

Preliminary Insights into Enhancing Human-Robot Teamwork

Kazuhiko Momose, Troy R. Weekes, and Thomas C. Eskridge

Department of Human-Centered Design, Florida Institute of Technology

ABSTRACT

Advancements in robotics, artificial intelligence, and other automation have highlighted the need for humans to work together with machines in a more flexible and collaborative fashion than previously possible. To formulate effective human-robot teams, it is critical to understand which factors play important roles in enhancing human-robot teamwork. To gain preliminary insights into key factors of effective human-robot teams, we carried out an experiment using an enhanced version of the “Lunar Lander” game, where the goal is to safely land a spacecraft on the moon’s surface in concert with an AI teammate. We specifically attempted to observe some patterns of communication across high-performing teams in the experiment. Due to the limited number of experimental participants, the results of the experiment did not definitively identify the factors that account for effective teams. Instead, the experiment indicated potential avenues to further investigate, including intent-oriented communication and trust in teams with human and non-human entities. This paper presents findings from the experiment and discusses future work to extend the scope of the study to include teleoperation of unmanned vehicles with communication delay.

Keywords: Human-Automation Teamwork, Trust in Automation.

1. Introduction

In 1951, Paul Fitts proposed a list comparing what people are better at and what machines are better at, also known as MABA-MABA [1]. Since the introduction of Fitts’ list, machine capabilities have advanced and are now more capable of handling a wide spectrum of tasks at the same level as humans or even better [2, 3]. Technological advancement in the form of AI, robots, and other types of automated agents highlight the need for humans to work together with machines in a more flexible and collaborative fashion than previously possible.

The tasks mentioned in Fitts’ list have been considered to be allocated either to human or machine, with few or no substitutions of performers expected during the course of an interaction. However, some human-automation teams can now tailor their collaboration by substituting task performers based on the performers’ capabilities, environmental conditions, etc [4, 5, 6]. In such a flexible collaborative relationship, it is imperative for the human-agent team to effectively assign each performer’s function in a timely manner during the course of a task to maximize overall team performance.

Communication is a key factor in teams, including human-human teams as well as teams with human and non-human entities. In the case of a human-human team, team members communicate with each other via verbal, facial, and/or other communication methods. Researchers have studied the traits of an effective team (e.g.,

[7, 8, 9]). Pentland [10] studied the significant factors in building a productive team by collecting data regarding team members’ communication behaviors from various types of teams. The data showed consistent patterns of communication across the teams that exhibited high productivity levels. Understanding trends of productive teams allows practitioners to develop strategies for enhancing overall team performance.

Research in human automation teams have investigated whether the patterns of communication of effective human teams could be observed in teams with human and non-human entities (e.g., [11, 12, 13, 14, 15]). For example, Chiou et al. [13] adopted social exchange theory in a human-agent cooperative scenario and investigated how two different social exchange structures (i.e., negotiated and reciprocal exchange) could affect joint performance of the human-agent team.

In this paper, we present our preliminary investigation aiming at gaining insights into patterns of communication for designing human-AI interaction systems that maximize team performance. We believe that an effective human-AI team should exhibit certain consistent patterns of communication during task interactions. To test this, we carried out a study using the “Lunar Lander” video game which asked participants to safely land a spacecraft on the moon’s surface while working in concert with AI agents executing different collaboration strategies. The second section of this paper describes related work. The third section addresses our hypotheses and experimental design of the study. Then, we present results from the study and discuss implications of patterns of communication between the humans and AI agents. The final section concludes with limitations of the study and suggests potential directions for future work.

2. Related Work

2.1 Levels of Automation and Dynamic Function Allocation

Fitts [1] and others [16, 17, 18] have proposed taxonomies or scales that describe the levels of automation (LOA) of a human-automation team. These teams can adapt to one of the automation levels during a cooperative task, and it is critical to switch each entity’s function in a timely and appropriate fashion for enhancing overall team performance. This requires a sound understanding of the effects of each LOA on human-automation interaction and how to facilitate dynamic function allocation for maximizing overall performance. Several strategies for adjusting the LOA have been investigated: (1) adaptive, (2) adaptable, and (3) hybrid functional allocation schemes (e.g., [19, 20, 21, 22, 4, 5]). One of the contributions to control systems and interface design for human-automation interaction is the Horse-Mode (H-Mode) interaction strategy [6]. In this

work, a human operator communicates with an automated agent via a haptic device (e.g., a force-feedback joystick), allowing the operator to convey their intention to the agent and detect the agent’s intention. The interaction strategy was either a “tight rein” mode or a “loose rein” mode. With the tight rein mode, the operator had a more dominant control authority whereas the loose rein mode provided the automated agent with a higher degree of control authority. Results of empirical studies indicated the usefulness of the H-Mode for keeping a human in the automation loop [23, 24].

2.2 Team Communication

Communication is a key factor in teams, including teams with humans and non-human entities as well as human teams, and it is promising to dissect process measures regarding communication in order to assess how effectively teams work together [8]. In the case of human teams, team members communicate with each other via verbal, facial, and/or other communication modalities, resulting in resilient, robust, and efficient communication [25]. In contrast, although both verbal and non-verbal communication approaches are available [26, 25], human-robot/AI teams are required to exchange information with less flexible communication methods, highlighting the importance of understanding how teams can effectively communicate within their members. To better understand communication of human-robot/AI teams, researchers have been attempting to apply knowledge of human teams, including social exchange theory, to the context of teams with human and non-human entities [25].

One of the patterns of communication in effective human teams is that they tend to “push” information rather than “pull” information, and research work has been done to investigate whether such a trend could be observed in human-robot teams [15, 14] and human-autonomy teams [11, 27]. Other precedents examined the effects of information types [28] and explicit and/or implicit communication on human-agent teamwork [29, 30]. These precedents provided fruitful insights into the applicability of findings from effective human teams to human-robot/AI teamwork and suggested considerations for further investigation.

2.3 Trust in Automation

Lee and See defined trust as “*the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability*” (p.51) [31] and argued that good calibration, high resolution, and high specificity are key enablers for achieving appropriate trust. In order for humans to achieve high specificity, they must possess a solid understanding of specific capabilities of automation systems that may fluctuate over time during the interaction, which could prevent humans from misuse and/or disuse of their automated teammates. Research work has been done to identify factors influencing trust in automation [31, 32, 33, 34], and researchers have investigated how such influencing factors could affect human trust in automation by employing different assessment methods of trust, including subjective and objective measures of trust. Reviews [35, 33, 36, 37] revealed a tendency to use subjective measures of trust.

3. Methodology

Our study is based on contributions made by Tan et al. [38] and uses a customized lunar lander game to investigate how human operators interact with agents of differing capabilities to achieve high performance in different collaboration conditions. The goal of the human-agent team is to safely land a spacecraft within a designated

landing zone on the moon surface (Figure 1). The following subsections address the experiment design, including development of the custom lunar lander game environment and experiment procedures.

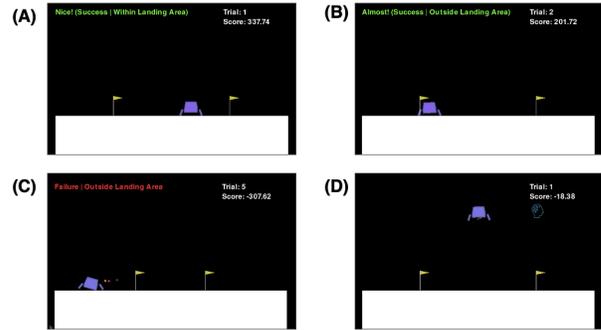


Figure 1. (A) a successful landing case, (B) a partially successful landing case, where the lander’s left side leg is outside the landing zone, (C) one of failure modes, where the lander’s base makes contact with the moon surface, and (D) a trial where a participant is playing the game with one of the two agents. Below the trial number and current score, the thinking head icon appears and flashes when the agent is providing inputs.

3.1 Independent Variables

Table 1 lists independent variables in the lunar lander game.

Factor	Levels
Agent Capability	1. More Capable 2. Less Capable
Task Difficulty	1. Easy 2. Medium 3. Hard
Control Scheme	1. Compositional 2. Non-Compositional
Control Input Ratio*	1. Human: 75 % / Agent: 25 % 2. Human: 50 % / Agent: 50 % 3. Human: 25 % / Agent: 75 %

*Control Input Ratio is only for the compositional case

3.1.1 Control Scheme

There are two control schemes: (1) compositional and (2) non-compositional. In the compositional control case, the human and the agent both have control inputs to the main engine and side engines of the lander. Each entity’s input is consolidated based on the control input ratio conditions (see 3.1.2). In contrast, in the non-compositional case, the human has only the control input to the side engines while the agent handles only the main engine.

3.1.2 Control Input Ratio

The control input ratio independent variable is a measure of control authority. It is used in the compositional control scheme trials and determines how to consolidate the inputs from the operator and agent. The 75% Human and 25% Agent condition is considered the *Tight Rein* mode, where human has more dominant control than the agent. The flipped ratio condition (i.e., 25% Human and 75% Agent) is the *Loose Rein* mode, where the agent has more dominant

control. The middle-ground condition splits the total control input into 50% Human and 50% Agent.

3.1.3 Task Difficulty

The goal of the lunar lander game is to safely land the spacecraft on the moon surface. The scoring policy modifies the original OpenAI Gym lunar lander game by adding additional points if the lander lands anywhere in the designated landing zone between the two yellow flags. Whereas the original lunar lander game modulates difficulty by presenting different terrain profiles and initial spacecraft conditions (i.e., position and velocity) in a random manner, this study employs the notion of Index of Difficulty (*ID*), which is derived from the Fitts' law paradigm [39]. Fitts' law states that the time it takes to move a cursor to an area and select it is a function of the distance to the target and the size of the target, and the movement time (*MT*) is calculated as follows:

$$MT = a + b * \log_2(1 + D/W) \quad (1)$$

where *a* and *b* are constants that are dependent on the input device, *D* is the distance to move, and *W* is the width of the target; the Index of Difficulty *ID* is defined by the logarithm term $\log_2(1 + D/W)$. In this study, *D* and *W* correspond to the shortest distance to the designated landing zone and landing pad size, respectively. Figure 2 shows the three levels of task difficulty. The lunar lander consistently starts from the left side, and difficulty is modulated by changing the distance to, and width of, the landing zone.

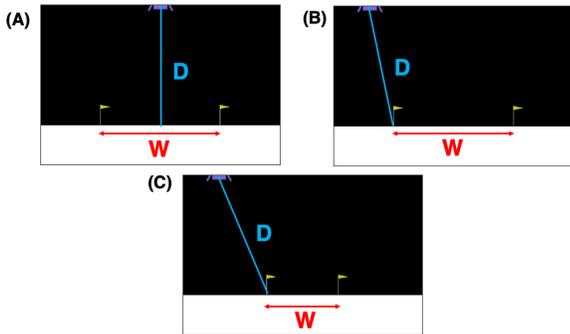


Figure 2. The three task difficulty levels used were: (A) easy case; the lander is set at the center with a wide landing zone, (B) medium case; the lander is set at the left side with a wide landing zone, and (C) hard case; the lander is set at the same position as the medium case, but the landing zone is narrow.

3.1.4 Agent Capability

The human participants interacted with two types of agents: (1) the more capable agent and (2) the less capable agent. The two agents were trained by employing methods developed by Tan et al [38]. The more capable agent was designed to consistently exhibit almost optimal performance levels whereas the less capable agent had a larger variance in capabilities, resulting in lower performance. Figure 3 presents each agent's performance level in a fully automated case (i.e., Human 0% and Agent 100%). This shows consistently higher performance in terms of game score and lower task completion times for the more capable agent than for the less capable agent.

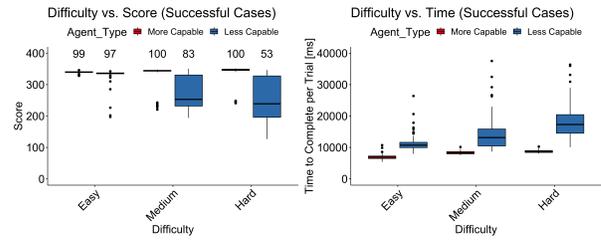


Figure 3. Comparison of agents' performance levels; each agent performs 100 trials across the three task difficulty levels (i.e., 300 trials in total). The left box plot also shows the number of successful landings across the three difficulty levels.

3.2 Dependent Variables

Table 2 lists dependent variables in the lunar lander game.

Table 2. Dependent Variables in the Lunar Lander Game

Category	Measures
Outcome	1. Game Score [-]
	2. Time to Complete [ms]
Process	1. Number of Human Inputs [-]
	2. Number of Disagreements [-]

3.2.1 Game Score

The team gains 100 points if the team safely lands the spacecraft on the moon surface; otherwise, the team receives -100 points (i.e., failure). The team gets an additional 100 points if the lander lands in the designated zone between the two yellow flags (see Figure 1). Whereas the original lunar lander game penalizes the horizontal distance from the center of the landing zone to the center of the lander, the custom environment eliminates that penalty by recognizing an equivalence in landing anywhere between the two flags. If the lander goes outside the game screen, the team receives -200 points. Also, a 50-second time limit was set for each trial, and if the team exceeds the time limit, the team receives -300 points. There were no fuel constraints applied during the trials. Except for these modifications, the custom environment employs the same scoring policy (i.e., each control input to the main engine and side engines results in -0.3 points and -0.03 points respectively). Typically, a successful landing in a landing zone results in 300 to 350 points.

3.2.2 Time to Complete

Time to complete is measured from the beginning of the trial until there is a successful landing or failure. The time to complete is used for normalizing the number of human inputs.

3.2.3 Number of Human Inputs

During a trial, the number of keystrokes that the participant makes is recorded and normalized for each trial using time to complete. The game continuously records participants' keystrokes if they hold down key(s) during a trial. The three arrow keys are used for the lander control: the up-arrow key is for activating the main engine (i.e., moving upward), the left-arrow key is for activating the right side engine (i.e., moving left and rotate in a counterclockwise direction), and vice versa for the right-arrow key. A combined input is recorded in a case where the participant simultaneously presses, for example, the up-arrow key and left- or right-arrow key.

3.2.4 Number of Disagreements

This metric counts the number of situations where a disagreement occurs between the human operator and the agent on the direction of lateral movement during a trial. For instance, if the human provides a right side engine input whereas the agent provides a left side engine input, the consolidated input is counted as a disagreement. This metric is also normalized for each trial using time to complete.

3.3 Experimental Procedures

The experiment consists of briefing, a familiarization session, compositional and non-compositional sessions, and debriefing, which takes approximately 1.5 hours in total. A total of nine volunteers from our internal research group participated in the study. The briefing included a short presentation to provide the study objective, experimental flow, and explain the lunar lander game environment. To mitigate the confounding factor of agent prejudice, the two agents were introduced without any explicit descriptions regarding their capabilities. The participant was asked to evaluate two agents' capabilities by playing the lunar lander game.

On completion of the briefing, the participant is asked to play 10 trials of the lunar lander game without any agent assistance. (i.e., manual case). This allows the participant to better understand the nature of the lunar lander game. Then, the participant observes the two agents' performance when each of the two agents plays 10 trials in a fully automated manner. After the demonstrations of the two agents, the participant has a chance to manually play another 10 trials. In the familiarization session, ten trials consist of four easy cases, three medium cases, and three hard cases, and they are presented in the order from easy to hard.

After the familiarization session, the participant plays the lunar lander game in the compositional control condition. The compositional case session consists of six experimental blocks based on the combinations of the two agent types and three control input ratios. The control input ratios are randomized within each agent type. Each difficulty level has 10 trials (i.e., 30 trials per block), and the three task difficulty levels are randomly presented. Following this, the non-compositional case session consists of two trials where the participant plays the lunar lander game with each agent type while only having the control of the side engines. Each experimental block contains 30 trials, and the difficulty levels are randomly presented in the same fashion as the compositional case session.

In debriefing, the participant is asked to verbally provide subjective feedback on the two agents' capabilities, the two control schemes, and the control input ratios for the compositional case.

4. Data Analysis

For data analysis, R (version 4.0.1) is used [40]. We ran a statistical analysis for each of the two control schemes (i.e., compositional and non-compositional cases) respectively. We performed a MANOVA test with the three factors (i.e., the agent capability, control input ratios, and task difficulty levels) for the compositional case and another MANOVA test for the non-compositional case with the two factors (i.e., the agent capability and task difficulty level). Furthermore, we utilized the participants' subjective feedback on their experience with the two agents for gaining additional insights. The following presents our hypotheses in this study:

- **Score:** As the agent's control input ratio increases, the team will exhibit higher game scores. Also, the team will exhibit better overall performance in the easy level condition when compared to the difficult level condition. These trends will be more pronounced in the more capable agent condition.

- **Human Inputs:** As the agent's control input ratio increases, the number of human inputs will decrease. Also, as the task difficulty level increases, the number of human inputs will increase. These trends will be more pronounced in the more capable agent condition.
- **Conflicted Inputs:** There will be fewer number of disagreements when human collaborates with the more capable agent.
- **Score and Human Inputs:** More variance will be observed in the number of human inputs vs. score scatter plot in the less capable agent case, and less variance along with more consistent and predictable patterns will be observed in the case of the more capable agent.

5. Results

To confirm the normality of the collected data, a Shapiro-Wilk test was performed, and QQ-plots and histograms were generated, suggesting the violation of the normality assumption. Therefore, a MANOVA test was carried out by adopting a method proposed by Friedrich et al. [41, 42], relaxing the normality requirement. The MANOVA test indicated that there was a significant main effect of the agent type for the compositional case ($p < 0.05$), and no significant effects were found for the non-compositional case. Although post-hoc pairwise comparisons were carried out for the compositional case, there were no significant differences. Figure 4 shows box plots of scores, completion time per trial, normalized number of human inputs, and normalized conflicted lateral inputs across the conditions. Figure 5 shows: (A) the normalized number of keystrokes that Participant ID 1 provided during the study and (B) the scatter plots regarding the normalized number of keystrokes and corresponding scores. Additionally, Figure 6 depicts a history of keystrokes of one trial provided by Participant ID 8 who worked on the hard difficulty level with the more capable agent in the loose rein mode.

6. Discussion

6.1 Implications

In this study, we attempted to gain preliminary insights into patterns of communication in the human-agent teams for achieving better team performance using the lunar lander game. Even though the post-hoc pairwise comparisons did not yield significant differences between the conditions, the MANOVA test indicated a significant difference between the two agents. This can be supported by subjective feedback collected during the debriefing; all participants preferred the more capable agent. Yet, the effects of the control input ratios as well as the task difficulty levels are not clear, making it difficult to confirm our hypotheses. Although the more capable agent did not exhibit consistent patterns of keystrokes as shown in Figure 4, Figure 5 implies the expected trend in the keystrokes and scores. In the lower right scatter plot in Figure 5, the upper left area could be a "sweet spot" for achieving successful landing and higher scores whereas plots were more distributed in case of the less capable agent. Participant ID 1 provided subjective feedback on the more capable agent; he let the agent do the landing task in the 75% of agent control input ratio condition, indicating trust and reliance on the agent.

Another participant also shared her strategy during the debriefing. She trusted in the more capable agent once the floating lander entered the landing zone; otherwise, she was trying to navigate the

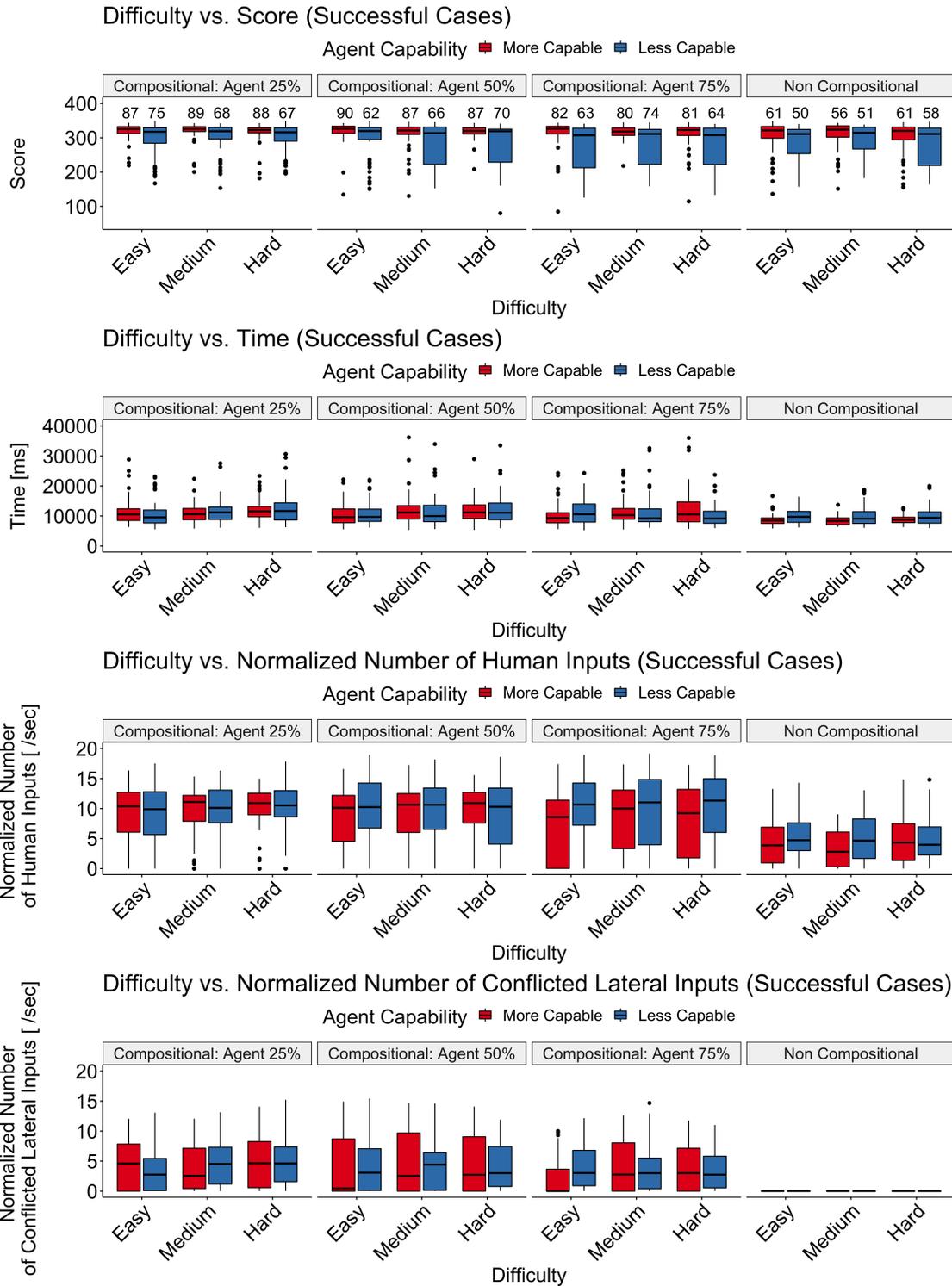


Figure 4. Comparisons of scores, completion time, and normalized number of human inputs across the conditions and control disagreements of lateral inputs. The top box plots also show the number of successful cases out of 90 trials (i.e., 10 trials x 9 participants).

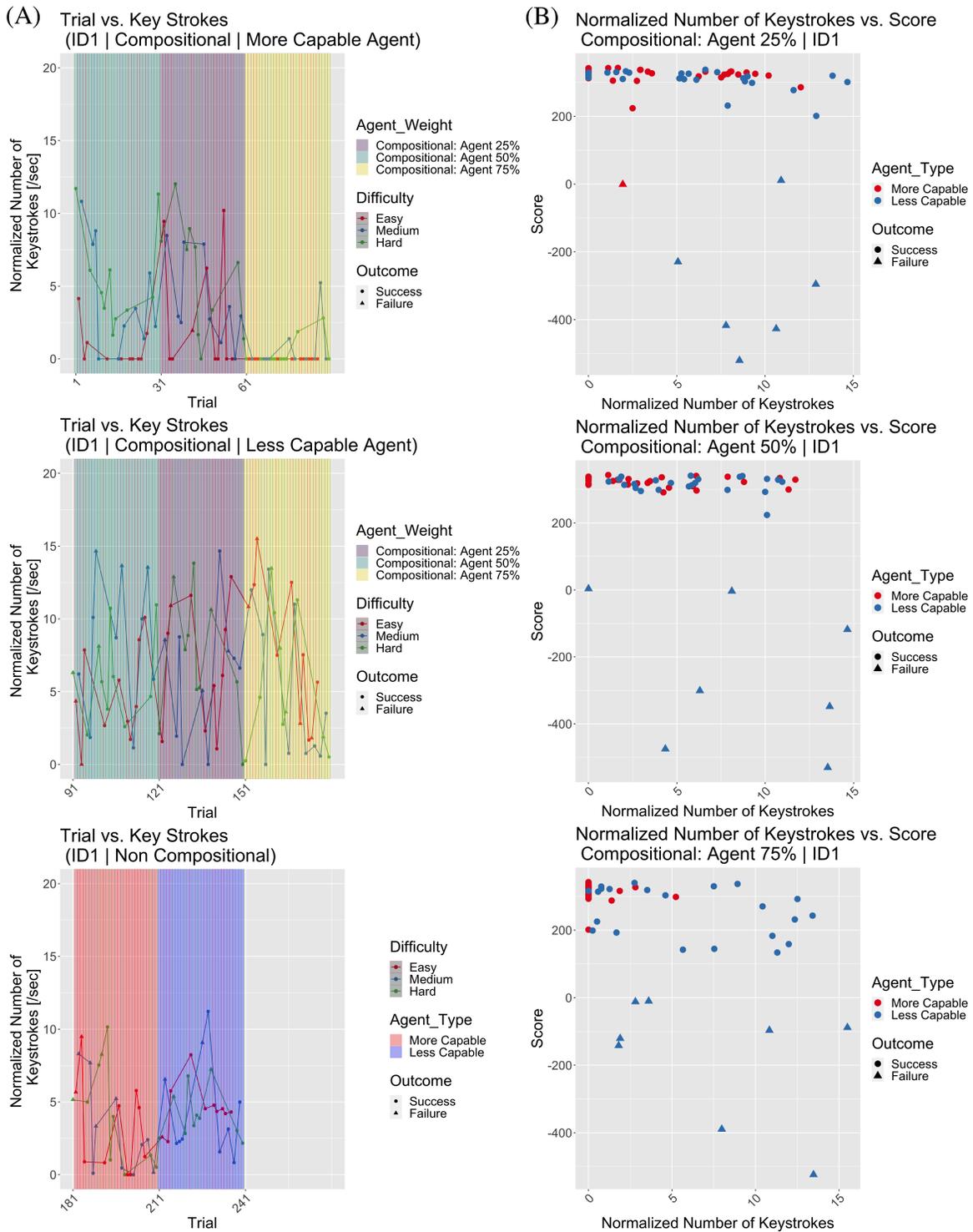


Figure 5. (A) Tracking of the number of keystrokes (normalized) that Participant ID 1 provided in the course of the study (i.e., Trial 1 to 240), and (B) scatter plots show the relationship between keystrokes provided by ID 1 and scores in the compositional case across the three control input ratios.

lander by herself. This highlights the contextual nature of trust. As seen in Figure 6, it seems she attempted to provide more inputs in the first half of the trial whereas a trend of the absence of inputs is observed in the second half of the trial. We believe that such a profile of the keystrokes could help to further investigate whether human possesses appropriate trust in an automated agent for a specific scenario and timing.

As the number of successful landing trials of the non-compositional case indicates (Figure 4), the participants mentioned that the non-compositional case was more challenging. One participant suggested providing another 10 trials in the familiarization session so that participants could feel more comfortable with the non-compositional control scheme. It was observed that participants who played well in the familiarization session felt more comfortable with the non-compositional case.

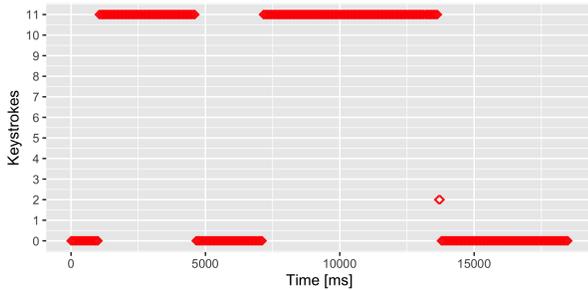


Figure 6. One of the successful hard trials performed by Participant ID 8 with the more capable agent in the loose rein mode (i.e., the agent had 75% of control); the assigned numbers, 0, 2, and 11, mean no inputs, move up, and move up and right, respectively.

6.2 Limitations and Future Work

The number of keystrokes provided by the participants were collected in order to gain insights into patterns of communication to achieve better team performance in the human-agent team. However, with the current game setting, the human-agent team members do not essentially “communicate” with each other. Rather, each entity reacts to the teammate’s input through the observation of the resulting lander behavior. They then provide their own inputs in response. In future studies using the same game environment, we could prompt team communication by introducing multiple landing zones with different widths and rewards. In this scenario, the team is required to reach an agreement on their targeted landing zone. For example, the team had two options: descending vertically (easy / low reward) and moving to the farthest landing zone (hard / high reward). If the human aimed at the vertical landing (i.e., the easy case), they would provide no lateral inputs. Based on the absence of human’s lateral inputs, the agent would understand the human’s intention and then move towards the “designated” landing zone. In such a scenario, the agent’s intention must also be transparent and easy-to-comprehend so that the human operator can recognize the agent’s intention [43]. This type of communication is similar to the communication performed in H-Mode collaboration schemes [6] and is one of the potential exploration directions for future work.

We acknowledge other limitations, including the sample selection and size, and the game environment. Another avenue to explore is to increase the task complexity by introducing a fuel usage constraint, which would require the human-agent team to negotiate and reach an agreement on landing zones and control strategies. One participant suggested alternately presenting the two agents with the same control input ratio so that participants could more easily

compare the two agents. Another participant suggested providing participants with time to write down their experience after each block so that they could revisit their notes during the debriefing. By considering these suggestions, we could administer a questionnaire for investigating participants’ trust, reliance, and workload in the next iteration.

We could also extend the scope of the study to include teleoperation of unmanned vehicles. One of the approaches for expanding the preliminary study in the human-robot teleoperation setting is to employ the steering law [44] and/or the cornering law [45], which serves as an analog to the Fitts’ law [39]. The human-robot team would be asked to conduct a scout mission in a maze and maximize total game scores. Figure 7 illustrates one of the examples, and the team needs to negotiate whether to negotiate one of the corners or go straight. Each pathway has different reward points depending on the task difficulty levels. This remote working environment may require the team to communicate well with each other to achieve a better level of situational awareness and exhibit better game scores. With this teleoperation human-robot teaming setting, we could present different collaboration configurations by introducing, for instance, communication delay, a secondary task for the human operator, and/or multiple robots.

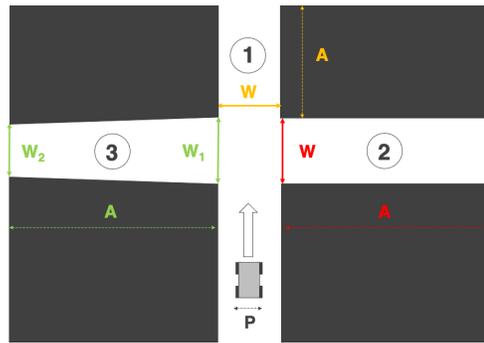


Figure 7. An example intersection presents three potential pathways with different task difficulty levels. The human-robot team is asked to communicate with each other to achieve a better level of situational awareness and exhibit better game scores. The human operator does not have access to this type of third-person view during the trials.

7. Conclusions

We conducted an experiment to gain preliminary insights into human-agent teamwork with a focus on patterns of communication. In the study presented here, the participants jointly played the lunar lander game with the two agents that were configured with different collaboration strategies. We expected that the participants would exhibit a more consistent pattern of inputs when collaborating with the more capable agent. Although the results indicated that the agent’s capability could impact the human-agent team performance, our questions regarding the human-agent teamwork, patterns of communication, and associated influencing factors still remain. One of the critical considerations for the next iteration of experimentation includes implementing a situation where the human-agent team is required to “communicate” with each other for achieving a shared goal, and we could employ the H-Mode-style (intention-based) communication. The subjective feedback from participants implied the notion of appropriate trust and distrust; their trust in the agent seemed to depend on situations, including the task difficulty, lander position, etc. The study presented

here serves as a steppingstone to the next round of experimentation, including human-robot teamwork setting (e.g., teleoperation of unmanned vehicles). Future work should address the human-robot/agent communication and teamwork in a more comprehensive fashion by incorporating the insights from this study.

References

- [1] P. M. Fitts, "Human engineering for an effective air- navigation and traffic-control system.," 1951.
- [2] P. Tschandl, C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, M. Janda, A. Lallas, C. Longo, J. Malveyh, J. Paoli, S. Puig, C. Rosendahl, H. P. Soyer, I. Zalaudek, and H. Kittler, "Human-computer collaboration for skin cancer recognition," *Nature Medicine*, vol. 26, no. 8, pp. 1229–1234, 2020.
- [3] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–489, jan 2016.
- [4] H. Li, N. Sarter, C. Wickens, and A. Sebok, "Supporting human-automation collaboration through dynamic function allocation: The case of space teleoperation," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, pp. 359–363, SAGE Publications Sage CA: Los Angeles, CA, 2013.
- [5] H. Ruff, G. Calhoun, E. Frost, K. Behymer, and J. Bartik, "Comparison of adaptive, adaptable, and hybrid automation for surveillance task completion in a multi-task environment," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 62, pp. 155–159, SAGE Publications Sage CA: Los Angeles, CA, 2018.
- [6] F. O. Flemisch, C. A. Adams, S. R. Conway, K. H. Goodrich, M. T. Palmer, and P. C. Schutte, "The h-metaphor as a guideline for vehicle automation and interaction," 2003.
- [7] E. Salas, D. E. Sims, and C. S. Burke, "Is there a "big five" in teamwork?," *Small group research*, vol. 36, no. 5, pp. 555–599, 2005.
- [8] C. R. Paris, E. Salas, and J. A. Cannon-Bowers, "Teamwork in multi-person systems: a review and analysis," *Ergonomics*, vol. 43, no. 8, pp. 1052–1075, 2000.
- [9] D. R. Ilgen, J. R. Hollenbeck, M. Johnson, and D. Jundt, "Teams in organizations: From input-process-output models to imoi models," *Annu. Rev. Psychol.*, vol. 56, pp. 517–543, 2005.
- [10] A. S. Pentland, "The new science of building great teams," *Harvard business review*, vol. 90, no. 4, pp. 60–69, 2012.
- [11] N. J. McNeese, M. Demir, N. J. Cooke, and C. Myers, "Teaming with a synthetic teammate: Insights into human-autonomy teaming," *Human factors*, vol. 60, no. 2, pp. 262–273, 2018.
- [12] N. J. McNeese, M. Demir, E. K. Chiou, and N. J. Cooke, "Trust and team performance in human-autonomy teaming," *International Journal of Electronic Commerce*, vol. 25, no. 1, pp. 51–72, 2021.
- [13] E. K. Chiou, J. D. Lee, and T. Su, "Negotiated and reciprocal exchange structures in human-agent cooperation," *Computers in Human Behavior*, vol. 90, pp. 288–297, 2019.
- [14] E. K. Chiou, M. Demir, V. Buchanan, C. C. Corral, M. R. Endsley, G. J. Lematta, N. J. Cooke, and N. J. McNeese, "Towards human-robot teaming: Tradeoffs of explanation-based communication strategies in a virtual search and rescue task," *International Journal of Social Robotics*, pp. 1–20, 2021.
- [15] S. G. Lakhmani, J. L. Wright, M. R. Schwartz, and D. Barber, "Exploring the effect of communication patterns and transparency on performance in a human-robot team," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, pp. 160–164, SAGE Publications Sage CA: Los Angeles, CA, 2019.
- [16] T. B. Sheridan and W. L. Verplank, "Human and computer control of undersea teleoperators," tech. rep., Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab, 1978.
- [17] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, 2000.
- [18] M. R. Endsley and D. B. Kaber, "Level of automation effects on performance, situation awareness and workload in a dynamic control task," *Ergonomics*, vol. 42, no. 3, pp. 462–492, 1999.
- [19] D. B. Kaber and M. R. Endsley, "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task," *Theoretical Issues in Ergonomics Science*, vol. 5, no. 2, pp. 113–153, 2004.
- [20] L. J. Prinzel III, F. G. Freeman, M. W. Scerbo, P. J. Mikulka, and A. T. Pope, "Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential p300 component," *Human factors*, vol. 45, no. 4, pp. 601–614, 2003.
- [21] B. Kidwell, G. L. Calhoun, H. A. Ruff, and R. Parasuraman, "Adaptable and adaptive automation for supervisory control of multiple autonomous vehicles," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, pp. 428–432, SAGE Publications Sage CA: Los Angeles, CA, 2012.
- [22] J. Sauer, C.-S. Kao, and D. Wastell, "A comparison of adaptive and adaptable automation under different levels of environmental stress," *Ergonomics*, vol. 55, no. 8, pp. 840–853, 2012.
- [23] F. Flemisch, J. Kelsch, C. Löper, A. Schieben, J. Schindler, and M. Heesen, "Cooperative control and active interfaces for vehicle assistance and automation," 2008.
- [24] K. Goodrich, R. Williams, and P. Schutte, "Piloted evaluation of the h-mode, a variable autonomy control system, in motion-based simulation," in *AIAA Atmospheric Flight Mechanics Conference and Exhibit*, p. 6554, 2008.
- [25] S. Kopp and N. Krämer, "Revisiting human-agent communication: The importance of joint co-construction and understanding mental states," *Frontiers in Psychology*, vol. 12, p. 597, 2021.
- [26] N. Mavridis, "A review of verbal and non-verbal human-robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, pp. 22–35, 2015.
- [27] C. J. Johnson, M. Demir, G. M. Zabala, H. He, D. A. Grimm, C. Radigan, A. T. Wolff, N. J. Cooke, N. J. McNeese, and J. C. Gorman, "Training and verbal communications in human-autonomy teaming under degraded conditions," in *2020 IEEE conference on cognitive and computational aspects of situation management (CogSIMA)*, pp. 53–58, IEEE, 2020.
- [28] M. Harbers, C. Jonker, and B. Van Riemsdijk, "Enhancing team performance through effective communication," in *Proceedings of the 4th Annual Human-Agent-Robot Teamwork Workshop*, pp. 1–2, 2012.

- [29] A. Butchibabu, C. Sparano-Huiban, L. Sonenberg, and J. Shah, "Implicit coordination strategies for effective team communication," *Human factors*, vol. 58, no. 4, pp. 595–610, 2016.
- [30] C. Liang, J. Proft, E. Andersen, and R. A. Knepper, "Implicit communication of actionable information in human-ai teams," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.
- [31] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human factors*, vol. 46, no. 1, pp. 50–80, 2004.
- [32] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Human factors*, vol. 57, no. 3, pp. 407–434, 2015.
- [33] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human factors*, vol. 53, no. 5, pp. 517–527, 2011.
- [34] A.-S. Ulfert and E. Georganta, "A model of team trust in human-agent teams," in *Companion publication of the 2020 international conference on multimodal interaction*, pp. 171–176, 2020.
- [35] M. Brzowski and D. Nathan-Roberts, "Trust measurement in human-automation interaction: A systematic review," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, pp. 1595–1599, SAGE Publications Sage CA: Los Angeles, CA, 2019.
- [36] T. Law and M. Scheutz, "Trust: Recent concepts and evaluations in human-robot interaction," in *Trust in Human-Robot Interaction*, pp. 27–57, Elsevier, 2021.
- [37] L. G. Dizaji and Y. Hu, "Building and measuring trust in human-machine systems," in *2021 IEEE International Conference on Autonomous Systems (ICAS)*, pp. 1–5, IEEE, 2021.
- [38] W. Tan, D. Koleczek, S. Pradhan, N. Perello, V. Chettiar, V. Rohra, A. Rajaram, S. Srinivasan, H. Hossain, and Y. Chandak, "On optimizing interventions in shared autonomy," *arXiv preprint arXiv:2112.09169*, 2021.
- [39] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement.," *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.
- [40] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [41] S. Friedrich, F. Konietschke, and M. Pauly, "Resampling-based analysis of multivariate data and repeated measures designs with the r package manova.rm.," *R J.*, vol. 11, no. 2, p. 380, 2019.
- [42] S. Friedrich, F. Konietschke, and M. Pauly, *MANOVA.RM: Resampling-Based Analysis of Multivariate Data and Repeated Measures Designs*, 2021. R package version 0.5.1.
- [43] K. E. Schaefer, E. R. Straub, J. Y. Chen, J. Putney, and A. W. Evans III, "Communicating intent to develop shared situation awareness and engender trust in human-agent teams," *Cognitive Systems Research*, vol. 46, pp. 26–39, 2017.
- [44] J. Accot and S. Zhai, "Beyond fitts' law: models for trajectory-based hci tasks," in *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pp. 295–302, 1997.
- [45] W. S. Helton, J. Head, and B. A. Blaschke, "Cornering law: The difficulty of negotiating corners with an unmanned ground vehicle," *Human factors*, vol. 56, no. 2, pp. 392–402, 2014.