2023 European Conference on Computing in Construction
40th International CIB W78 Conference
Heraklion, Crete, Greece
July 10-12, 2023

ATTRIBUTING RESPONSIBILITY FOR PERFORMANCE FAILURE ON WORKER-ROBOT TRUST IN CONSTRUCTION COLLABORATIVE TASKS

Woei-Chyi Chang[1], Sophia Marie Ryan[1], Sogand Hasanzadeh[1], and Behzad Esmaeili[1]
[1]Purdue University, West Lafayette, USA

Abstract

Recent advances in construction automation increased the need for cooperation between workers and robots, where workers have to face both success and failure in human-robot collaborative work, ultimately affecting their trust in robots. This study simulated a worker-robot bricklaying collaborative task to examine the impacts of blame targets (responsibility attributions) on trust and trust transfer in multi-robots-human interaction. The findings showed that workers' responsibility attributions to themselves or robots significantly affect their trust in the robot. Further, in a multi-robots-human interaction, observing one robot's failure to complete the task will affect the trust in the other devices, aka., trust transfer. This study calls for the necessity of educating current and future workers regarding safe and productive collaboration with robots.

Introduction

Although robots are expected to have tremendous potential to enhance construction efficiency, they might also impose some latent uncertainties and complexities on construction sites known to be dynamic, hazardous, and uncontrollable. Therefore, it is vital to cultivate trust between workers and robots to boost harmonious interaction and team dynamics.

Human trust is a multifaceted concept that varies over time and can be affected by different factors (Demir et al. 2021). Previous literature has classified the influential factors into three categories: human, robot, and environmental (Hancock et al. 2011). The first category refers to the factors related to human users, such as gender, age, personality, and self-confidence (e.g., Hu et al. 2019; Sanchez et al. 2014). For example, in the study that investigated the effect of gender on trust, Ghazali and her colleagues reported that males manifested higher trust in a robotic advisor than females (Ghazali et al. 2018). Partner factors represent the attribute-based (e.g., anthropomorphism) and performance-based (e.g., reliability and transparency) characteristics of the agent with whom the human interacts. For instance, Liu and his co-workers mentioned providing the current actions of an autonomous vehicle (i.e., transparency) encouraged pedestrians to trust the vehicle (Liu et al. 2021). Finally, environmental factors are associated with the contexts and environments of the interactions between humans and partners, such as task nature and time pressure (e.g., Robinette et al. 2017; Sanders et al. 2019). This taxonomy perfectly reflects the three integral entities (i.e., human, robot, and environment) that can impact workers' trust. This complex development of trust creates uncertainty for future construction sites when incorporating novel robots.

Specifically, failure in the human-robot interaction (e.g., robot's malfunction), as a performance-based factor, has been identified to influence human trust level (e.g., Abd et al. 2017; Kraus et al. 2020; Salem et al. 2015). For example, Kraus and his colleagues found drivers decreased their trust in an autonomous vehicle when it took over tasks without permission (i.e., malfunction) (Kraus et al. 2020). Although the adverse effect of the failure on trust could be readily anticipated, a follow-up question regarding who should be responsible for the failure also arises. In other words, humans might take responsibility for the failure or consider the failure a robot's fault. These two perspectives could trigger different fluctuations in human trust levels. This issue is essential for the context that workers need to be in the loop to cooperate with robots instead of merely supervising robots. However, the effect of taking responsibility for robot failure and its relations with trust are still understudied.

Further, human's perception of robots could be not only based on taking responsibility but also transferred from other scenarios (i.e., trust transfer). Multi-task trust transfer, which denotes the trust is increased/decreased by the robot's success/failure in another similar task, has been proposed by prior studies (e.g., Shu et al. 2018; Soh et al. 2018). For example, in the study that examined the trust transfer in a multi-functional robot, Stewart found humans would initially build trust in the robot being able to pick up a can based on viewing it to grab a plastic bottle (Stewart 2003). Moreover, the multi-task trust transfer could be expanded to multi-agent trust transfer (i.e., trust in an agent can be transferred from the trust in another agent). This type of transfer should play an important role in future construction sites because workers will need to interact with various robots simultaneously or be exposed to different robots across construction projects. Nonetheless, previous literature has not yet paid much attention to the exploration of multi-agent trust transfer.

This study aims to examine how trust is affected by workers' responsibility attributions for the failure in human-robot interaction and how workers transfer trust across multiple agents on future construction sites. Specifically, a VR environment was developed to simulate the future bricklaying task in which workers are required to perform bricklaying, collaborating with a cobot and interacting with various AI agents (e.g., drones delivering materials or messages). These issues are critical to the construction in which workers are unfamiliar with the newly-introduced robots and must interact simultaneously with multiple agents. The expected contributions of this study lie in validating the effect of human responsibility attributions on trust and supporting the multi-agent trust transfer in worker-robot interaction.

Background

Attributing responsibility for failure in human-robot interaction

Failures in human-robot interaction refer to unexpected incidents in the interaction (Honig and Oron-Gilad 2018), where the behaviors performed by the robots are inconsistent with the ideal, standard, and correct functionality (Brooks 2007). This failure can decrease human trust levels in robots (e.g., Abd et al. 2017; Kraus et al. 2020; Salem et al. 2015). For example, in the study investigating the effect of a robot's competence on trust, Abd and his colleagues reported that participants lowered their trust in a bottle-delivery robot when the robot dropped the bottle (Abd et al. 2017). Likewise, Salem and his colleagues suggested that participants who interacted with a faultless robot (i.e., arriving at the destination efficiently) manifested a higher trust level compared to the ones who experienced a faulty robot (i.e., taking a detour to the destination) (Salem et al. 2015).

Previous studies have also proposed taxonomies for classifying different types of failures. For example, Carlson and Murphy categorized the failures into physical failures (i.e., the errors caused by the system's effector, sensors, etc.) and human failures (i.e., the errors caused by human mistakes or slips) (Carlson and Murphy 2005). Honig and Oron-Gilad's taxonomy included technical failures (i.e., errors related to the robot's software and hardware) and interaction failures (i.e., errors pertaining to social norm violation, human errors and environment or other agents) (Honig and Oron-Gilad 2018). These classifications implied that humans are vital parts of the interaction and could not be excluded in case of failure.

Although previous literature indicated the adverse impact of failure in the human-robot interaction on human trust in them, the linkage between taking responsibility for failures and workers' trust in AI-agents in future construction jobsite is unclear.

When failures occur in the interaction, humans might perceive themselves to take responsibility for it or attribute the responsibility to the robot. Previous studies have employed the self-serving bias (SSB) to support the latter statement. SSB refers to the phenomenon that people tend to take credit for the success of the interaction but blame their partners for the failure (Miller and Ross 1975). SSB was initially observed in human-human interaction, and researchers found it could be applied to the interaction between humans and robots (e.g., Moon 2003; You et al. 2011). For example, You and his colleagues conducted an experiment in which participants were asked to follow the motions taught by a robot, and the researchers provided the evaluation of the performance (You et al. 2011). The result indicated that participants tended to question the robot for lower-than-expected performance feedback. In the context of SSB, humans are anticipated to attribute the robot to take responsibility for failure and decrease their trust in the robot.

However, recent studies have also reported that people tend to blame themselves for failures more than robots. In the study that investigated the interaction between two humans and one robot, Lei and Rau found participants attributed more credit and less blame to the robot, triggering the reverse self-serving bias (reverse SSB) (Lei and Rau 2021). An alternative concept supporting reverse SSB is human wishful thinking (Ullrich et al. 2021). Individuals with wishful thinking would believe the robot is perfect when using it and exhibit overreliance on it. The reverse SSB would lead humans to take responsibility for failures and maintain their trust in the robot. Due to the inconsistency of findings in the literature, considering human's taking responsibility as a factor in investigating the effect of failure on trust is essential.

Trust transfer

Apart from taking responsibility for failure, human trust in the current situation can also be affected by trust in another similar situation, called trust transfer. Trust transfer refers to the transition from a known source of trust to an unknown target in which humans need to develop trust (Stewart 2003). Literature has explored the trust transfer in multi-functional robots and emphasized that the transfer is based on the similarity of task category and difficulty (e.g., Shu et al. 2018; Soh et al. 2018; Soh et al. 2020; Xie et al. 2019). In the study conducted by Shu and his colleagues (Shu et al. 2018), after seeing the robot successfully pick and place a cup, participants (1) trusted that it could pick and place a bottle task while distrusting it could perform another simple navigation task (i.e., guiding subjects to a door) and (2) trusted that it could pick and place a bottle task while distrusting it can pick and place an apple. The two findings supported the transfer was triggered by the similarity of task category and difficulty, respectively.

Although multi-agent trust transfer is still understudied, humans will exhibit trust transfer across agents if the similarity between multiple agents is identified. A recent study provided insights into trust transfer across different sources. By conducting a survey, Renner and his colleagues found that respondents' trust in an autonomous car can be transferred from their trust in advanced vehicle technologies and AI because they are integral components of an autonomous car (Renner et al. 2022). In the near future, a worker needs to interact with various AI agents and other workers simultaneously. Therefore, multi-worker-multi-robot-teaming is anticipated to be the dominant working mode on future construction sites. This calls for taking into account the impact of failure on distrust transfer.

Point of departure

It is envisioned that future construction workplace will be comprised of workers, robots, and job sites. Human-robot interaction entails appropriate trust-building between workers and robots, which can be impacted by robot failure or malfunctions. With the growth of human-robot teaming, it is crucial to understand how workers allocate responsibilities (blame for failure) in mixed worker-robot teams, and how these responsibility attributions affect their trust evaluations in robots (H_1 and H_2):

H_1: There is a significant change in worker-robot trust assessment after the robot failure.

H_2: Attributing responsibility (blaming self or robot for failure) significantly impacts worker-robot trust assessment.

Furthermore, with the proliferation of robots in construction jobsites, workers soon need to work alongside and interact concurrently with numerous robots to accomplish complicated construction tasks. Therefore, a robot's performance failure may impact humans' overall reliability and trust in AI agents. Thus, it is critical to explore whether workers' trust transfers across other agents in the construction context in case of failure of an agent (H_3 and H_4):

H_3: When there are multiple agents/robots in the construction environment, the observation of other robots' failure (drone failure) in completing a task will affect the trust in the other agents (i.e., trust transfer).

H_4: Given humans taking responsibility for the robot's failure, the decreasing trust in other agents significantly impacts worker-robot trust assessment (i.e., trust transfer).

Table 1 illustrates the groupings of participants for three hypotheses.

Methodology

Participants

A total of 35 healthy subjects (22 male and 13 female) were recruited to participate in this study. All the subjects are from Civil Engineering and Construction Engineering and Management majors at Purdue University, representing the next generation of the workforce. Participants' age ranged from 19 to 31 years (M= 23.86, SD = 3.32). About 46% of the subjects had over one year of work experience in the construction industry. All participants had normal or corrected-to-normal vision, and the final analyses were based on all 35 participants. All participants received compensation for their participation.

Table 1: Groupings for testing hypotheses

	Group A	Group B
H_1	Subjects' trust level after the Baseline module (i.e., T_b). (A1)	Subjects' trust level after the Error module (i.e., T_e). (B1)
H_2	The changes in trust level (T_e- T_b) of subjects who blamed the cobot for the failure. (A2)	The changes in trust level (T_e- T_b) of subjects who took responsibility for the cobot's failure. (B2)
H_3	The changes in trust level (T_e- T_b) of subjects who lowered their trust in drones and AI-assistant (A3)	The changes in trust level (T_e- T_b) of subjects who retained their trust in drones and/or AI-assistant. (B3)
H_4	The changes in trust level (Te- Tb) of who took responsibility for the cobot's failure and who lowered their trust in drones and AI-assistant (A4)	The changes in trust level (T_e- T_b) of who took responsibility for the cobot's failure and who retained their trust in drones and/or AI-assistant. (B4)

Experimental design

To investigate the effect of blame attribution on trust assessments in human-robot collaboration in the future construction environment, a within-subject experiment was designed to simulate a bricklaying task in an immersive mixed-virtual-reality environment. Participants were asked to complete a bricklaying task while interacting with various AI agents, namely a bricklaying cobot (i.e., called MULE), various drones, and an AI assistant. Figure 1 shows the research framework of the VR experiment.

Specifically, MULE is a semi-autonomous cobot that assists workers with lifting/dropping heavy concrete blocks while workers still have to apply mortar and manually move MULE to the correct positions to pick up and place the blocks. MULE would not drop blocks correctly if participants misplaced MULE or forgot to apply mortar.

This study also included three types of drones: (1) surveillance, (2) delivery, and (3) inspection drones. The surveillance drone was employed to either monitor the status of the construction site by patrolling the job site or facilitate communication between workers by conveying the message of change orders. The delivery drone aimed to deliver materials (i.e., deliver a new mortar bucket for participants and collect the empty one) for workers standing on an elevated platform. The inspection drone was utilized to examine the work progress, monitor the safety behaviors and productivity of workers, and report to the manager. The drone would hover overhead from

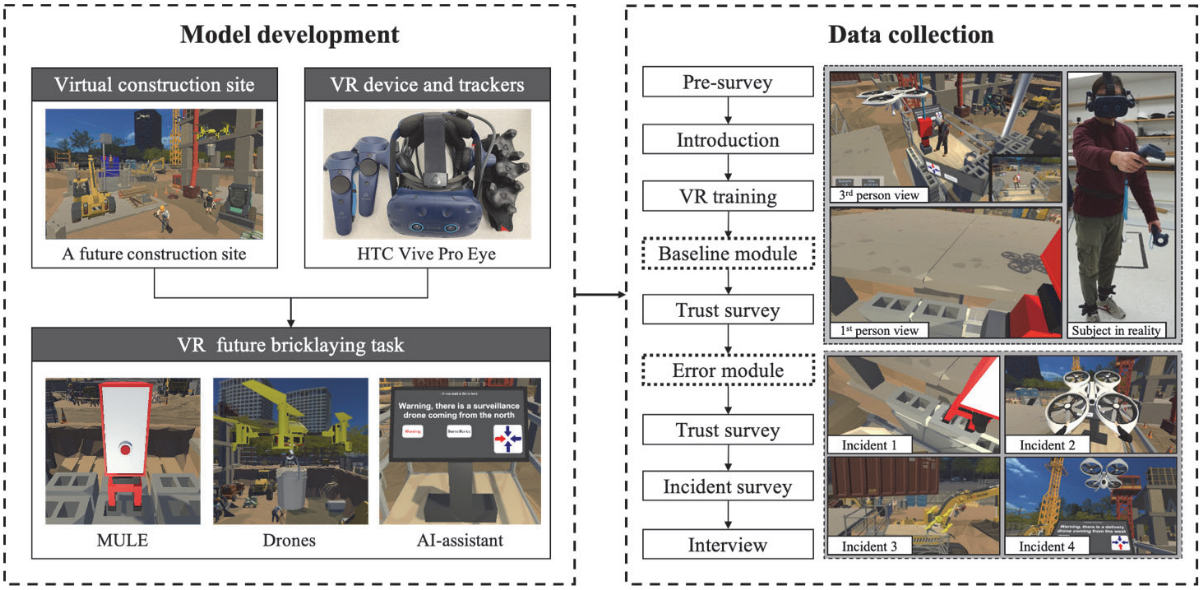

Figure 1: Research framework.

randomly ordered directions while keeping various distances from the subjects. Lastly, AI assistant (i.e., an intelligent screen) is an information delivery platform informing workers about the dynamic objects (e.g., the type and direction of drones) approaching them on the construction site.

To test the hypotheses, this research conducted a within-subject study by asking all subjects to complete two modules (i.e., Baseline and Error modules). The baseline module referred to the scenario that all the agents (i.e., MULE, drones, and AI-assistant) exhibited normal behaviors with ideal performance. However, the error module embodied four incidents happening to the agents: (1) MULE cobot malfunction: it did not drop blocks as expected, (2) Drone malfunction #1: the participant was struck by a drone, (3) Drone malfunction #2: a drone struck another worker at the job site, and (4) AI-assistant malfunction: it provided incorrect information about the type and direction the approaching drone. Noteworthy, this study mainly considered MULE's malfunction as a failure in human-robot interaction to examine human responsibility attributions because both humans (i.e., misplacing MULE) and MULE (i.e., malfunctioning) could be the responsible agent for this incident. For the rest of the incidents (Incidents 2-4), drones and AI-assistant should undoubtedly take responsibility for the failures.

Experimental procedure

Initially, all the subjects were asked to sign a consent form and complete a demographic pre-survey. Then, they were provided with an introduction to the experiment. The introduction presented an overview of the bricklaying task and all types of agents (i.e., MULE, drones, and AI-assistant). In addition, the training offered an opportunity for subjects to familiarize themselves with the VR environment and practice the bricklaying task. Participants were equipped with a VR headset, two controllers, three motion trackers, and neuro-psychophysiological wearable sensors.

After ensuring participants fully understood the experiment process, they were asked to complete baseline and error modules; each taking approximately 7 mins. In both modules, participants needed to perform the bricklaying task in collaboration with MULE and interact with three types of drones while the AI-assistant provided assistive information. The Error module included the incidents mentioned above. A widely-used 5-point Likert-scale trust questionnaire (Muir 1994) was administered to collect their trust levels in three agents separately after each module (i.e., T_b = post-trial trust assessment after the baseline module; T_e= post-trial trust assessment after the error module). After completing the error module, additional questions were asked to examine whether they have noticed the incidents and their responsibility attributions. Finally, the experimenter conducted a brief interview to obtain feedback on the experimental design from the participants. All the procedures were approved by the Purdue Institutional Review Board (IRB).

Apparatus

The selected VR device was the HTC Vive Pro Eye (manufactured by HTC Corporation, Taoyuan, Taiwan), which contains the built-in Tobii eye tracker with a refresh rate of 90 Hz and a field of view of 110°. The calibration system embedded in the headset was developed to calibrate eye-tracking data for each participant. The experiment was run on an Alienware PC with an AMD Ryzen 9 5950X 16-Core processor and an NVIDIA GeForce RTX 3090 graphics card. Due to the page limit, the data from wearable sensors were not considered in the final analyses of this paper.

Results

Normality checks and Levene's test were carried out, and the assumptions were met. A paired t-test was used to examine H_1. H_2, H_3, and H_4 were tested by conducting a one-tailed two-sample t-test. Table 4 provides a summary of the results, and Figure 2 shows a graphical overview of findings based on independent variables (i.e., groupings) and dependent variables (i.e., trust).

Table 4: Results of hypotheses tests (H1- H4)

Hypothesis	Group	N	Mean	STD	*t*-value	*p*-value
H_1	A1	35	4.614	0.439	1.546	0.131
	B1	35	4.471	0.514		
H_2	A2	16	-0.313	0.616	-1.734	0.046*
	B2	19	0.000	0.414		
H_3	A3	19	-0.329	0.507	-2.334	0.013*
	B3	16	0.078	0.490		
H_4	A4	8	-0.250	0.433	-2.531	0.026*
	B4	11	0.182	0.284		

* $p < 0.05$

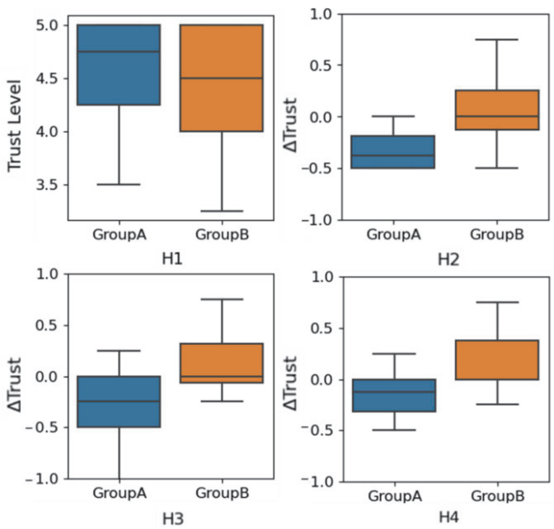

Figure 2: Graphical box plots representing trust changes in H1- H4

H_1: Effect of cobot failures on worker-robot trust

According to the subjective report, all participants noticed the MULE malfunction in the error module. To analyze the effect of the failure on trust, a within-subject paired t-test was utilized to compare the difference in trust levels measured after baseline, Group A_1 ($Mean_{A1}$=4.614; STD_{A1}=0.439) and after error, Group B_1 ($Mean_{B1}$=4.471; STD_{B1}=0.514). The result indicated no significant difference between the two trust levels ($t_{A1\ vs\ B1}$=1.546; p=0.131>0.05). Thus, overall, the cobot failure did not significantly impact workers' trust in robots.

H_2: The effect of workers' responsibility attributions for failures on trust

Among 35 participants, who all noticed the MULE malfunction, 16 subjects reported attributing the responsibility for the failure to the robot (blame target was robot, $Mean_{A2}$=-0.313; STD_{A2}=0.616), while 19 subjects perceived themselves to take responsibility (blame target was human: $Mean_{B2}$=0.000; STD_{B2}=0.414). The results of the t-test revealed a significant effect of responsibility attribution (blame target) on trust level changes (Δtrust= T_e- T_b; $t_{A2\ vs\ B2}$=-1.734; p=0.046<0.05). This indicates that blame attribution by a human influence trust, and subjects who took responsibility for the failure retain their trust in the robot compared to those who regarded the failure as the robot's fault.

H_3/H_4: Trust transfer in human-robot interaction

The trust transfer in this study represented that lowering trust in other agents would be transferred to reduce the trust in MULE. Among all participants, 19 subjects lowered their trust in drones and AI-assistant ($Mean_{A3}$=-0.329; STD_{A3}=0.507) while 16 subjects retained trust in drones and/or AI-assistantafter incidents ($Mean_{B3}$=0.078; STD_{B3}=0.490). The result showed a significant effect of lowering trust in other agents on trust change in MULE (Δtrust= T_e- T_b; $t_{A3\ vs\ B3}$=-2.334; p=0.013<0.05). Although the trust transfer could explain the trust change in MULE, an alternative is human blaming MULE for the failure.

To further investigate whether workers perform multi-agent trust transfer, Group B2 was categorized into Group A4 (i.e., subjects who reduced trust in drones and AI-assistant when taking responsibility for the MULE's failure: $Mean_{A4}$=-0.250; STD_{A4}=0.433) and Group B4 (i.e., subjects who retained trust in drones and/or AI-assistant when taking responsibility for MULE's failure: $Mean_{B4}$=0.182; STD_{B4}=0.284). Moreover, a permutation test was conducted to compensate for the small sample size of the sub-categories (<15). The results showed that distrust in other agents had a significant negative impact on workers' trust in MULE ($t_{A4\ vs.\ B4}$=-2.531 p=0.026<0.05). Hence, the multi-agent trust transfer was corroborated in this study.

Discussion

Due to the proliferation of robots at construction jobsites and as the cooperation between workers and robots increases, individuals have to face both success and failure in human-robot collaborative work, ultimately affecting their trust in robots. To explore the relationship between workers' responsibility attributions (blame for failure) and their subsequent trust in the robot, the present study developed a future bricklaying experiment that participants had to team up with MULE cobot to execute bricklaying while interacting with various drones and an AI-assistant.

The findings indicated that not all participants manifested a significant reduction in trust in MULE cobot after discerning the failure (i.e., MULE's malfunction). This result was inconsistent with the findings suggested by previous literature in other disciplines (e.g., Abd et al. 2017; Kraus et al. 2020; Salem et al. 2015; van den Brule et al. 2014), reporting that all subjects' trust levels were highly influenced by the robot's performance on the task. In the most relevant literature, the failures in human-robot interaction could be readily attributed to the robot's faults by human users. This phenomenon was also related to the experimental tasks that researchers designed. That is, although participants needed to interact with a robot to perform the designated task, they usually played the role of a supervisor or a person being provided service (e.g., McNeese et al. 2021), not collaborating with the robot as was done in this paper. The previous studies also considered technical failures embedded in the robot and disregarded the failures related to human users included in the proposed taxonomies (e.g., Carlson and Murphy 2005; Honig and Oron-Gilad 2018). However, the failure in this research could be attributed to either robot's malfunction or the human's faults. Unlike previous studies, MULE is a semi-autonomous system incapable of solely completing the task, and human intervention was necessary. Therefore, one reason for these inconsistent results is related to moderating effects of responsibility attribution in case of robot failure or malfunction.

The findings of this study demonstrated that most of the participants attributed more responsibility to themselves (blame themselves) than to the MULE cobot for failure. These attributions then considerably affect their trust in robots. The participants who took responsibility for failure would increase their trust in the cobot, while those who attributed the blame to MULE cobot would decrease their trust. This outcome was consistent with the literature that mentioned that failure adversely affected trust (e.g., Abd et al. 2017; Salem et al. 2015).

The findings indicated that workers who have decided to take responsibility, given that the robot caused the failure, would still retain trust in a faulty robot. This over-trust is due to the fact that the implementation of AI agents in the construction industry is still in its infancy compared to other industries and to workers who are unfamiliar with the newly-introduced robots on the jobsites. And often, a less competent agent is more likely to be blamed for the failure in human-robot interaction (Lei and Rau 2021). In addition, transparency (i.e., the comprehension of a robot's intention, ability, and limitations) has been highlighted by literature as a critical factor facilitating human's appropriate trust level (e.g., Clare et al. 2015; Du et al. 2020; Kaniarasu et al. 2013; Kraus et al. 2020). For example, in the study examining the effect of transparency on trust, Kraus and his colleagues suggested informing users of the limitations of an autonomous vehicle in advance can expedite their trust building (Kraus et al. 2020). In other words, when incorporating novel technologies into construction sites, workers might overestimate the robots' capabilities and build inappropriate trust levels. This suggests that more transparency in educating workers regarding potential unexpected behavior, new risks, or robot failures allows them to attribute the responsibility more logically.

In the foreseeable future, workers will still be in the loop to collaboratively interact with robots to perform dynamic and complex construction tasks (e.g., human-centered robot interaction) (Emaminejad and Akhavian 2022). In this context, both humans and robots might be responsible for the failure, and the findings of this study showed that worker perception and responsibility attributions of failure affect their trust in robots.

Further, the results also demonstrated that humans would transfer the decreasing trust in drones and AI-assistant to lower their trust in MULE cobots, even if they have decided to take responsibility for MULE's failure. This multi-agent transfer could be inferred from the multi-task transfer. Previous literature also shed light on the relationship between multi-task trust transfer and the similarity within different tasks (e.g., Shu et al. 2018; Soh et al. 2020; Xie et al. 2019). Humans may adjust their trust in a robot to perform a task based on their trust in another similar task. Extending to multi-agent trust transfer, humans would transfer from the trust in one agent to the trust in another similar agent. In this study, participants have bridged the connection between MULE cobots, drones, and AI-assistant because, from a technical perspective, all of the agents included an element of automation. For example, MULE can automatically lift concrete blocks for workers, while AI assistants can automatically detect the drone's type and coming direction. Hence, the similarity of agents' functionality facilitated the multi-agent trust transfer.

In the future construction industry, workers must collaborate with various robots to execute complicated construction work. This multi-robots-human interaction necessitates workers' trust-building in multiple agents simultaneously. In this context, the performance of one agent might affect workers' trust in another agent. the findings illustrated the interaction between workers and multiple robots and endorsed the multi-agent trust transfer in human-robot interaction.

There are some limitations in this study worth noting. First, while the recruited participants in this experiment represent the next generation of the construction workforce, it is worthwhile to also examine the current experienced workforce who might experience more complacency and reluctance to embrace technologies. Second, the measurement of participants' technology adoption was not included in this research. The extent to which workers accept and utilize the technologies might exert a significant effect on trust. Third, this study only considered the responsibility attribution of MULE cobot failure and the short-term impact of this failure on trust. Future researchers are recommended to explore the long-term impact of failures or the effect of frequent failures on trust. Last, while multiple objective data were collected in the experiment, due to the page limit, it was not used in this paper. The objective trust data could provide more insights into the trust dynamics of workers.

Conclusions

Construction environments are changing rapidly, and with an increasing demand for human-robot collaboration in the construction industry, it is crucial to understand how workers allocate responsibilities (blame for failure) in mixed worker-robot teams. In addition, this might affect workers' trust in their robot partners, which is crucial for successful worker-robot teaming and effective collaboration. The present study investigated the impacts of blame targets (responsibility attributions) on trust in a collaborative bricklaying task simulating the multi-robots-human interaction in future construction jobsites. Results showed that participants who attributed more responsibility to themselves (reverse SSB) than to the robot for the failure retained their trust. However, those who perceived robots to be accountable for their failure (SSB) reduced their trust in the robot significantly. In the future multi-robot multi-human construction work environment, the negative effect on trust gets exacerbated by being transferred to other agents, making them less reliable for workers.

These findings indicate the need for (1) further studying effective communication strategies for robots (non-human agents) in case of failures without compromising the trust relationships with their human partners; and (2) educating current and future workers regarding safe and productive collaboration with robots. This study also provides possible design insights for future construction robots; and calls for continued work in this area to enhance the likelihood of robots being accepted as true teammates, not only as a tool, by current and future construction stakeholders, and attribute with appropriate responsibility for given unexpected situations or failure.

Acknowledgments

This study was supported by the FW-HTF award from National Science Foundation (NSF 2128867 and 2128970). Any findings and recommendations in this material are those of the authors and do not necessarily reflect the views of NSF.

References

Abd, M. A., Gonzalez, I., Nojoumian, M., & Engeberg, E. D. (2017) Trust, satisfaction and frustration measurements during human-robot interaction. In Proceedings of the 30th Florida Conference on Recent Advances in Robotics May 11-12 (Vol. 2107).

Brooks, D. J. (2017) A Human-Centric Approach to Autonomous Robot Failures, Ph.D. dissertation, Department of Computer Science, University. Lowell, MA.

Carlson, J., & Murphy, R. R. (2005) How UGVs physically fail in the field. IEEE Transactions on robotics, 21(3), 423-437.

Clare, A. S., M. L. Cummings, & N. P. Repenning. (2015) Influencing trust for human-automation collaborative scheduling of multiple unmanned vehicles. Hum Factors, 57 (7): 1208–1218. SAGE Publications Inc.

Demir, M., N. J. McNeese, J. C. Gorman, N. J. Cooke, C. W. Myers, & D. A. Grimm. (2021) Exploration of Teammate Trust and Interaction Dynamics in Human-Autonomy Teaming. IEEE Trans Hum Mach Syst, 51 (6): 696–705. Institute of Electrical and Electronics Engineers Inc.

Du, N., K. Y. Huang, & X. J. Yang. (2020) Not All Information Is Equal: Effects of Disclosing Different Types of Likelihood Information on Trust, Compliance and Reliance, and Task Performance in Human-Automation Teaming. Hum Factors, 62 (6): 987–1001. SAGE Publications Inc.

Emaminejad, N., & R. Akhavian. (2022) Trustworthy AI and robotics: Implications for the AEC industry. Autom Constr, 139: 104298.

Ghazali, A. S., J. Ham, E. I. Barakova, & P. Markopoulos. (2018) Effects of robot facial characteristics and gender in persuasive human-robot interaction. Frontiers Robotics AI, 5 (JUN). Frontiers Media S.A.

Hancock, P. A., D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, & R. Parasuraman. (2011) A meta-analysis of factors affecting trust in human-robot interaction. Hum Factors, 53 (5): 517–527.

Honig, S., & Oron-Gilad, T. (2018) Understanding and resolving failures in human-robot interaction: Literature review and model development. Frontiers in psychology, 9, 861.

Hu, W. L., K. Akash, T. Reid, & N. Jain. (2019) Computational Modeling of the Dynamics of Human Trust during Human-Machine Interactions. IEEE Trans Hum Mach Syst, 49 (6): 485–497. Institute of Electrical and Electronics Engineers Inc.

Jeelani, I., & Gheisari, M. (2021) Safety challenges of UAV integration in construction: conceptual analysis and future research roadmap. Safety science, 144, 105473.

Kaniarasu, P., A. Steinfeld, M. Desai, & H. Yanco. (2013) Robot confidence and trust alignment. ACM/IEEE International Conference on Human-Robot Interaction, 155–156.

Kraus, J., D. Scholz, D. Stiegemeier, & M. Baumann. (2020) The More You Know: Trust Dynamics and Calibration in Highly Automated Driving and the Effects of Take-Overs, System Malfunction, and System Transparency. Hum Factors, 62 (5): 718–736. SAGE Publications Inc.

Lei, X., & Rau, P. L. P. (2021) Should I Blame the human or the robot? Attribution within a human–robot group. International Journal of Social Robotics, 13(2), 363-377.

Liu, H., T. Hirayama, & M. Watanabe. (2021) Importance of Instruction for Pedestrian-Automated Driving Vehicle Interaction with an External Human Machine Interface: Effects on Pedestrians' Situation Awareness, Trust, Perceived Risks and Decision Making. 2021 IEEE Intelligent Vehicles Symposium (IV), 748–754. IEEE.

McNeese, N. J., Demir, M., Chiou, E. K., & Cooke, N. J. (2021) Trust and Team Performance in Human–Autonomy Teaming. International Journal of Electronic Commerce, 25(1), 51-72.

Miller DT, Ross M (1975) Self-serving biases in the attribution of causality: fact or fiction? Psychol Bull 82:213

Moon Y (2003) Don't blame the computer: when self-disclosure moderates the self-serving bias. J Consum Psychol 13:125–137

Muir, B. M. (1994) Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics, 37 (11): 1905–1922.

Renner, M., S. Lins, M. Söllner, S. Thiebes, & A. Sunyaev. (2022) Understanding the Necessary Conditions of Multi-Source Trust Transfer in Artificial Intelligence. Proceedings of the 55th Hawaii International Conference on System Sciences.

Robinette, P., A. M. Howard, & A. R. Wagner. (2017) Effect of Robot Performance on Human-Robot Trust in Time-Critical Situations. IEEE Trans Hum Mach Syst, 47 (4): 425–436. Institute of Electrical and Electronics Engineers Inc.

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015) Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 1-8). IEEE.

Sanchez, J., W. A. Rogers, A. D. Fisk, & E. Rovira. (2014) Understanding reliance on automation: Effects of error type, error distribution, age and experience. Theory Issues Ergon Sci, 15 (2): 134–160. Taylor and Francis Ltd.

Sanders, T., A. Kaplan, R. Koch, M. Schwartz, & P. A. Hancock. (2019) The Relationship Between Trust and Use Choice in Human-Robot Interaction. Hum Factors, 61 (4): 614–626. SAGE Publications Inc.

Shu, P., C. Min, I. Bodala, S. Nikolaidis, D. Hsu, & H. Soh. (2018) Human Trust in Robot Capabilities across Tasks. Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, 241–242. New York, NY, USA: ACM.

Soh, H., Y. Xie, M. Chen, & D. Hsu. (2018) Multi-Task Trust Transfer for Human-Robot Interaction. Robotics: Science and Systems.

Soh, H., Y. Xie, M. Chen, & D. Hsu. (2020) Multi-task trust transfer for human–robot interaction. International Journal of Robotics Research, 39 (2–3): 233–249. SAGE Publications Inc. https://doi.org/10.1177/0278364919866905.

Stewart, K. J. (2003) Trust Transfer on the World Wide Web. Organization Science, 14 (1): 5–17.

Ullrich, D., A. Butz, & S. Diefenbach. (2021) The Development of Overtrust: An Empirical Simulation and Psychological Analysis in the Context of Human–Robot Interaction. Front Robot AI, 8.

Van den Brule, R., R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, & P. Haselager. (2014) Do Robot Performance and Behavioral Style affect Human Trust?: A Multi-Method Approach. Int J Soc Robot, 6 (4): 519–531. Kluwer Academic Publishers.

Xie, Y., I. P. Bodala, D. C. Ong, D. Hsu, & H. Soh. (2019) Robot Capability and Intention in Trust-Based Decisions Across Tasks. 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 39–47. IEEE.

You, S., Nie, J., Suh, K., & Sundar, S. S. (2011) When the robot criticizes you... Self-serving bias in human-robot interaction. In Proceedings of the 6th international conference on human-robot interaction (pp. 295-296)