# Trust and Reliance in Compositional Control Teams

KAZUHIKO MOMOSE and TROY R WEEKES, Human-Centered Design, L3Harris Institute for Assured Information, Florida Institute of Technology, USA

THOMAS C ESKRIDGE, Computer Science, Human-Centered Design, L3Harris Institute for Assured Information, Florida Institute of Technology, USA

DANIEL KIDWELL, US Department of Defense, USA

Humans can benefit from the increasingly more advanced Artificial Intelligence (AI)-powered systems that are being deployed in safety-critical systems and in their daily lives. They can benefit even more when humans and intelligent systems begin to work together as a team by exploiting the strengths of each other to improve overall performance. Trust is believed to play a critical role in formulating effective human-agent teams. Still, prior studies suggest that further research is needed to better understand how trust affects human-agent teamwork. In this paper, we aim to establish a foundation for a better understanding of trust in the context of human-agent teamwork. First, we address two types of human-agent teams: compositional control (e.g., automated driving assistance, collaborative text editing) and non-compositional control (e.g., warehouse autonomous robots, cybersecurity) teams. Then, we propose an extension to a risk-taking relationship trust and decision model that exploits the unique characteristics of compositional control teams.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models**.

Additional Key Words and Phrases: Trust, compositional control, human-agent teamwork, intent

**ACM Reference Format:**
Kazuhiko Momose, Troy R Weekes, Thomas C Eskridge, and Daniel Kidwell. 2023. Trust and Reliance in Compositional Control Teams. In *In Workshop on Trust and Reliance in AI-Human Teams, at CHI2023, April 23, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 11 pages.

## 1 INTRODUCTION

Artificial Intelligence (AI)-powered systems such as ChatGPT [1], Cicero [14], and others have demonstrated a rapid increase in AI capability and applicability. The advanced systems are used in a wide spectrum of contexts ranging from our daily lives to safety-critical systems. Although these intelligent systems can benefit society, there have been some unfortunate accidents that occurred because of the difficulty in using such systems (e.g., autonomous car [30, 43], aviation [38]). These difficulties underscore the need to change the way that humans work with AI systems. A key part of this change is the transition from AI system as tool to AI system as teammate [24]. With the advent of the fifth industrial revolution, humans and machines are expected to work together by leveraging the strengths of each other [39]. Research work has been done in various domains, including AI, human factors, and Human-Computer Interaction (HCI) to better understand human-agent team dynamics and how to improve teamwork [10, 12, 24]. Some research aims to apply knowledge from human-human teams, including social exchange theory to teamwork between


Manuscript submitted to ACM

humans and non-human entities [8, 9, 35]. Trust is also one of the key research areas aiming at improving teamwork between humans and AI systems; however, its multifaceted nature makes it difficult to understand the effect of trust on human-agent teamwork [17, 28]. It is an open question whether lower levels of trust lead to lower team performance or low team performance results in lower levels of trust [34]. To fulfill effective human-agent teams, it is critical to gain a sound understanding of what contributes to trust, what is the effect of trust on human-agent teamwork, and how trust is measured and shaped.

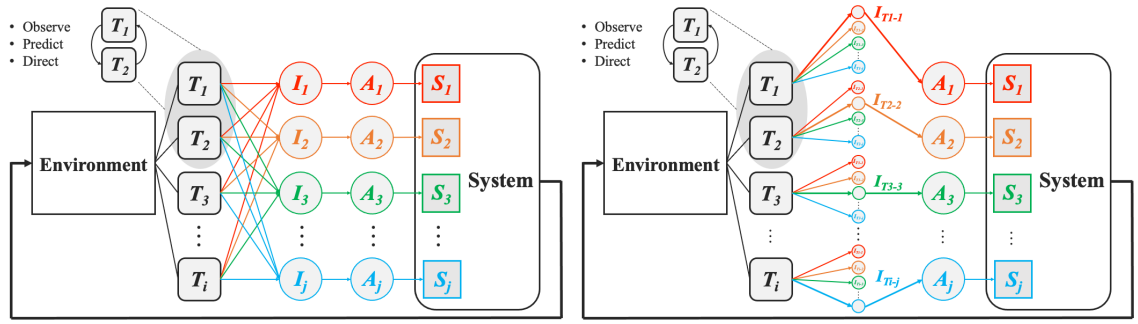
This paper is intended to serve as a foundation for better understanding and trust and teamwork in human-agent teaming. First, we propose a taxonomy classifying human-agent teams into two types: compositional and non-compositional control teams. We explain the two types of human-agent teams in relation to some examples in research and real-world contexts. Second, we propose a compositional control trust model, which is built on trust models by Mayer [33] and Johnson and Bradshaw [20]. This paper concludes with a discussion on our future work centered around our proposed compositional control trust model using a compositional control game used in our original experiment [36].

## 2 TYPES OF HUMAN-AGENT TEAMS

Key features of a team are generally understood to include members, common goals, interdependence, and roles/functions although there is not a consistent use of these terms in the literature [13, 20, 25, 27, 42]. In this paper, we define a team as follows: a team consists of multiple entities who engage in activities interdependently to achieve goals while performing their roles in a dynamic, timely, and context-specific manner. Also, we refer to human-agent team as a team with humans and non-human entities with a certain degree of agency that are capable of sensing environments, collecting information, communicating with people, taking actions, learning, and evolving [11]. The term "human-agent team" is used in this paper to refer to human-AI/human-automation teaming as well. Teams can be characterized by a number of different factors including skill, authority differentiation and experience of working as a team [18], the amount of communication and task interdependence [41], and perceived human-likeness, autonomy, and interdependence [29].

We classify human-agent teams into two types: *compositional control* and *non-compositional control* human-agent teams, which indicates that ways that control is exerted by the team. In the compositional control team setting, team members simultaneously possess the same action channels, and the effect on the system is dependent on all of the entities' inputs (Figure 1a). As real-life examples, automobile driving assistance, and aircraft and boat autopilots fall into the compositional control team category. Each team member in a human-agent team executes the Observe, Orient, Decide, and Act (OODA) loop process [4] where situation awareness [12] and interdependence [22] play a critical role. In other words, each team member ($T_i$) observes other teammates as well as the working environment, comprehends the collected information, decides what to do, and makes an action. In the compositional control team (Figure 1a), all the team members possess the same action channels (i.e., $I_j$ and $A_j$) at the same time and their individual actions are combined to create one action effect. Suppose that the team works on a system consisting of subsystems ($S_j$) that are controlled by actions ($A_j$). Each team member's input ($I_j$) contributes to each action ($A_j$), and then each action affects the whole system, leading to a new state (i.e., environment).

In contrast, non-compositional control teams can have different action channels that affect the system independently; each entity can have independent effects on the system (Figure 1b). Non-compositional control teams in a real-world setting include intelligence analysis in cybersecurity, cyber-physical-human systems in space missions, and package management work with autonomous warehouse robots. For example, a team with an intelligence analyst and an AI system might work together on a variety of tasks such as data collection via queries, extracting key information, and




(a) Compositional Control Teams (b) Non-Compositional Control Teams

Fig. 1. Schematic of how each team member's action affects the system in (a) compositional control and (b) non-compositional control teams. Team members ($T_i$) can be either humans or agents. $I_j$ is an input determined by team member(s), resulting in an action ($A_j$) which affects the corresponding subsystem ($S_j$) and ultimately the whole system. Also, team members ($T_i$) can observe, predict, and direct each other [22] (highlighted with a gray oval); for the sake of readability, interactions between all the team members ($T_j$) are not depicted.

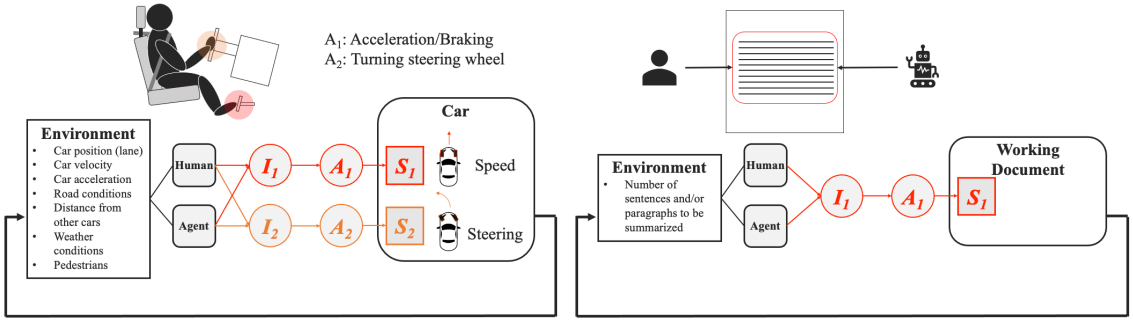
generating a summary report. The following subsections present some examples of compositional and non-compositional control teams from research studies employing simple games as microworld domains.

## 2.1 Compositional Control Teams

*2.1.1 Real-World Examples.* Figure 2 presents two real-world examples of compositional control human-agent teams: automated driving assistance and collaborative text editing (text summarization tasks). Figure 2a illustrates a driving scenario where a human drives with driving supporting features such as brake/acceleration and steering (i.e., SAE Level 0-2 [40]). In this scenario, both human driver and automated driving assistance system share the same action capabilities (i.e., brake/acceleration and steering) over the course of driving, and the effect on the car movement is dependent on inputs from both entities. Researchers have investigated how to accomplish collaborative steering control between a human driver and automation, for instance, by focusing on haptic-oriented communication [15, 32, 37].

Whereas the automated driving assistance scenario employs continuous inputs, compositional control human-agent teams can also take discrete form of inputs. A collaborative text editing task (e.g., summarizing an article with an agent assistant [7]) is one real-world example of such type of human-agent teams (Figure 2b). An agent assistant provides suggestions in a real-time manner while a human writer summarizes. Then, inputs from both entities are used to determine the team's action. Other examples of discrete input-oriented compositional control teams include AI-supported diagnosis [45].

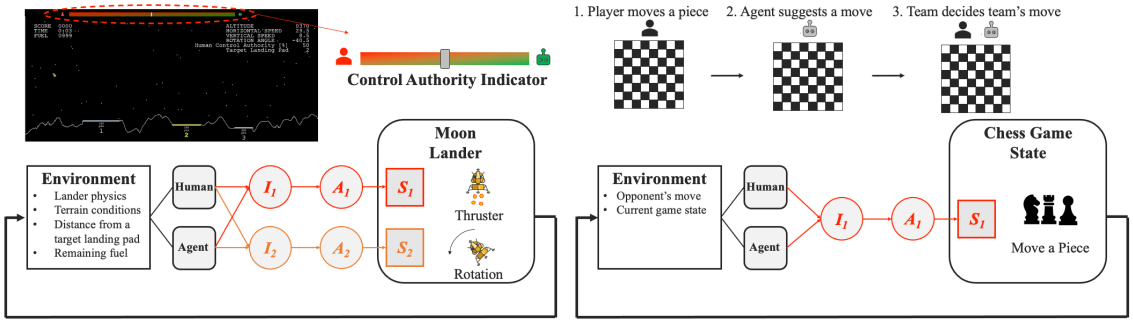
*2.1.2 Microworld Examples.* A moon lander game (Figure 3a) used in our original study [36] is one of the examples of compositional control teams, which can serve as an analog to automobile driving assistance. In the moon lander game, a human player is asked to work together with an agent on a moon lander maneuvering task. The human player and the agent share two action channels: a thruster control to change the lander's speed and a rotation control to change the lander's attitude. The human player can customize how to compose his/her input with the input of the agent to determine the input to the action to be executed. The control authority indicator (i.e., a horizontal bar with a color gradient with red and green colors in Figure 3a) displays the current control mix ratio ranging from manual control to




(a) Automated Driving Assistance [2, 40] (b) Collaborating with Agent on Text Summarizing Task [7]

Fig. 2. Real-World Examples of Compositional Control Human-Agent Teams

fully automated configuration, which was inspired by an interaction strategy called "Horse-Mode" or H-Mode [15]. The 50-50 middle-ground configurations allow the player to share the lander control with the determined ratio. Suppose the player sets the ratio at 75% of human control and 25% of agent control and rotates the lander in a clockwise direction while the agent provides an opposite direction input. In this case, the rotational inputs from the player and the agent are conflicting each other. Yet, the lander rotates in a clockwise direction with 50% of the magnitude of the full rotational speed because of the more dominant human's control authority level. With a split configuration, both human player and agent have 50% of lander control. If the player rotates the lander in a clockwise direction while the agent provides a counterclockwise direction input, then both offset each other, and the team's lander rotational input becomes zero.




(a) Moon Lander Game [36] (b) Chess Game [46]

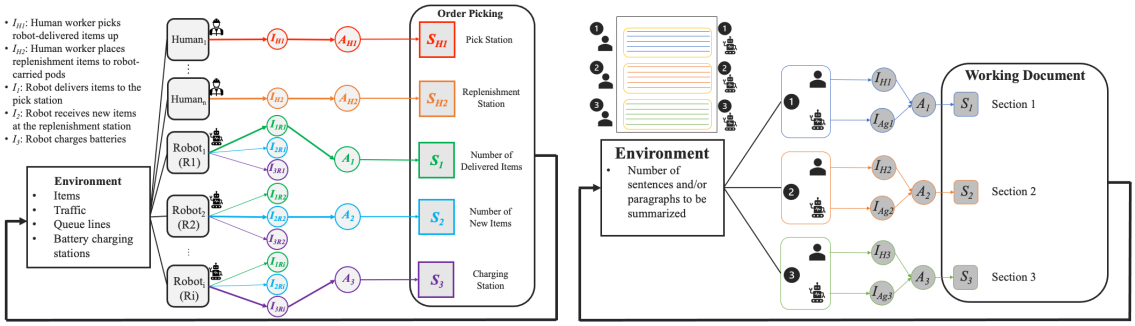
Fig. 3. Examples of Compositional Control Human-Agent Teams

Whereas the moon lander game employs continuous inputs, inputs used in compositional control human-agent teams can also take discrete form, and a chess game with an agent teammate [46] is one of the simple examples of discrete input-oriented compositional control human-agent teams (Figure 3b). In the collaborative chess game, first, a human player indicates how to move a piece. Next, an agent teammate offers a recommendation, including an alternative piece movement. Finally, the team decides the team's action. In this case, the team explores multiple alternative options and reaches an agreement on the team's piece move by integrating inputs from the player and the agent.

## 2.2 Non-Compositional Control Teams

*2.2.1 Real-World Examples.* Figure 4 shows two real-world examples of non-compositional control human-agent teams. Autonomous robots are employed in a warehouse to facilitate the order picking process, and Figure 4a presents an example of a warehouse system [3]. Human workers are responsible for picking robot-delivered items up at a pick station or placing items into cases carried by a robot at a replenishment station. Deployed autonomous robots have three actions: (i) deliver items to the pick station, (ii) receive new items at the replenishment station, and (iii) charge a battery. Although the all deployed robots have the three action capabilities, their individual inputs are not consolidated and independently affect the order picking process.

There may be a non-compositional control team setting in a collaborative text editing task, where a set of human-agent dyads are working on a common document (Figure 4b). Although each human-agent team is operated in a compositional control team manner (i.e., Figure 2b), each dyad is working on a different section, meaning that each team independently affects the same working document.

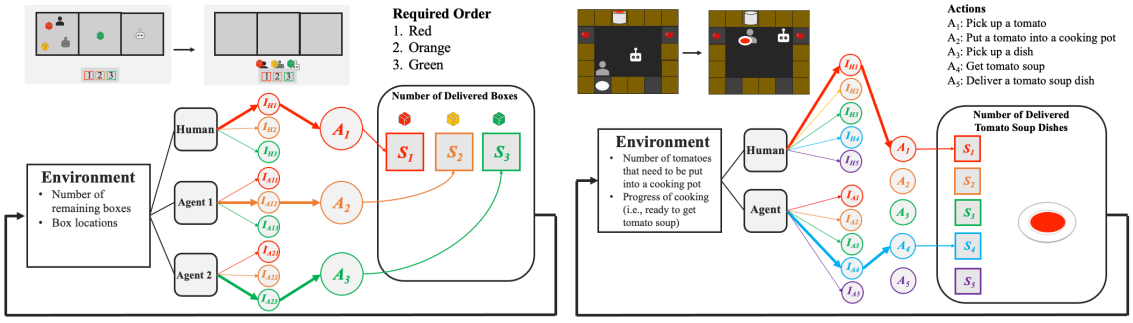



(a) Warehouse Order Picking [3] (b) Text Summarizing Task with Human-Agent Dyads

Fig. 4. Real-World Examples of Non-Compositional Control Human-Agent Teams

*2.2.2 Microworld Examples.* The Block World for Teams (BW4T) testbed [23] (Figure 5a) is one of the game environments to investigate human-agent teamwork (e.g.,[5, 16]), which falls into the non-compositional control team category. In the BW4T testbed, team members deliver colored boxes and drop them off at a designated area. The colored boxes are randomly located in rooms, and each room is partitioned by walls. Each entity can carry one colored box at a time, and the team is required to deliver a colored box based on a pre-determined sequence of colors. In Figure 5a, the team is required to drop three boxes off in order of red, orange, and green boxes.

A cooking game used for research on human-agent coordination [6, 44] is another example of non-compositional control teams (Figure 5b). In the cooking game, a human player and an agent are asked to cook tomato soup and deliver tomato soup dishes to a designated counter. To cook tomato soup, three tomatoes need to be put in a cooking pot. Once three tomatoes are placed in a cooking pot, the cooking process starts, and it takes some time in order for tomato soup to be ready to pick up. Upon the completion of the cooking process, one of them needs to pick up a tomato soup dish and deliver it to the designated counter. Each of them can pick up, drop off, or deliver one item at a time, and each action affects the whole system independently.




(a) BW4T [23] (b) Cooking Game [44]

Fig. 5. Examples of Non-Compositional Control Human-Agent Teams

## 2.3 Key Distinctions between Compositional and Non-Compositional Control Teams

*2.3.1 Effect of Adding Team Members.* One of the key distinctions between compositional control and non-compositional control teams is that the effect of increasing the number of team members on the system. A compositional control human-agent team has one single input for each action channel regardless of the total number of team members. With the increased number of team members, the team may improve their input quality although the efficiency is not significantly changed. In contrast, adding team members helps non-compositional control teams improve their task efficiency because of the fact that each team member can affect the system independently.

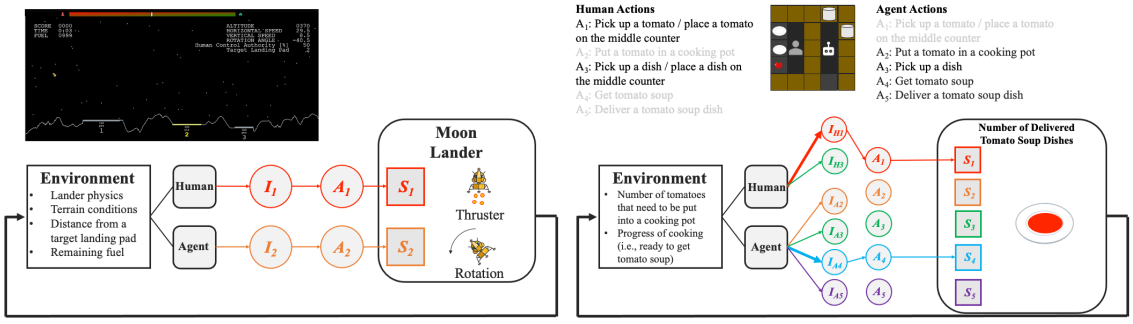
*2.3.2 Hard Interdependence in Non-Compositional Control Teams.* Another key distinction is that there is a situation where hard interdependence [22] is required to complete team's goal in non-compositional control teams. Figure 6 presents examples of non-compositional control teams with hard interdependence. Figure 6a shows the moon lander game; however, the player has only the capability of engaging the thruster input while the agent can only rotate the lander, meaning that both do not have the same action channels (i.e., non-compositional control). In this scenario, it is impossible for the team to successfully land without any inputs from the partner, requiring hard interdependence.

Figure 6b shows the cooking game; yet, the layout of the kitchen is different from Figure 5b. The layout is set for simulating forced coordination [44], where a team member in the right kitchen has to wait for inputs from the partner on the left side (i.e., pass a tomato or a dish), requiring hard interdependence.

## 3 TRUST, RELIANCE, AND INTENT

Previous research on the interactions between trust and reliance in human-agent teams has found that humans develop trust in their AI teammates through understanding how the agent works [21], gaining experience with the agent [31], or by reputation [19]. However, the distinction between compositional and non-compositional control human-agent teams can have implications for the determination of trust of and reliance on team mates. In this section, we propose an extension to the model presented in [20] to apply specifically to compositional control teams.

In compositional control teams, the human operator can use their own task model to estimate why the agent was built (it's *purpose*) and how it works (it's *process* [28]). The unknown in the determination of trust and the decision to rely on the agent is understanding in which contexts the agent performs well (it's *performance*). Without the experience of using the agent, the human operator can only guess at the performance capabilities of the agent, resulting in the decision to




(a) Hard Interdependence in Moon Lander Game (b) Hard Interdependence in Cooking Game [44]

Fig. 6. Examples of Hard Interdependence in Non-Compositional Control Human-Agent Teams

use the agent being based on user propensity and preference, the recognition of the demands of the task environment, and the ability to determine the real-time performance of the agent. For example, having the initial attempts at enabling driving assistance on smooth, straight roadways will make deviations from the operator's intended route (i.e., straight and consistent lane position) evident, making it easier to recognize good and bad agent performance. Trust in the agent is calibrated through continued experiences in differing contexts, enabling the human user to determine when and in which contexts the agent teammate can be relied on to enhance team performance. Reliance depends both on trust and on the risk-reward of relying on the AI teammate when task loads increase.

The trust model proposed by [33] was developed for human-human teams, but has been widely cited and applied in analysis of human-agent teams [20, 26]. A key contribution of the model is recognition of the importance of making the decision to enter into a risk-taking relationship (RTR) between team mates to accomplish a task. The trust held in team mates is a key component in making this decision. The decision to enter into an RTR between team mates indicates the reliance of team mates on each other to accomplish a task. Reliance must be calibrated to ensure that the team neither over-relies nor under-relies on team mates. In this model, trust is the perception of one agent about another agent, and reliance is the interdependence between agents in an RTR to work to achieve a goal. In other words, "... trust is an attitude, reliance is a behavior." [28, p.53].

Johnson and Bradshaw [20] extend Mayer et al's model to highlight the importance of alternatives in making risk assessments in determining the best way to accomplish work. They introduce "Perceived Risk/Reward" and "Activity Context" to refine the components of the decision to enter into an RTR.

The trustor factors that guide the estimation of trust and the decision to rely on a team mate are the propensity of the trustor to trust and the preference of the trustor to trust. Propensity can be viewed as a personality trait that reflects how cautious a trustor is, in general, regardless of the context. The preference to trust acknowledges that in different situations, the preference of the trustor to enter an RTR may change. For example, the trustor may not want to rely on another team mate to perform an task in order to gain more personal experience or because the task is enjoyable.

Johnson and Bradshaw [20] extended the Mayer et al's model by including the context in which the activity is taking place and the perceived risk/reward of the activity. These situational factors affect the assessment of both the degree of trust and benefits of entering into an RTR.

The trustee factors present in both the Mayer and Johnson and Bradshaw models include the benevolence and integrity of the trustee, and their perceived ability. The benevolence is "the extent to which a trustee is believed to want

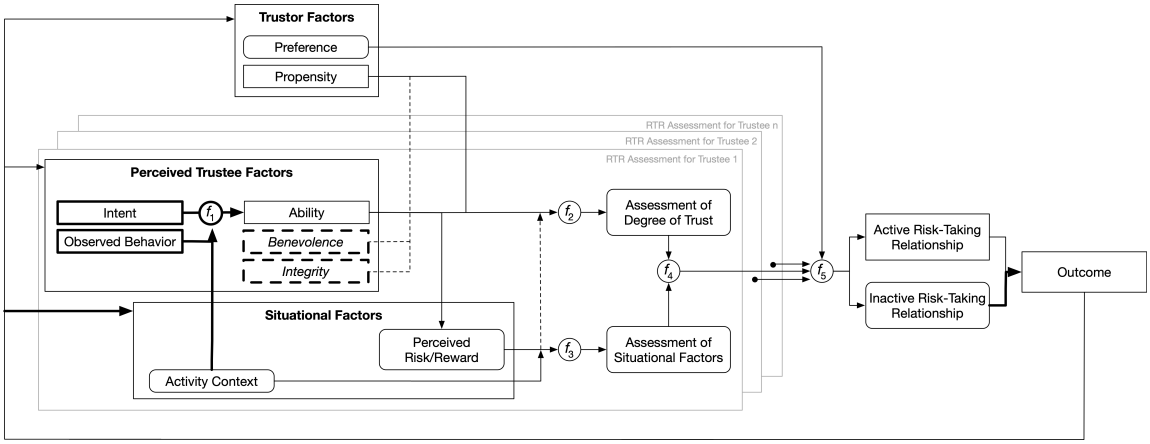



Fig. 7. Proposed compositional control trust model refined from models by Mayer [33] and Johnson and Bradshaw [20]. Mayer contributions are rectangles, Johnson and Bradshaw are rounded rectangles and proposed changes are in bold.

to do good *to the trustor*, apart from an egocentric profit motive."[33, p.718]. Integrity is the perception of the trustee that they operate under a set of rules and guidelines that are acceptable to the trustor. These perceptions of the trustee combine with an estimate of the trustee's competence and ability to perform tasks in the activity context as input to the trustor's overall assessment of trust in the trustee.

### 3.1 Enhancements to the Model

Figure 7 shows the enhancements made to the Mayer et al and Johnson and Bradshaw models in bold for additions and in dotted lines for deletions. These changes are driven by the particular characteristics of compositional control teams, and are hoped to be a useful tool for determining trust and reliance for human-agent teams.

The structure of the compositional control team as a group of entities that can take actions and control the system in ways that are transparent to their team mates makes the estimates of benevolence and integrity somewhat redundant; all the team members are working towards a common goal while possessing the same action channel(s). Consider the case of a human driving with automated lane-keeping assistance. Because interactions between driver and agent happen continuously, there are many opportunities for the human to revise their estimates of the agent abilities, which would encompass explicit attempts to cause harm through malevolence or errors through lack of integrity. Because of this, the connections between the trustor factors and these components has also been removed.

The key contribution of our enhancements is in using *intent* and the *observed behavior* seen during execution combined with the activity context to calculate trustee ability (shown as the computation circle $f_1$ in Figure 7). The recognition of ability in this model is the degree of closeness between what the performer's intent is and what they actually did. For example, if the intent of the driving agent is to maintain a constant two-foot distance from the road centerline, we could compare the actual distances from the road centerline as the driving agent is performing to estimate the agent ability. The trustor propensity to trust determines the closeness necessary for a high ability rating from the trustor. Those having a lower propensity to trust require there to be few deviations from the specified intent in order to get a high score. Trustors with a higher propensity to trust will accept larger deviations.

A second enhancement is the feedback of the outcome of the RTR decision regardless if the decision was to enter into the RTR or not. In both of these circumstances, the outcome of the decision will have an effect on the overall system performance, which can be used to determine if adjustments need to be made to the assessment factors, particularly to the perceived risk/reward assessment. As performance suffers in a particular activity context, the threshold for entering into an RTR may decrease, enabling additional task support.

The key functions to be computed in this model are:

$f_1$: This function computes the ability of the trustee based on the communication of the trustee's intent and the subsequent observation of the actual behavior. The activity context is used to influence the ability based on the context the actor is operating in.
$f_2$: This function computes the assessment of trust based on the ability of the trustee and the trustor's propensity.
$f_3$: This function computes the expected gain or loss from working with the current team assignments versus adding or removing additional team mates.
$f_4$: This function combines the assessment of trust with the assessment of the situational factors to produce an overall measure of the ability of the trustee to perform usefully with respect to the current goals.
$f_5$: this function either selects the best output of $f_4$ for any trustee when the trustor has no preference, or responds to the trustor preference by not entering into any RTR if the trustor wants to perform the task themselves, or entering into the trustor preferred RTR.

## 4 FUTURE WORK

In our original experiment using the compositional moon lander control game [36], we attempted to investigate patterns of interactions of effective human-agent teams. Although our preliminary results are inconclusive, the experimental data implied that participants who better understood when to rely on an agent teammate exhibited higher team performance. As part of our future work, we will conduct another experiment using the moon lander game environment and attempt to validate the proposed trust model for compositional control teams.

## 5 CONCLUSION

With the rise of the fifth industrial revolution, humans and increasingly more sophisticated intelligence systems are expected to work together as a team [39]. Although trust is considered a key enabler to improve human-agent teams, more research is required to better understand the mechanism of trust in the context of human-agent teamwork. This paper was intended to establish a cornerstone for further investigating trust and human-agent teamwork. First, we addressed two types of human-agent teams: compositional control and non-compositional control teams. Then, we proposed a trust model of compositional control human-agent teams, which is built on the trust models by Mayer [33] and Johnson and Bradshaw [20]. We will investigate our proposed trust model using our moon lander game environment [36], and we believe that our proposed model could help us better understand trust, reliance, and teamwork in compositional control human-agent teams.

## ACKNOWLEDGMENTS

Any opinions, findings and conclusions or recommendations presented in this material are those of the authors and do not necessarily reflect the views of the Department of Defense.

## REFERENCES


[1] Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. 2020. Towards a Human-like Open-Domain Chatbot. *CoRR* abs/2001.09977 (2020). arXiv:2001.09977 https://arxiv.org/abs/2001.09977

[2] National Highway Traffic Safety Administration. 2022. Levels of Automation. https://www.nhtsa.gov/sites/nhtsa.gov/files/2022-05/Level-of-Automation-052522-tag.pdf Last accessed 11 April 2023.

[3] Ali Bolu and Ömer Korçak. 2021. Adaptive task planning for multi-robot smart warehouse. *IEEE Access* 9 (2021), 27346–27358.

[4] John R Boyd. 1996. The essence of winning and losing. *Unpublished lecture notes* 12, 23 (1996), 123–125.

[5] Abhizna Butchibabu, Christopher Sparano-Huiban, Liz Sonenberg, and Julie Shah. 2016. Implicit coordination strategies for effective team communication. *Human factors* 58, 4 (2016), 595–610.

[6] Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems* 32 (2019).

[7] RUIJIA CHENG, ALISON SMITH-RENNER, KE ZHANG, JOEL R TETREAULT, and JAIMES ALEJANDRO. 2022. Trust and Reliance in Human-AI Collaborative Text Summarization. (2022).

[8] Erin K Chiou, Mustafa Demir, Verica Buchanan, Christopher C Corral, Mica R Endsley, Glenn J Lematta, Nancy J Cooke, and Nathan J McNeese. 2021. Towards Human–Robot Teaming: Tradeoffs of Explanation-Based Communication Strategies in a Virtual Search and Rescue Task. *International Journal of Social Robotics* (2021), 1–20.

[9] Erin K Chiou, John D Lee, and Tianshuo Su. 2019. Negotiated and reciprocal exchange structures in human-agent cooperation. *Computers in Human Behavior* 90 (2019), 288–297.

[10] Nazli Cila. 2022. Designing Human-Agent Collaborations: Commitment, responsiveness, and support. In *CHI Conference on Human Factors in Computing Systems*. 1–18.

[11] Nazli Cila, Iskander Smit, Elisa Giaccardi, and Ben Kröse. 2017. Products as agents: Metaphors for designing the products of the IoT age. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 448–459.

[12] Mica R Endsley. 2022. Supporting Human-AI Teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior* (2022), 107574.

[13] Mica R Endsley, Betty Bolté, and Debra G Jones. 2003. *Designing for situation awareness: An approach to user-centered design*. CRC press.

[14] Meta Fundamental AI Research Diplomacy Team (FAIR)†, Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. 2022. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science* 378, 6624 (2022), 1067–1074.

[15] Frank Flemisch, Johann Kelsch, Christan Löper, Anna Schieben, Julian Schindler, and Matthias Heesen. 2008. Cooperative control and active interfaces for vehicle assitsance and automation. (2008).

[16] Maaike Harbers, Jeffrey M Bradshaw, Matthew Johnson, Paul Feltovich, Karel van den Bosch, and John-Jules Meyer. 2011. Explanation and coordination in human-agent teams: a study in the BW4T testbed. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Vol. 3. IEEE, 17–20.

[17] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.

[18] John R Hollenbeck, Bianca Beersma, and Maartje E Schouten. 2012. Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. *Academy of Management Review* 37, 1 (2012), 82–106.

[19] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. 2006. Certified Reputation: How an Agent Can Trust a Stranger. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (Hakodate, Japan) *(AAMAS '06)*. Association for Computing Machinery, New York, NY, USA, 1217–1224. https://doi.org/10.1145/1160633.1160854

[20] Matthew Johnson and Jeffrey M Bradshaw. 2021. How Interdependence Explains the World of Teamwork. In *Engineering Artificially Intelligent Systems*. Springer, 122–146.

[21] Matthew Johnson and Jeffrey M Bradshaw. 2021. The role of interdependence in trust. In *Trust in human-robot interaction*. Elsevier, 379–403.

[22] Matthew Johnson, Jeffrey M Bradshaw, Paul J Feltovich, Catholijn M Jonker, M Birna Van Riemsdijk, and Maarten Sierhuis. 2014. Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction* 3, 1 (2014), 43–69.

[23] Matthew Johnson, Catholijn M Jonker, M Birna van Riemsdijk, Paul J Feltovich, and Jeffrey M Bradshaw. 2009. Joint Activity Testbed: Blocks World for Teams (BW4T).. In *ESAW*, Vol. 9. Springer, 254–256.

[24] Matthew Johnson and Alonso Vera. 2019. No AI is an island: the case for teaming intelligence. *AI magazine* 40, 1 (2019), 16–28.

[25] Jon R Katzenbach and Douglas K Smith. 2015. *The wisdom of teams: Creating the high-performance organization*. Harvard Business Review Press.

[26] Spencer C. Kohn, Ewart J. de Visser, Eva Wiese, Yi-Ching Lee, and Tyler H. Shaw. 2021. Measurement of Trust in Automation: A Narrative Review and Reference Guide. *Frontiers in Psychology* 12 (2021), 604977. https://doi.org/10.3389/fpsyg.2021.604977

[27] Carl Larson, Carl E Larson, and Frank MJ LaFasto. 1989. *Teamwork: What must go right/what can go wrong*. Vol. 10. Sage.

[28] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[29] Joseph B Lyons, Katia Sycara, Michael Lewis, and August Capiola. 2021. Human–autonomy teaming: Definitions, debates, and directions. *Frontiers in Psychology* 12 (2021), 589585.

[30] Carl Macrae. 2021. Learning from the Failure of Autonomous and Intelligent Systems: Accidents, Safety, and Sociotechnical Sources of Risk. *Risk Analysis* 42 (2021).

[31] P. Madhavan and D. A. Wiegmann. 2007. Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science* 8, 4 (2007), 277–301. https://doi.org/10.1080/14639220500337708 arXiv:https://doi.org/10.1080/14639220500337708

[32] Mauricio Marcano, Sergio Díaz, Joshué Pérez, and Eloy Irigoyen. 2020. A review of shared control for automated vehicles: Theory and applications. *IEEE Transactions on Human-Machine Systems* 50, 6 (2020), 475–491.

[33] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

[34] Nathan J McNeese, Mustafa Demir, Erin K Chiou, and Nancy J Cooke. 2021. Trust and team performance in human–autonomy teaming. *International Journal of Electronic Commerce* 25, 1 (2021), 51–72.

[35] Nathan J McNeese, Mustafa Demir, Nancy J Cooke, and Christopher Myers. 2018. Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors* 60, 2 (2018), 262–273.

[36] Kazuhiko Momose, Troy R Weekes, Rahul Mehta, Cameron Wright, Josias Moukpe, and Thomas C Eskridge. 2023. Patterns of Effective Human-Agent Teams. In *In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. 1–13.

[37] Tomohiro Nakade, Robert Fuchs, Hannes Bleuler, and Jürg Schiffmann. 2023. Haptics based multi-level collaborative steering control for automated driving. *Communications Engineering* 2, 1 (2023), 2.

[38] Jack Nicas, Natalie Kitroeff, David Gelles, and James Glanz. 2019. Boeing Built Deadly Assumptions Into 737 Max, Blind to a Late Design Change. *The New York Times* (2019). https://www.nytimes.com/2019/06/01/business/boeing-737-max-crash.html

[39] Stephanie M Noble, Martin Mende, Dhruv Grewal, and A Parasuraman. 2022. The Fifth Industrial Revolution: How harmonious human–machine collaboration is triggering a retail and service [r] evolution. *Journal of Retailing* 98, 2 (2022), 199–208.

[40] Society of Automotive Engineers (SAE) International. 2021. SAE Levels of Driving Automation™ Refined for Clarity and International Audience. https://www.sae.org/binaries/content/assets/cm/content/blog/sae-j3016-visual-chart_5.3.21.pdf Last accessed 11 April 2023.

[41] Elizabeth Phillips, Kristin E Schaefer, Deborah R Billings, Florian Jentsch, and Peter A Hancock. 2016. Human-animal teams as an analog for future human-robot teams: Influencing design and fostering trust. *Journal of Human-Robot Interaction* 5, 1 (2016), 100–125.

[42] Eduardo Salas, Terry L Dickinson, Sharolyn A Converse, and Scott I Tannenbaum. 1992. Toward an understanding of team performance and training. (1992).

[43] Faiz Siddiqui, Rachel Lerman, and Jeremy B. Merrill. 2022. Telsas running Autopilot involved in 273 crashes reported since last year. *The Washington Post* (2022). https://www.washingtonpost.com/technology/2022/06/15/tesla-autopilot-crashes/

[44] DJ Strouse, Kevin McKee, Matt Botvinick, Edward Hughes, and Richard Everett. 2021. Collaborating with humans without human data. *Advances in Neural Information Processing Systems* 34 (2021), 14502–14515.

[45] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, John Paoli, Susana Puig, Cliff Rosendahl, H. Peter Soyer, Iris Zalaudek, and Harald Kittler. 2020. Human–computer collaboration for skin cancer recognition. *Nature Medicine* 26, 8 (2020), 1229–1234. https://doi.org/10.1038/s41591-020-0942-0

[46] Guanglu Zhang, Leah Chong, Kenneth Kotovsky, and Jonathan Cagan. 2023. Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation. *Computers in Human Behavior* 139 (2023), 107536.