONAL JOURNAL OF SOCIAL ROBOTICS, 12(2), 459–478. doi: 10.1007/s12369-019-00596-x

[5] B. Wu, Bin Hu and H. Lin, "Toward efficient manufacturing systems: A trust based human robot collaboration," 2017 American Control Conference (ACC), 2017, pp. 1536-1541, doi: 10.23919/ACC.2017.7963171.

[6] Hopko, S. K., & Mehta, R. K. Neural Correlates of Trust in Automation: Considerations and Generalizability Between Technology Domains. *Frontiers in Neuroergonomics*, 26.

[7] Zhu, Y., Rodriguez-Paras, C., Rhee, J., & Mehta, R. K. (2020, June). Methodological Approaches and Recommendations for Functional Near-Infrared Spectroscopy Applications in HF/E Research. Human Factors: The Journal of the Human Factors and Ergonomics Society, 62(4), 613–642. doi: 10.1177/0018720819845275

[8] Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000, March). Foundations for an Empirically Determined Scale of Trust in Automated Systems. International Journal of Cognitive Ergonomics, 4(1), 53–71.

[9] R. M. Taylor, “Situation awareness rating technique (SART): The development of a tool for aircrew systems design,” Situational Awareness in Aerospace Operations (Chapter 3), no. Frace: Neuilly sur-Seine, pp. 111–128, 1990.

Financial Disclosures

This work was supported by the NSF under Grant 1900704 and SecureAmerica Institute Grant 28-235950-00001.