The Role of Rejection within the Trust Calibration Process: Insights from a Mixed-Methods Human-Robot Teaming Experiment

Tony Nguyen
School of Psychology & Exercise Science
Murdoch University
Perth, Australia
Tony.Nguyen745@gmail.com

Justin Fidock
Land Division
Defence Science & Technology
Adelaide, Australia
Justin.Fidock@dst.defence.gov.au

Graeme Ditchburn
School of Psychology & Exercise Science
Murdoch University
Perth, Australia
Graeme.Ditchburn@murdoch.edu.au

***Abstract*—An important influence on the appropriate exploitation of robotic and autonomous systems will be the human-RAS trust calibration process in the context of transitioning between different modes of control. Understanding the dynamic interaction of factors that influence calibration in such use contexts is an important research question. This paper sheds light on these constructs through a simulated experiment in which participants (n = 11) alternated between teleoperation (TO) and supervisory control (SC) whilst monitoring the robot's performance and locating objects concurrently. Applying an integrative mixed-methods approach involving semi-structured interview questions, behavioural observations, and quantitative data from a parallel study, a narrative account of the trust calibration process and the primacy of rejection relating to trust calibration was extracted. Implications for the design of robotic systems and the training of human-robot teams are discussed.**

Keywords—human-robot trust; mixed-methods research; autonomous systems; naturalistic decision-making; rejection

I. INTRODUCTION

Due to advances in science and technology, robotic systems have increased in utility and ubiquity in recent years. Indeed, as these systems become smarter and more adaptive, the idea of a human-robot team moves from science fiction to reality. The advantages of human-robot teams have been demonstrated through increased operational capability for unsafe environments and complex activities that impose hazardous levels of workload and complex information integration [1,8]. However, for this potential to be fully realised, existing challenges within the scope of human-robot interaction need to be addressed. One of these challenges is the degree to which the human operator trusts and subsequently relies on the robot. Another is a deeper understanding of the cognitive processes that underlie the transition between modes of control, particularly in a shared-control context (e.g., joint control of a vehicle).

II. BACKGROUND

A. Trust and Trust Calibration

Trust is a complex and multidimensional psychological construct that drives many facets of human behaviour. Trust is an important topic for human robot teaming because at a basic level, an untrusted system is an unused, or worse, misused system [10]. Trust therefore needs to be appropriate or *calibrated* appropriately to the system. This calibration of trust is the correspondence between a person's trust in the robotic autonomous system (RAS) and its capabilities. When trust correctly corresponds to system reliability and capability, reliance on the system is appropriate to the context and performance, and by extension, is maximised [8]. Current models of trust calibration suggest that the construct is affected by a multitude of factors including who you are (i.e., individual differences), your background (e.g., early experiences with a product or system), and what situation (i.e., a mission context) you are in. Examples of these factors include attentional capacity, operator workload, training, personality traits, automation transparency, and user experience to name just a few [4].

Individually, there is an understanding of how these factors influence trust, however, the dynamic interaction between these factors is less understood [11]. For example, the existing body of knowledge has mostly examined the momentary state of trust and much of the research have been driven by experimental designs attempting to extrapolate the degree to which trust is optimal for given outcomes (e.g., reliance on the robot [2]). What is missing within the extant research is a perspective of the evolution of trust over time and where and when specific factors can manifest or are most prominent (i.e., at what point in the calibration process or in the presence of what variables). For instance, under a shared-control scenario where human and robotic agents take turns at being the active and passive user, the phenomena of transitioning itself has largely been unexamined within the

literature [12]. In a shared control context where the human operator and robot takes over and gives up control, the human operator experiences periods where they are on-the-loop, in-the-loop, or out-of-the-loop [3]. Under these conditions, human operators shift between periods of engagement and disengagement which could potentially affect the trust calibration process.

B. *Rational and Current Study*

The background above alluded to gaps in the literature in relation to the dynamic interaction of the factors that affect the trust calibration process and the dearth of research around trust calibration as it relates to the phenomena of transitioning. Guided by these reasons, the research questions for the current study are as follows:

1. What are the cognitive mechanisms that underlie the transition process from teleoperation (TO) to supervisory control (SC) and vice versa.
2. How does trust calibrate across transitions between SC and TO within a shared-control context in a human-robot team?

III. METHODOLOGY

A. *Design and Framework*

The current study adopted a mixed-methods research (MMR) approach to understanding the phenomena of trust calibration in the context of transitioning between two levels of control (TO and SC). The methodology was also guided by a pragmatic paradigm which advocates for a way of understanding the world that is guided by what is practical and useful. Centrality is placed on the practical utility of the research findings above methodological and epistemological purity [11]. Under this framework, the constructs under investigation are examined within a computer-simulated virtual environment in which a human operator and robot work together in a humanitarian relief mission. Due to technical capabilities and budgetary constraints, a confederate driver was used as the robot, unbeknownst to the participant. The confederate driver was trained to behave like a robot based on advice drawn from subject matter experts to create consistency and accuracy.

B. *Materials*

1) *Tactical Team Simulator*

Participants interface with the robotic vehicle through the Tactical Team Simulator (TTS). The TTS is configured with 15 motion actuated seats with a monitor mounted on the front wall of each cubical and a Windows Surface tablet configured to support switching between TO and SC (Figure 1). Within the current study, one of the seats was configured for TO of the robotic vehicle and another was configured for the confederate robot driver.

2) *The Simulated Scenario*

The simulated scenario was created using Virtual Battle Space (VBS: Bohemia Interactive) to create an environment where a human and robot was working together on a mission. In the current study, the simulation was made to resemble a humanitarian relief mission whereby the participant was tasked with two tasks; a primary task of locating supply drops that have been scattered throughout the map and a secondary task in monitoring the robotic vehicle as it traversed the fictional island. Within the simulation the participant is represented as an avatar that is situated within an optionally crewed platform capable of autonomous driving on a fixed pre-defined path (waypoint navigation). Within the mission, participants were exposed to two modes of operation: SC and TO. Throughout the designated route, the robotic vehicle encountered 12 scenarios involving unsealed roads, obstructed roads, and harsh terrain that caused the robot to become erratic (e.g., veers off course or stops entirely). In such cases, participants were required to attend to the situation and make the decision to manually work through the obstacle (assume control), typically by driving around the obstacle, before returning to SC (relieving control).

Figure 1. The TTS (Left) and the Tablet display with robotic interface (Right)

3) *Trust in Automation Assessment*

To measure how trust was calibrated over time, a 12-item trust questionnaire was used [7]). The scale consists of 12 items that included statements such as “the system is deceptive” and “the system is reliable”. Participants ranked their agreeableness from 1(not at all) to 7(extremely). All scores from the twelve items were added to provide a total between 12 and 84 (low and high trust in automation, respectively). The Scale was used after exposure to the pre-trial training and once again post-trial. Differences in scores were used to guide the interview schedule.

4) *Interview Schedule*

The interview schedule comprised of two components; (1) semi-structured Critical Decision Method (CDM) questions which focus on the cognitive aspects of decision-making at the point of transition and (2), open-ended interview questions relating to technology usage which provided an inductive perspective on how participants viewed and approached technology usage (e.g., tell me a story about a time when a technology system has been unreliable or reliable). In this instance, the critical incidents refer to the transition between SC and TO, and vice versa.

C. Participants

Using a convenience sample, participants were at least 18 years old, had corrected-to-normal vision, and were proficient in English. Additionally, participants were required to possess a current driver's licence and had no reported history of adverse symptomatology relating to screen use. Based on this criterion, a total of 11 participants (M = 7, F = 4) was included in the final analysis. Ethics approval was obtained prior to conducting the study.

Procedure

Before commencing the trial, participants filled out a pre-experiment survey containing a demographic questionnaire and the Trust in Automation assessment. Additionally, each participant was taken through a tutorial on how to operate the vehicle and interface with the robot. On commencement of the mission, one investigator assumed the role of the robot by teleoperating within a separate non-visible station within the TTS. The other investigator alternated between monitoring the participant on a third TTS unit that was configured to provide a god's eye view and observing the participant's behaviour. At the end of the trial, observations by both investigators were exchanged and discussed (e.g., near-crash and detour frequencies). On average, trials lasted 40 minutes and upon completion, participants were required to fill out post-trial surveys and took part in an interview lasting approximately 30 minutes.

The Qualitative CDM data was analysed in accordance with best practice and was theoretically guided by the Recognition-Primed Decision (RPD) Model [7]. To integrate the qualitative and quantitative data, the Pillar Integration Process (PIP; [6]) was utilized as a guiding framework. The PIP is a recent iteration of the joint display technique used for cases where both quantitative and qualitative data exist for the same case and are available for joint examination. Using a four-step process involving listing, matching, checking, and pillar building, the PIP was completed following the initial quantitative and qualitative analyses.

IV. Results

Data from the qualitative CDM analysis revealed differences in the types of cues as well as the cognitive strategies and behavioural responses for each transition. For the human operators, vehicle speed and the severity of the obstruction (e.g., a tree branch in the way versus an unfinished road) held primacy for the transition from SC to TO. In contrast, with the decision to give back control (i.e., TO to SC), the visual cue of a straight road, (as indicated by the primary display and the topographic mini-map), was combined with the interpretation of previous rejection patterns. To clarify, a rejection is an instance where the human operator incorrectly initiates a request to give back control (TO to SC), which may be contingent upon factors such as an incomplete navigation of the obstacle. In relation to rejection, the CDM analysis also revealed that the majority of participants also expressed a strong desire for the robot to display the reason for rejection alongside the error message. The following section discusses an emergent theme in the context of a narrative process that was derived from the PIP.

A. Trust Calibration

One of the major themes that emerged was the concept of trust and how it calibrates with increased exposure to the robot across multiple transitions. This calibration process was found to be multifaceted, with participants actively engaging in strategies to "feel out" the robot by testing its capabilities. One of these strategies is the decision to hold onto TO for a longer period. One of the distinct findings from this process is the role that rejection plays, particularly in the transition from TO to SC. This is exemplified in the following quote:

> After using it a couple of times, I can't recall which ones [referring to transitions] exactly, but after there was some sharp bends. Having tried to relieve control a few times and it rejected me, I used information on how it doesn't like to be handed over control for some situations, and assumed the thinking behind it. Once I realised that it didn't like that, I kept driving until it was a straight road and then handed it over. Yea trial and error from the rejections. That was the main thing.

B. Pillar Integration

By applying the PIP, a narrative timeline of the trial was constructed (Figure 2) A high number of the rejections occurred before transitions 3 and 4, or temporally between 5 and 20 minutes into the trial. The initial rejections provided an initial baseline of expectations which would influence how long participants would stay in TO for. Rejection rates also affected the participant's estimation of the robot's capabilities such that participants with higher rejection rates in the early stages of the experiment remained in a stage of guessing (i.e., should the robot take this obstacle, or should I take over) well into the later stages.

By phase 5 of narrative timeline, where transitions were routine and there were no more cases where the robot rejected a takeover request, rejection no longer factored into the calibration process. In this way, it is plausible to assert that the calibration process reached an equilibrium point (a stability of rejection rate), and rejection becomes less relevant by this point.

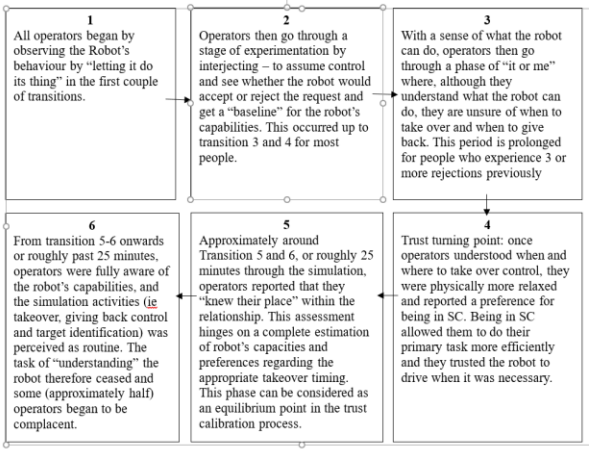

Figure 2. A narrative timeline of the 6 phases of trust calibration

V. Discussion

This study examined the phenomena of trust calibration in the context of transitioning between TO and SC. One of the key findings is the role that rejection plays within the trust calibration process, particularly within the early stages of interaction. Within the current study, the experience of rejection served as a prompt for a re-assessment of robot's capabilities because it was interpreted by human operators as an incorrect judgement of what the robot can do. For instance, a high frequency of rejections was interpreted as the robot being insufficient to handle a given obstacle, and therefore decreases confidence and trust in the robot for that task. As a corollary, a lower frequency of rejections may speed up the trust calibration process because the robot is viewed as competent. In this way, the experience of rejection provides a systematic feedback loop for the assessment of expectancy (an important construct relating to both trust and trust calibration [4,11]), and exists as a salient heuristic for transition-specific decision-making. This line of research can be extended in several ways. For example, future research could examine the influence of the frequency of rejections on the time it takes to calibrate trust and the influence of pre-trial training that incorporates a rejection module (e.g., the correct information-processing of rejection) on the time it takes to calibrate trust. This would assist in clarifying whether the rejection dynamics exists as a product of training effects or the novelty of the scenario.

This study also revealed that the frequency of rejections has a role to play within the early phases of the trust calibration process. More specifically, experiencing a high number of rejections meant that human operators spent more time trying to understand the robot's capabilities through testing and observing. This had two outcomes: an increase in time spent in TO and more cognitive effort spent to assess the robot's capabilities. As a noted caveat, cognitive effort was not measured directly but instead was inferred from the narrative accounts of individuals who expressed frustration with experiencing a high number of rejections. Although there are minor consequences for these outcomes within the current study, there are safety and performance considerations when extrapolated into mission-critical settings. For example, an increase in time in TO implies that the human operator is in-the-loop for an extended period, or more than what is required [3]. In this way, the limited attentional bandwidth that human operators are bound to is spent on calibrating trust instead of mission-related tasks, thus potentially affecting performance.

Finally, the current findings also have implications for communication within a human-robot teaming context. In a shared decision-making paradigm where humans and robots have joint control, each party needs to have insight into the decision-making process [13]. In human-human teams, this process is normally facilitated through bi-directional communication and a give-and-take process where teammates query for information related to the ongoing mission. Within human-robot teams, the capacity for extensive dialogue is limited to the design choices within the robotic interface; in this case through the communication of rejection. In the current study, there was consensus amongst the participants that the robotic interface needed to display the reasoning behind each rejection. Combined with the salience of rejection, the findings imply that, transparency (a correlate of trust calibration [5,11] can be facilitated through design and interface decisions to expedite the trust calibration process. Future research could examine the effects of different modalities of rejection-specific communication on the trust calibration process.

VI. Conclusion

The trust calibration process is an important barrier towards the appropriate exploitation of RAS. By employing a mixed-methods framework, the current study provided a narrative account of the trust calibration process and highlighted the role that rejection plays within that process. Collectively, the findings suggest that rejection has a prominent role within the trust calibration process and thus warrants further research.

Acknowledgement

The authors wish to acknowledge Defence Science and Technology, the United States Army Research Laboratory, Adelaide University, and Murdoch University.

References

[1] Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Human factors, 58(3), 377-400.

[2] Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517-527.

[3] Endsley, M. R. (2017). From Here to Autonomy: Lessons Learned from Human-Automation Research. *Human Factors*, *59*(1), 5–27.

[4] Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust . *Human Factors: The Journal of the Human Factors and Ergonomics Society* , *57*(3), 407–434.

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

[6] Johnson, R. E., Grove, A. L., & Clarke, A. (2017). Pillar Integration Process: A Joint Display Technique to Integrate Data in Mixed Methods Research. *Journal of Mixed Methods Research*,

[7] Klein, G. (2008). Naturalistic Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 456–460.

[8] Lee, J. D., & See, K. a. (2004). Trust in automation: designing for appropriate reliance. *Human Factors*, *46*(1), 50–80.

[9] Mertens, D. M. (2012). What Comes First? The Paradigm or the Approach? *Journal of Mixed Methods Research*, *6*(4), 255–257.

[10] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2), 230-253.

[11] Pop, V. L., Shrewsbury, A., & Durso, F. T. (2015). Individual differences in the calibration of trust in automation. *Human Factors*, *57*(4), 545–56.

[12] Wong, B. L. W. (2003). Data analysis for the critical decision method. *Task Analysis for Human Computer Interaction*, (January 2004).

[13] Barnes, M. J., Chen, J. Y., & Hill, S. (2017). Humans and Autonomy: Implications of Shared Decision Making for Military Operations (No. ARL-TR-7919). US Army Research Laboratory Aberdeen Proving Ground United States.