Quantifying Trust Evolution Through Predictability of Compliance Behavior: A Dynamical Systems Perspective

Matthew A. Peel[1,2], Matthew J. Scalia[1,2], Jayci T. Landfair[2,3], Myke C. Cohen[1,2], Matthew M. Willett[1,2], Dr. Jamie C. Gorman[1,2], Dr. Nancy J. Cooke[1,2]

[1]Human Systems Engineering, Ira A. Fulton Schools of Engineering at Arizona State University, Mesa, AZ, USA
[2]Center for Human, Artificial Intelligence, and Robot Teaming, Arizona State University, Tempe, AZ, USA
[3]Department of Psychology, Arizona State University, Tempe, AZ, USA

INTRODUCTION

For decades, researchers have explored the concept that humans and technologies may form effective and resilient teams. Within the context of human-automation teams (HATs), trust influences the exchange and credibility of information among team members and facilitates effective delegation of task roles based on each teammate's strengths and weaknesses. Trust, or the belief that an agent will achieve one's goals and a willingness to rely on them in the face of risk and uncertainty, is commonly measured through self-report surveys, behavioral observations, and psychophysiological responses (Mayer et al., 1995; Lee and See, 2004; Schaefer et al., 2021). An issue with these trust measures is that they capture different but related facets of trust, sometimes leading to conflicting conclusions about trust. This is possibly due to how these measures are rarely analyzed to capture the momentary fluctuations that characterize the dynamic process of trust*ing*, in which trust is formed, dissolved, and repaired over time (Chiou and Lee, 2021; de Visser et al., 2019; Yang et al., 2022).

In response to these challenges, recent efforts have sought to understand trust dynamics through time-series representations of repeated trust measurements and the use of computational modeling to predict trust as it develops over time (Yang et al., 2022). Additionally, complex adaptive systems science techniques have been leveraged to frame teaming as a dynamical system and apply non-linear analysis techniques to investigate aspects of team interactions that are associated with trust (Cooke et al., 2013; Gorman et al., 2017).

Recurrence quantification analysis (RQA) is a nonlinear technique for modeling dynamic patterns in complex systems (Riley & Van Orden, 2005). The application of RQA spans many fields, but notably, has been used to analyze communication patterns and psychological data to understand team behaviors and social dynamics, such as coordination, cooperation, and trust (Mitkidis et al., 2015; Tolston et al., 2018; Grimm et al., 2018; Demir et al., 2021). A key output metric of RQA is percent determinism (%DET), which quantifies the predictability of dynamic patterns within a system (Riley & Van Orden, 2005). This metric has shown promise in predicting trust development in HATs; for example, Demir et al. (2021) calculated %DET over HAT communication flow dynamics and predicted the development of trust in an autonomous teammate. Despite these advancements, the potential of RQA to dynamically characterize trust measures remains largely unexplored (Landfair et al., 2021). These may reveal novel insight into how trust changes in time and across context.

In this paper we address this gap in application by investigating compliance decisions over time to quantify the predictability of trust's nonlinear development in dynamic HAT interactions. We hypothesize that: (1) the predictable patterning of compliance behavior will be associated with perceived trust, and (2) the predictable patterning of compliance behaviors will be sensitive to the same factors that influence trust in HAT, such as communication style and automation confidence levels (Guznov et al., 2020; Wang et al., 2016).

METHOD

A remote study was conducted in an urban search and rescue (USAR) simulation (Raimondo et al., 2022) where 66 participants worked with a simulated autonomous robot to navigate a collapsed building and administer medical care to survivors. We used a mixed experimental design, focusing on communication style as a between-subjects variable, and confidence information presence and mission sequence as within-subjects variables. Participants were randomly assigned to a communication style condition in which the robot provided medical care recommendations either graphically, textually, or both. Participants completed two 20-minute missions comprising 20 decisions, with robot recommendations having a 70% reliability rate. Mission order was held constant, but confidence indicators accompanied robot recommendations in only one mission per participant, assigned randomly in a counterbalanced manner.

Participants explored the environment freely, and the order in which participants experienced either correct or incorrect recommendations was determined by which areas they went to first, and the survivors they discovered there. Compliance behaviors were recorded as a time-series and reflected the sequence of participant decisions to comply with the recommendation or not. Self-reported trust was measured using an adapted version of the Chancey et al., (2016) trust survey administered after each Mission. RQA was conducted on each participant's unique series of compliance behaviors, and %DET was calculated. %DET provided a value of the

predictability of compliance behavior which could be used to assess how these trust manipulations affected the patterning of behavior in the team and therefore trust throughout the missions.

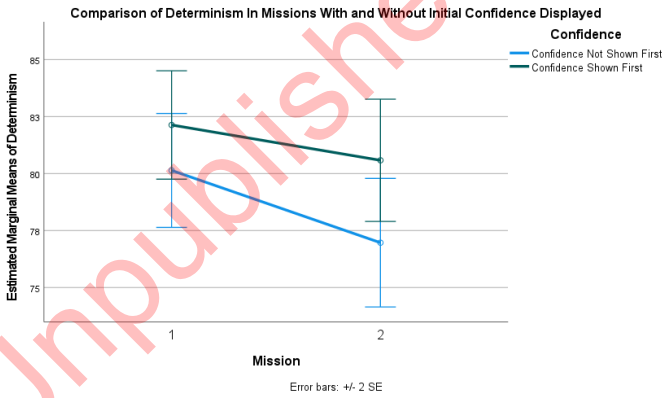
RESULTS

To understand the overlapping trust constructs between patterning of compliance behavior and perceived trust, we first calculated the Pearson correlation coefficient between the %DET of participant compliance behaviors and self-reported trust. %DET and trust scores were not significantly correlated, $r(128) = .067$, $p = .450$, suggesting that perceived trust and pattern predictability of compliance behavior may capture different facets of trust.

Next, a 3 (communication style: text, graphic, both) × 2 (confidence order: confidence shown in Mission 1 or in Mission 2) × 2 (Mission: 1st and 2nd) mixed repeated measures analysis of variance (ANOVA) on %DET was run. We found a significant relationship between-subject main effect of confidence order, $F(1, 59) = 4.39$, $p < .05$. This suggested that the predictability of compliance behaviors was different as a function of whether a robot's confidence level was first shown to their human teammate in Mission 1 and then removed in Mission 2 ($M = 81.36$) versus not shown to their human teammate in Mission 1 and then shown in Mission 2 ($M = 78.55$). No other significant main or interaction effects were found, indicating that %DET was similar from Mission 1 to Mission 2 and that the communication style of the robot did not affect the predictability of compliance behaviors in its human teammate.

Figure 1

%DET by Confidence order and Mission



DISCUSSION

Our findings support the utility of compliance pattern predictability as a unique measure of trust over time. The absence of correlation between compliance pattern predictability and self-reported trust does not negate the possibility of a relationship between the predictability of compliance behavior and trust; rather, it may suggest that each measure could be sensitive to different aspects of trust. Notably, the predictability of compliance patterns was significantly influenced by the order in which confidence information was received. Participants exposed to confidence information from the outset demonstrated greater predictability in their compliance behavior and this trend remained stable after the robotic teammate no longer provided this information. In contrast, those who received confidence information in mission two exhibited less predictability across both missions. This observation implies that patterns of compliance behavior may persist, even after the capabilities of a robotic teammate change. This persistence has been demonstrated in other research that investigated trust over time (Mendoza, 2023).

Because the patterning of compliance behavior is affected by confidence information, a recognized means of manipulating trust (Wang et al., 2016), it is likely sensitive to some dimension of trust. These findings suggest that the calculation of RQA and its %DET variable on compliance behaviors could offer a novel approach to analyzing the fluctuations of trust over time. Future research should explore ways dynamic measures can be used to quantify dynamic aspects of team interactions to predict trust, and how they may be sensitive to aspects of trust not detected in surveys.

REFERENCES

Chancey, E. T., Bliss, J. P., Yamani, Y., & Handley, H. A. (2017). Trust and the compliance–reliance paradigm: The effects of risk, error bias, and reliability on trust and dependence. *Human factors*, 59(3), 333-345.

Chiou, E. K., & Lee, J. D. (2023). Trusting automation: Designing for responsivity and resilience. *Human factors*, 65(1), 137-165.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive science*, 37(2), 255-285.

Demir, M., McNeese, N. J., Gorman, J. C., Cooke, N. J., Myers, C. W., & Grimm, D. A. (2021). Exploration of teammate trust and interaction dynamics in human-autonomy teaming. *IEEE transactions on human-machine systems*, 51(6), 696-705.

De Visser, E. J., Peeters, M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerincx, M. A. (2020). Towards a theory of longitudinal trust calibration in human–robot teams. *International journal of social robotics*, 12(2), 459-478.

Gorman, J. C., Dunbar, T. A., Grimm, D., & Gipson, C. L. (2017). Understanding and modeling teams as dynamical systems. *Frontiers in psychology*, 8, 1053.

Grimm, D., Demir, M., Gorman, J. C., & Cooke, N. J. (2018, June). The complex dynamics of team situation awareness in human-autonomy teaming. In *2018 IEEE conference on cognitive and computational aspects of situation management* (CogSIMA) (pp. 103-109). IEEE.

Guznov, S., Lyons, J., Pfahler, M., Heironimus, A., Woolley, M., Friedman, J., & Neimeier, A. (2020). Robot transparency and team orientation effects on human–robot teaming. *International Journal of Human–Computer Interaction*, 36(7), 650-660.

Landfair, J., Nguyen, T., Magaldino, C., Amazeen, P. G., Huang, L., & Demir, M. (2022, September). Dynamic Modeling of Trust in Automation in Human-Autonomy Teaming. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 66, No. 1, pp. 1174-1178). Sage CA: Los Angeles, CA: SAGE Publications.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1), 50-80.
Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, 20(3), 709-734.
Mendoza, AS. (2023) Impact of Changing Automation Reliability on User Trust. Masters Thesis. San Jose State University.
Mitkidis, P., McGraw, J. J., Roepstorff, A., & Wallot, S. (2015). Building trust: Heart rate synchrony and arousal during joint action increased by public goods game. *Physiology & behavior*, 149, 101-106.
Raimondo, F. R., Wolff, A. T., Hehr, A. J., Peel, M. A., Wong, M. E., Chiou, E. K., ... & Cookea, N. J. (2022, September). Trailblazing Roblox Virtual Synthetic Testbed Development for Human-Robot Teaming Studies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 66, No. 1, pp. 812-816). Sage CA: Los Angeles, CA: SAGE Publications.
Riley, M. A., and Van Orden, G. C. (2005). Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences. Available at: https://www.nsf.gov/pubs/2005/nsf05057/nmbs/nmbs.jsp.
Schaefer, K. E., Perelman, B. S., Gremillion, G. M., Marathe, A. R., & Metcalfe, J. S. (2021). A roadmap for developing team trust metrics for human-autonomy teams. In *Trust in Human-Robot Interaction* (pp. 261-300). Academic Press.
Tolston, M. T., Funke, G. J., Alarcon, G. M., Miller, B., Bowers, M. A., Gruenwald, C., & Capiola, A. (2018, September). Have a heart: Predictability of trust in an autonomous agent teammate through team-level measures of heart rate synchrony and arousal. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 714-715). Sage CA: Los Angeles, CA: SAGE Publications.
Wang, N., Pynadath, D. V., & Hill, S. G. (2016, March). Trust calibration within a human-robot team: Comparing automatically generated explanations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction* (HRI) (pp. 109-116). IEEE.
Yang, X. J., Guo, Y., & Schemanske, C. (2022). From trust to trust dynamics: Combining empirical and computational approaches to model and predict trust dynamics in human-autonomy interaction. *In Human-Automation Interaction: Transportation* (pp. 253-265). Cham: Springer International Publishing.