

It's Not Teaming Until It's Teaming

The Essential Role of Bi-Directional Influence in Human-Automation Teams

Matthew A. Peel^{1,3,†}, Jayci T. Landfair², Matthew J. Scalia^{1,3}, Myke C. Cohen^{1,3}, Matthew M. Willett^{1,3},
Nancy J. Cooke^{1,3} and Jamie C. Gorman^{1,3}

¹Human Systems Engineering, Ira A. Fulton Schools of Engineering at Arizona State University, Mesa, AZ, USA

²Department of Psychology, Arizona State University, Tempe, AZ, USA

³Center for Human, Artificial Intelligence, and Robot Teaming, Arizona State University, Tempe, AZ, USA

ABSTRACT

Modeling humans and robots together as a team can alleviate many of the challenges of human supervisory control, however, the construct suffers from several technological limitations. In this paper, we argue that mutual adaptation through bi-directional influence is a necessary component of all members of a team and is required for true human-automation teaming to exist. Bi-directional influences require that all members of a team possess the ability to change as a result of team interaction (plasticity) and be capable of detecting changes in others and the environment and dynamically adapting as a result of them (responsiveness). Trust at the team level and an understanding of trust dynamics may provide a promising way forward for imbuing these technologies with the capability of mutual adaptation.

KEYWORDS

Interactive Team Cognition, Human-Automation Teaming, Human-Robot Interaction, Trust Dynamics, Trust in Automation, Relational Trust, Mutual Adaptation, Bi-Directional Influence, Autonomy, Artificial Intelligence

ACM Reference format:

Matthew A. Peel, Jayci T. Landfair, Matthew J. Scalia, Myke C. Cohen, Matthew M. Willett, Nancy J. Cooke and Jamie C. Gorman. 2023. It's Not Teaming Until It's Teaming: The Essential Role of Bi-Directional Influence in Human-Automation Teams. In *Proceedings of Human Robot Interaction conference (HRI'23)*. Boulder, CO, USA, 4 pages. <https://doi.org/>

1 Introduction

Over the past 30 years, the idea that human interaction with increasingly autonomous automation technologies could be designed for and classified as “teaming” has grown in popularity [1]. For simplicity, we will use the term automation to refer to any embodied or disembodied increasingly autonomous automation, AI, or technology capable of advanced information processing, decision-making, and independent action. A key distinction between Human-Automation Teaming (HAT) and other interaction paradigms is that humans and technology interact as

peers rather than in hierarchies of monitors/supervisors and subordinates [2]. Modeling humans and robots together as a team can alleviate many of the challenges of human supervisory control, such as poor vigilance, automation bias, skill degradation, and out-of-the-loop phenomena [1].

Despite the conceptual benefits, achieving HAT has suffered from many of the same challenges described by the “AI Effect”, where the complexity of achieving true interdependent and resilient HAT is often underestimated [3] [4]. This paper leverages ideas from Interactive Team Cognition Theory, Relational Trust, and Trust Dynamics to discuss a major challenge overlooked in the design of technology for HAT [5] [6] [7]. In this paper, we argue that for technology to achieve true, interdependent teaming, *all members* must be capable of dynamic mutual adaptation based on their interactions with one another, and that trust at the *team* level is a fundamental component that drives adaptation over time.

Dynamic mutual adaptation requires bi-directional influence between all members of a team, where influence refers to the mutual impact that team members have on each other's thoughts, behaviors and performance within a dynamic team [8]. Two technological limitations of bi-directional influence have thus far precluded automation from teaming. The first is “plasticity”, which refers to technology's ability to change as a result of evolving team interaction. The second is “responsiveness”, which refers to the ability to detect changes in the environment and individuals and adapt one's goals, intentions, interaction strategies, and cognitive processes to better meet the needs of the team or influence it. Plasticity is therefore changes directed by external sources, while responsiveness refers to changes instigated by internal direction as a result of information collected. Together, these two concepts form the necessary components for bi-directional influence.

2 Team Cognition

Team cognition, or the system-level cognitive processes originating from interactions of individuals toward common goals, distinguishes teams from other groups [9]. Similar to streams of inner monologue and conscious thought that occur *in the heads* of individuals, Interactive Team Cognition (ITC) theory posits that

team cognition is the streams of bi-directional, dynamic interaction that occurs *between the heads* of team members over time [5]. When teams interact to overcome challenges, situations, or events, they share information, knowledge, and experiences; leverage the skills, and strengths of individuals while seeking to compensate for weaknesses; and achieve resilience through dynamic interactions [5] [10] [11]. Team cognition can be observed at the team level in activities such as planning, reasoning, and problem-solving.

Taking the ideas of ITC a step further, we posit that the interaction over time changes the team, at the team *and* individual levels. During these interactions, members exert bi-directional influence, iteratively changing the individual *and* collective states of the team as they progress toward achieving the team's goal, as shown in Figure 1a [12]. That is, the team, and the individuals that comprise it, demonstrate both plasticity and responsiveness. It is not enough for individuals to engage with one another. A group of individuals who do not influence one another, and change over time would not be classified as a team. Furthermore, it is possible that in a group of individuals, some members form a team, while others do not (cf., “teamness”;) [9]. These distinctions may seem obvious for human teams, but become less clear when considering the limitations of automation.

3 Automation Rigidity

To date, the design of automation has focused on taskwork by creating independent and reliable technology capable of learning and adapting to tasks and contexts, and extending its capabilities as changing task contexts require achieving dynamic goals (i.e., graceful extensibility) [11]. However, the research efforts that have been made at advancing automation capable of learning and adapting to the humans that they interact with have not been widely adopted, resulting in a majority of HAT research that uses rigid automation. Rigidity, the opposite of plasticity, specifically refers to the inability of automation to change *as a result of interactions with others* over time. Rigidity also implies that automation will not instigate internal changes in response to information acquired through interaction (responsiveness).

The effects of rigid automation on team cognition are demonstrated in Figure 1. In Figure 1a, we see three human teammates interacting over time. As a result of this interaction, each individual's state and the state of the team evolves. This evolution over time as a result of interaction is the team's evolving cognition. Alternatively, in Figure 1b, two humans form a team in which the individual states and the collective state of the team evolve, but because the automation is rigid, its state does not change. Its goals, intentions, computational processes, and strategies remain relatively static. Downstream, the human team possesses an evolved understanding based on a history of interaction that allows for emergent team cognition, but when humans are paired with automation they must adjust their

interactions to the automation because it does not evolve with them.

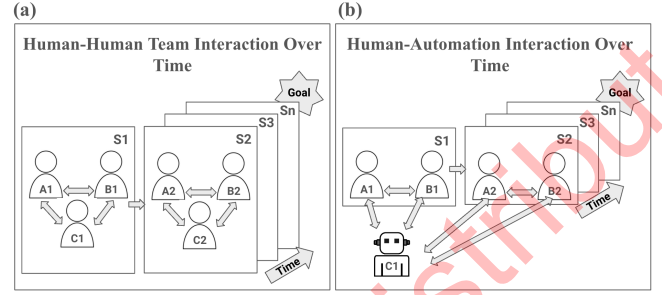


Figure 1: Comparison of two teams within an environment working towards a goal. Figure 1a is a team of three humans. Each member is labeled A, B, and C representing different sets of knowledge, skills, and experience. The team state is denoted by the letter S and is shown to evolve as a result of team interactions. Individuals are also shown to change as a result of the interaction as depicted in their numbers changing. Figure 1b depicts a two-person team and an automation where the humans are changing (in the same manner as 1a) but the automation remains in its original state.

4 A Focus on Team Process

The limitations of automation are evident in the extant body of research in the human-automation/ human-robot interaction literature. One example is automation transparency, which seeks to overcome the challenges of interacting with complex technology by attempting to make information about its processes sufficiently observable [13] [14]. Transparency may be beneficial for technology that is relatively simple and engaged in a supervisory manner, however, it has been criticized in HAT contexts. This is because the dominant black-box approaches to improving the capabilities of automation have made the internal workings and processes increasingly inscrutable [15]. Furthermore, novel challenges and rapidly unfolding contexts create an exponential number of behavioral contingencies. Making all of these internal workings observable results in a bottleneck in terms of the workload required to understand them at any given moment. Designing for transparency can create rigid interactions with technology where the only solution is to provide more and more information [16] [17].

Similar to human teams, arguments have been made that the information intended to be conveyed through transparency, should instead be distributed throughout team interactions [17]. As team interactions occur over time, a history is developed within the team about the automation that serves to supplant the need for immediate transparency of all of its internal workings. To support teammates' history-based knowledge of its trustworthiness through teaming interactions, automation must also be capable of bi-directional influence. Automation capable of

bi-directional influence in this example could detect instances where the team needs to understand information about its internal processes and dynamically instigate team interaction that distributes that information. These are not hard coded events, but dynamic activities that occur within the ongoing dialect of team interaction. They become a component of the team's cognition.

While mutual adaptation through bi-directional influence is required for HAT, the exact means by which they can occur is still lacking. Team trust and an understanding of trust dynamics may provide a way forward for achieving bi-directional influence.

4 A Way Forward:

Trust dynamics refers to the continuous and evolving nature of trust, including the momentary fluctuations, changes, and developments of trust over time, as influenced by various factors and interactions [18] [19]. At the team level, members' trust in a team's ability to achieve the team's goal acts as a heuristic, or shorthand indicator for members to adapt their behavior or instigate team interaction. For automation to have effective bi-directional influence, it must be able to perceive, predict, and understand how different actors' trust changes in time and across contexts. Understanding trust dynamics creates a way to signal to automation when it should adapt goals, intentions or interaction strategies to be more consistent with team expectations and needs or dynamically engage in the trust calibration process through the use of strategies for trust repair or dampening [6] [20]. Several methods for modeling trust dynamics and using them to create trust-aware automation have already been demonstrated with promising results [21] [22].

Furthermore, in HATs, it is important to note that humans will not be infallible, and the technology may not be required to adapt to or adhere to the human's expectations. Instances may arise wherein a non-human agent is required to influence a human towards more realistic or advantageous goals or expectations. An understanding of trust dynamics makes it possible for technology to, in turn, influence human counterparts in ways similar to human-human teams.

5 Conclusion

A fundamental component of teams is that they possess team cognition through interaction. The activity of team cognition occurs over time and results in mutual adaptation at the individual and team level. Mutual adaptation occurs through bi-directional influences exerted between team members. Bi-directional influences require that all members of a team possess the ability to change as a result of team interaction (plasticity) and be capable of detecting changes in others and the environment and dynamically adapting as a result of them (responsiveness). For HAT to exist, technology must be capable of both plasticity and

responsiveness. To date, HAT engineering and research have yet to address these issues fundamental to teaming. Leveraging the growing body of knowledge on trust dynamics, automation may use this information as an input to sense and instigate changes dynamically with the team as they evolve through interaction.

REFERENCES

- [1] National Academy of Sciences. 2021. Human-AI Teaming: State of the Art and Research Needs. <https://doi.org/10.17226/26355>
- [2] Thomas O'Neill, Nathan McNeese, Amy Barron, and Beau Schelble. 2022. Human-autonomy teaming: A review and analysis of the empirical literature. *Human Factors* 64, 5 (Sep. 2022), 904-938. <https://doi.org/10.1177/00187208211063011>
- [3] Marvin Minsky. 2000. Steps toward artificial intelligence. In *Artificial Intelligence: Critical Concepts*, 102.
- [4] Ray Kurzweil. 2005. The singularity is near. In *Ethics and Emerging Technologies*, London: Palgrave Macmillan UK, 393-406.
- [5] Nancy J. Cooke, Jamie C. Gorman, Christopher W. Myers, and Jasmine L. Duran. 2013. Interactive team cognition. *Cognitive Science* 37, 2, 255-285.
- [6] Erin K. Chiou and John D. Lee. 2023. Trusting automation: Designing for responsiveness and resilience. *Human Factors* 65, 1, 137-165.
- [7] Jessie X. Yang, Yaohui Guo, and Christopher Schemanske. 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*, Cham: Springer International Publishing, 253-265.
- [8] Julie L. Harrison, Jamie C. Gorman, Jason G. Reitman, and Sidney D'Mello. 2023. Toward the Validation of a Novel Measure of Individuals' Influence During Team Collaborations. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 67, no. 1, Sage CA: Los Angeles, CA: SAGE Publications, pp. 290-295.
- [9] Nancy J. Cooke, Myke C. Cohen, Walter C. Fazio, Laura H. Inderberg, Craig J. Johnson, Glenn J. Lematta, Matthew Peel, and Aaron Teo. 2023. From Teams to Teamness: Future Directions in the Science of Team Cognition. *Human Factors*, 00187208231162449.
- [10] Eduardo Salas, Terry L. Dickinson, Sharolyn A. Converse, and Scott I. Tannenbaum. 1992. *Toward an understanding of team performance and training*.
- [11] David D. Woods. 2015. Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering & System Safety* 141, 5-9. DOI:<https://doi.org/10.1007/3-540-09237-4>.
- [12] Jamie C. Gorman, Terri A. Dunbar, David Grimm, and Christina L. Gipson. 2017. Understanding and modeling teams as dynamical systems. *Frontiers in Psychology* 8, 1053.
- [13] Jessie YC Chen, Shan G. Lakhmani, Kimberly Stowers, Anthony R. Selkowitz, Julia L. Wright, and Michael Barnes. 2018. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science* 19, 3, 259-282.
- [14] Koen van de Merve, Steven Mallam, and Salman Nazir. 2024. Agent transparency, situation awareness, mental workload, and operator performance: A systematic literature review. *Human Factors* 66, 1, 180-208.
- [15] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1, 5, 206-215.
- [16] Guznov, Svyatoslav, J. Lyons, Marc Pfahler, A. Heironimus, Montana Woolley, Jeremy Friedman, and A. Neimeier. "Robot transparency and team orientation effects on human-robot teaming." *International Journal of Human-Computer Interaction* 36, no. 7, 650-660.
- [17] Christopher A. Miller. 2021. Trust, transparency, explanation, and planning: Why we need a lifecycle perspective on human-automation interaction. In *Trust in Human-Robot Interaction*, Academic Press, 233-257.
- [18] Mengyao Li, Sofia I. Noeovich, Ernest V. Cross, and John D. Lee. 2023. Explaining Trust Divergence: Bifurcations in a Dynamic System. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 67, no. 1, Sage CA: Los Angeles, CA: SAGE Publications, pp. 139-144.
- [19] Jessie X. Yang, Yaohui Guo, and Christopher Schemanske. 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*, Springer International Publishing, Cham, 253-265.
- [20] Ewart J. De Visser, Marieke MM Peeters, Malte F. Jung, Spencer Kohn, Tyler H. Shaw, Richard Pak, and Mark A. Neerinx. 2020. Towards a theory of longitudinal trust calibration in human-robot teams. *International Journal of Social Robotics* 12, 2, 459-478.

- [21] Anqi Xu and Gregory Dudek. 2015. Optimo: Online probabilistic trust inference model for asymmetric human-robot collaborations. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, 221-228.
- [22] Shreyas Bhat, Joseph B. Lyons, Cong Shi, and X. Jessie Yang. 2023. Effect of Adapting to Human Preferences on Trust in Human-Robot Teaming. In Proceedings of the AAAI Symposium Series, vol. 2, no. 1, 5-10.

Unpublished Work, Do Not Distribute