Exploring Human Compliance Toward a Package Delivery Robot

Andrew Washburn, Kripash Shrestha, Habib Ahmed, Dave Feil-Seifer, Hung Manh La
Computer Science and Engineering
University of Nevada Reno, USA

***Abstract*—Human-robot interaction (HRI) studies have found people overtrust robots in domestic settings, even when the robot exhibits faulty behavior. Cognitive dissonance and selective attention explain these results. To test these theories, a novel HRI study was performed in a university library where participants were recruited to follow a package delivery robot. Participants then faced a dilemma to deliver a package in a private common room that might be off-limits. Then, they faced another dilemma when the robot stopped in front of an Emergency Exit door, and they had to trust the robot whether to open it or not. Results showed individuals did not overtrust the robot and open the Emergency Exit door. Interestingly, most individuals demurred from entering the private common room when packages were not labeled, whereas groups of friends were more likely to enter the room. Then, selective attention was demonstrated by stopping participants in front of a similar Emergency Exit door and assessing whether they noticed it. In one condition, only half of participants noticed it, and when the robot became more engaging no one noticed it. Additionally, a malfunctioning robot is exhibited, showing what kind of negative outcome was required to reduce trust.**

I. Introduction

Trust is the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability [1]. A significant factor in trust of an autonomous agent is its performance [2]. For example, three S.W.O.R.D. remotely-operated, combat robots were deployed to Iraq in 2007, but soldiers refused to take them on patrols because they had a faulty reputation [3], [4]. Previously during acceptance testing, the robots unexpectedly moved because of a loose wire, broken solder joint, and an overheated motor. The design was hardened prior to safety certification and deployment. Nonetheless, soldiers would not trust taking the robots alongside dangerous patrols. Thus, trust can greatly affect adoption of new autonomous platforms, and distrust might retard adoption for years.

Overtrust occurs when a system is improperly relied upon. For example, fatal outcomes happen when drivers overtrust autonomous vehicles. In 2018, an Uber autonomous vehicle collided with and killed a pedestrian while the driver was watching TV on their phone. The Arizona governor immediately suspended testing of self-driving cars in the state and public sentiment decreased [5]. While autonomous vehicles promise to reduce the thousands of car related fatalities each year, regulators and the public need to build trust in the systems before widespread adoption. Crashes and unexpected movements increase scrutiny on companies and their practices, such as when a TuSimple autonomous truck veered into a concrete highway barrier because of a software error [6]. Safety will become increasingly important for companies to increase trust and adoption in autonomous vehicles.

A less dangerous situation involves domestic service robots, such as the Amazon Astro. The robot can autonomously map, navigate, and monitor homes. Overtrust could occur when the robot is left alone and fails to identify intruders. However, fatal outcomes are unlikely. A small robot does not possess enough lethal energy or dangerous weaponry to encourage disuse. Therefore overtrust might be more common in a domestic environment. Furthermore, performance issues with domestic robots might have less effect on distrust compared to robots operating in more dangerous situations.

Human-Robot Interaction (HRI) studies in domestic situations demonstrate humans overtrust robots [7]–[9]. Each experiment introduced the participant to the autonomous robot, then the robot built (or failed to build) trust with the person, and finally the robot made a request. Compliance with the request was considered a trustworthy act. Across subjects and experiments, participants followed the robot's request. Faulty behavior did not significantly alter compliance. These studies found people would comply with unusual requests from robots, but they did not identify relevant psychology literature addressing why they complied with the unusual requests, and why they ignored faulty behavior.

This paper identifies the psychology theories of cognitive dissonance [10] and selective attention [11] as confounding factors of overtrust in HRI studies of domestic robots. Our hypothesis is: participants interacting with a package domestic service robot would not overtrust it without understanding clearly what the robot's goal was, and that a more engaging robot would cause more selective attention. An experiment was conducted in a university library where participants followed a package delivery robot and encountered increasing dilemmas. The following sections describe related works in HRI and psychology, the experiment design, and finally results and conclusion.

II. Related Works

A. Human-Robot Interaction

Salem, *et. al.* [7], experimented with a robot in a home setting. The robot introduced itself to participants, and then made unusual requests such as "pour orange juice on a plant." Out of 40 recruited participants, 27 poured orange

978-1-6654-5238-0/22/$31.00 ©2022 IEEE

juice on the plant. No one cited the robot's authority as a reason for compliance. In an attempt to manipulate trustworthiness, the robot would malfunction with faulty navigation and verbal responses. Participants noticed the malfunctions, yet the compliance rates were similar to when the robot performed correctly. This result was surprising because robot performance is the most significant factor in human-robot trust [2].

Robinette, *et. al.* [8], employed a similar malfunctioning robot technique in an emergency evacuation scenario. After following a robot through a hallway to a meeting room, smoke filled the hallway and a fire alarm sounded, prompting an evacuation. Participants saw the robot pointing towards an unknown exit, and every single one followed its direction. Even when the robot made a faulty navigation error, participants still followed its direction. Only a third of participants noticed the additional doorway where they entered from. Later trials showed participants would not trust the robot only if it never worked at all, or if it pointed towards a dark room, obstructed by a couch, with no exit. Thus, even in a potentially dangerous situation, people would follow a faulty emergency guide robot.

In contrast to removing trust, Booth, *et. al.* [9], showed that robot could increase trust by disguising itself. Students entering a dormitory were greeted by a robot that wanted to piggyback inside. Most individuals did not let the robot inside. The robot was much more successful piggybacking along a group of friends. But when the robot toted a box of cookies, individuals were just as likely as groups to allow the robot in the dormitory. They overtrusted the robot, because in post-hoc interviews, many participants who let the robot inside recognized it as a security risk.

In summary, existing HRI studies show people are likely to comply with unusual requests from robots despite poor performance. The literature does not address why people comply with unusual requests, or why faulty robots have little effect on compliance.

B. Psychology

Cognitive dissonance explains people's attitudes while performing tasks [10]. If the task is long and laborious, but well compensated monetarily, a change in attitude follows from the compensation. However, if the compensation is very little, the attitude does not follow from either the task or compensation, so a dissonance occurs. In order to reduce the resulting dissonance, they change their attitude to become more favorable to the task.

An application of cognitive dissonance is advertising via the foot-in-the-door effect. In one experiment [12], housewives were asked to allow a survey team of 5 or 6 men into their homes for two hours and classify household products. In one condition, the housewives were contacted three days prior via telephone survey to answer questions about they household products they used. Agreeing with the smaller request (telephone survey) doubled the compliance rate to the larger request (men entering their home). By saying "yes" a change in attitude occurred towards the activity, which increased the likelihood of agreeing to larger requests.

Selective attention occurs while observers inspect a certain object in their environment and then fail to perceive obvious features in their surroundings. In one experiment [11], participants watched a video of people passing a basketball and were instructed to count the number of passes. Midway through the video a gorilla walked onscreen, waved his arms in the air, and then left. Only half of participants noticed the man in the gorilla suit because they were focused on counting basketball passes.

In summary, cognitive dissonance encourages compliance by changing attitudes in response to agreeing to perform a task. Volunteering to be in a HRI study could be considered agreeing to a small request, therefore agreeing to any further requests would be much more likely. Selective attention narrows the amount of perception people have while engaged in other tasks, such as noticing alternative exits while evacuating a burning building.

This paper addresses the confounding factor of cognitive dissonance in HRI studies by presenting increasingly large requests upon the participant in order to see what level of risk they were willing to accept. In one condition, the robot exhibited faulty behavior and successfully modified a person's decision-making. Finally, selective attention is demonstrated by increasing the engagement with a robot.

III. Experiment Methodology and Design

Methodology of an HRI trust experiment generally follows as:

1) Person(s) introduced to a robot where it demonstrates competence, or lack thereof.
2) Robot makes request.
3) Person(s) comply with request (implying trust), or declines.

In this study, the robot was remotely controlled Wizard-of-Oz style.

A. Package Delivery Routes

This study took place in a public university library setting, shown in Fig. 1. An experimenter recruited participants from the student body studying in the area, and requested them to test a package delivery robot's navigation software by following it. After gaining consent, the experimenter led the participant to the start location seen in Fig. 2. There he explained the task of following the robot to three goal and deliver a package respectively.

The first goal (G1) was a Production help desk for printing posters. This location was in public view, and designed to reinforce the package delivery hand-off. Upon arriving at the goal location, the robot played a stop arrival sound. Then the participant placed one package on the counter. After waving in front of the robot's lidar sensor, it played an acknowledgement sound and moved onward.

Next, the robot stopped outside a common room (G2). No signs denoted what kind of room it was, and participants

Fig. 1. This study took place in a university library. Figure shows the package delivery robot near the experiment's starting point. Participants followed the robot to three goals, where they respectively delivered one of three packages held on top of the robot.

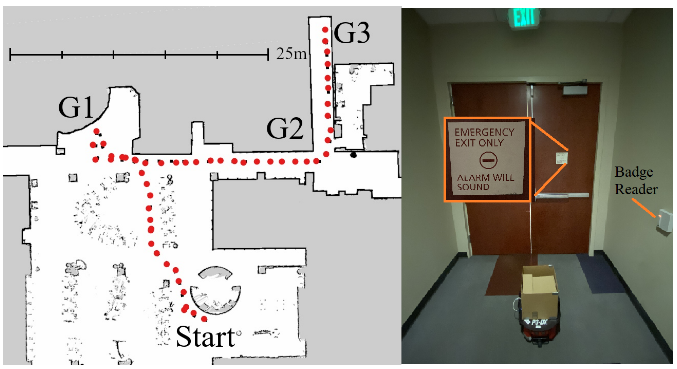

Fig. 2. Package delivery Route 1 with three goals. G1: a production desk in a public area. G2: robot stops outside a common room which is not a public space. G3: an emergency exit door at the end of a hallway. Robot shown in front of the door adorned with the sign "Emergency Exit Only. Alarm Will Sound." The door is always unlocked. Opening the door would sound an alarm, and library personnel could swipe the badge reader to disable the alarm.

described it as "private". An office with a nameplate is within three feet of the door's entrance, as well as a table which could hold one of the packages. Participants faced a dilemma of potentially violating privacy placing the package on the table in the common room, or leaving it outside on the floor.

Finally (G3), the robot led the participant down a long hallway blocked by double doors marked "Emergency Exit Only. Alarm will Sound." Fig. 2 shows the robot at (G3). If the door opened, a loud alarm would sound, but no building systems or security were alerted. Library personnel could freely open the door by disabling the alarm with a badge swipe. Although, several times we heard the alarm go off because a worker opened the door without a badge swipe. The participant's dilemma was to determine whether the robot wanted to open the door, or place the package on the ground. Afterwards, the robot led the participant back to the start, and

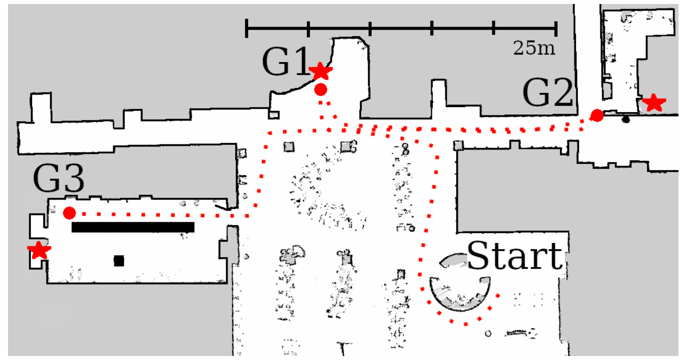

Fig. 3. Modified delivery route to test selective attention. The first (G1) and second (G2) stops are the same as in Fig. 2. The third stop (G3) is in front of another Emergency Exit door in a computer lab. Participants are questioned afterwards if they noticed the Emergency Exit door at G3. Red circles indicate robot's stop location. Red stars are intended package delivery spot.

they answered a post-trial questionnaire.

In order to test selective attention, a modified route maintained the first two stops, while changing the third delivery to be in a public computer lab seen in Fig. 3. Inside the computer lab was another Emergency Exit door identical to the one shown in Fig. 2. A red star in Fig 3 shows the intended delivery location several feet away from where the robot stopped. Thus the participant must observe their environment to determine the correct delivery location. Afterwards, they are questioned whether they noticed the emergency exit door.

B. Between Subjects Conditions and Hypotheses

a) Robot communication: In Route 1 (Fig. 2) the robot communicated using only two non-verbal sounds: one for stop arrival (`ROS sound_play builtin 1`) and the other acknowledging package delivered by the participant (`ROS sound_play builtin 1`). These sounds were informative enough to communicate between the robot and the person, but belies no further information where to place the package.

During Route 2 (Fig. 3), speech was added on after the sounds: the stop arrived phrase indicated the respective location, "Stop arrived. Please deliver the package to [Production]" and the acknowledgment phrase was, "Package delivered. Let's move to the next stop."

b) Non-verbal Prods to Request Opening the Emergency Exit Door: Opening the emergency exit door in Fig. 2 represented a large request. The robot escalated non-verbal prods between subjects in order to convey it wanted to open the door:

1) Return as soon as participant signaled package delivered.
2) Wait one minute at the emergency exit door.
3) Nudge the emergency exit door.

c) Package Labeling: In Route 1 none of the packages were labeled. Participants randomly choose which package to place at each stop. For Route 2 all the packages were labeled respectively for "Production" and the two offices.

d) Individuals and Groups: Recruited participants were chosen as individuals or as a small group composed of a pair of friends.

TABLE I
TRIAL COUNTS PER STUDY CONDITION.
*WEEK BEFORE FINAL EXAMS.
**INVALID MEANS NO PACKAGES DELIVERED.

Date	Condition	Population	Trials	**Invalid
April 20-22	Route 1 Non-Verbal	Individual	12	2
		Group	2	-
*April 28 - May 3	Route 2 Non-Verbal	Individual	4	7
		Group	3	-
	Route 2 Speech	Individual	3	1
		Group	2	-
		Total	26	10

The study's hypotheses were:

1) Non-verbal Prods increase compliance to open the Emergency Exit door.
2) Labeling packages would prod more participants to enter the common room at G2.
3) A speaking robot would encourage more compliance than non-verbal communication.

IV. RESULTS AND DISCUSSION

A. Participants

No one was harmed from participating in the study. One person remarked it was the most interesting activity they had done all day. Table I describes the number of trials conducted over two weeks in late April 2022. All participants were undergraduates studying in the library. They were not familiar with the common room or emergency exit doors. They all understood how to communicate with the robot, and everyone followed the robot to all three delivery stops and filled out the post-trial questionnaire. Some participants observed previous trials of people delivering packages around the library, but they were unaware of the study's intention to test overtrust and selective attention.

Some individuals did not deliver any packages, resulting in an invalid trial. Regardless, they reported the instructions were clear, followed the robot to all three stops, and answered the questionnaire This occurred more often during the week before final exams than the prior week. All groups of participants delivered packages. It's unclear why there were more invalid trials the week before final exams.

B. Route 1 Emergency Exit Door

No one opened the emergency exit door at the end of Route 1. An individual remarked, "the robot just stopped and beeped, so I put the package on the ground. An Amazon delivery driver wouldn't open someone's door. It turned around when I put my hand in front." When the robot waited one minute at the emergency exit door, participants were unsure what the robot wanted. "While waiting we thought the robot would signal us to open the door, but we were not sure. We were trusting the technology." All participants waited the entire minute until the robot turned back to the start. In two trials, someone opened the emergency exit door from the other side. The experimenter immediately drove the robot back to the start location. Participants associated the robot turning around with

TABLE II
DELIVERY LOCATIONS AT GOAL 2.

Condition	Population	Floor	Table	Office
No Label	Individual	9	2	-
	Group	-	2	-
Label	Individual	1	2	1
	Group	-	2	1
Label + Speech	Individual	1	-	2
	Group	-	-	2

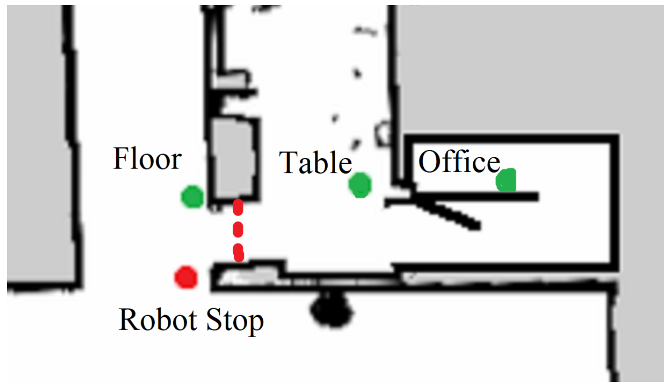

Fig. 4. Goal 2 stopped in front of a common room. A package with no label provided no initiative for most individuals to enter the room, whereas groups were more likely to enter the room to deliver the package. Labeling the package prodded most individuals to enter the room.

the door opening, and wondered if that was the robot's true intention. Thus, when the robot arrived at the emergency exit door and made a simple stop arrival sound, participants did not presume to open the door. Waiting one minute did not encourage participants to open the door either. In a separate experiment, the robot attempted to non-verbally communicate it wanted to enter the door to the computer lab, but participants did not understand what it wanted unless the robot used speech, "please open the door for me." In conclusion, non-verbal sounds were insufficient the robot to make a new, complicated request, and participants did not overtrust the robot and open an emergency exit door to deliver a package.

However, when the robot nudged the emergency exit door, all three individuals thought it was malfunctioning. "It was running into a door like it didn't recognize it." One participant ran back to the start exclaiming the robot was running into a door. In contrast to previous HRI studies [7], [8], this robot's faulty behavior deterred participants from trusting it because the negative outcome was highly noticeable.

C. Delivering a Package in the Common Room

In both Routes 1 and 2 the robot stopped outside a common room (G2). Participants faced a dilemma to enter the common room to deliver the package, or leave it in the hallway on the floor. Fig 4 shows the relative location where the robot stopped to the intended delivery locations. The door to the common room was always open, but sometimes the office door was closed when unoccupied.

TABLE III
PARTICIPANT NOTICED THE EMERGENCY EXIT DOOR, CONDITIONED ON ROBOT COMMUNICATION LEVEL.

Condition	Population	Yes	No
Sound	Individual	5	6
	Group	1	2
Speech	Individual	-	4
	Group	-	2

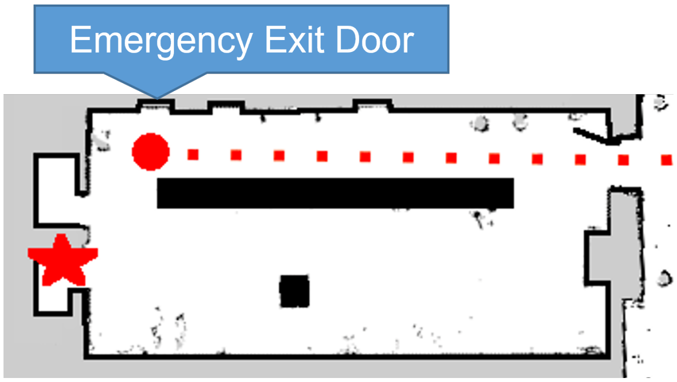

Fig. 5. Route 2 - Goal 3 stopped in front of an emergency exit door that was similar in appearance as the one seen in Fig. 2. The robot followed the dotted trajectory and stopped at the red dot, and the red star was the delivery destination. Participants were facing away from the emergency exit door, and they had to observe their surroundings in order to find the office to leave the package at.

When packages had no labels, most individuals did not enter the common room. They described the room as “eerie. I didn’t know this person. The room could have been off limits.” One individual entered the common room then quickly backtracked because they noticed it was occupied. In contrast, groups were more likely to deliver an unlabeled package inside the common room (Fisher’s Exact test p=0.076). This result aligns with groups being more likely to allow piggybacking robots into dormitories [9].

Labeling the packages with the office holder’s name was sufficient to prod most individuals into the common room. Instead of fearing a privacy violation, participants were confident enough to walk all the way into the office and hand the package to the worker at his desk. In this case the robot used only simple sounds for communication, and increasing communication to speech had no significant effect.

D. The Effect of Robot Communication

Table IV shows participant responses to a post-trial questionnaire administered after completing Route 2. In these trials, the robot’s communication was modified from only using sounds to employing speech instructions. A speaking robot was found to communicate more clearly, want the participant’s best interest, and they thought it was less likely to be remote controlled and less faulty. Higher scores for a speaking robot indicate it was more engaging.

Route 2 ended inside a public computer lab, shown in Fig. 3 and Fig. 5. The robot stopped in front of an Emergency Exit door similar to Fig. 2, and the intended delivery destination was an office several feet away. Participants had to observe their surroundings to see find where to deliver the package. When the robot communicated with only sounds, Table III shows about half of participants noticed the Emergency Exit door immediately to their right. When the robot used speech no one reported noticing the emergency exit door. One group of participants didn’t know where the office was, so they placed the package right on the door itself, but they didn’t record noticing it. The questionnaire scores and the selective attention results indicate a more engaging robot increases selective attention. Thus, an emergency evacuation guide robot [8] providing direction would likely blind a person’s perception of their surroundings.

TABLE IV
POST-TRIAL QUESTIONNAIRE RESULTS CONDITIONED ON ROBOT COMMUNICATION LEVEL. SCORES ON A 5-POINT SCALE. ORDERED BY F SCORE ($p^{**} < 0.05, p^{*} < 0.1$).

	Sound (n=17)		Speech (n=8)	
	Mean	SE	Mean	SE
The robot communicates clearly.	3.53	1.18	4.75^{**}	0.46
The robot wants my best interest.	3.41	1.00	4.25^{**}	0.71
The robot is remote controlled.	3.41	1.06	2.25^{*}	1.83
The robot is faulty.	2.12	1.11	1.38^{*}	0.52
I know what the robot wants.	3.59	1.00	4.13	0.64
The robot understands me.	3.24	1.09	3.88	1.13
The robot is autonomous.	4.00	1.06	3.38	1.19
I trust the robot.	4.00	1.06	4.38	0.92
I understand the robot.	3.76	1.09	4.13	0.99
The robot navigates successfully.	4.47	0.62	4.63	1.06
The robot can hear me.	2.65	1.11	2.88	1.25
The robot is erratic.	2.35	1.11	2.13	1.46
The robot can see me.	4.29	1.10	4.25	0.71
The robot is trustworthy.	4.00	0.87	4.00	0.93

E. Limitations

Public spaces are noisy. For example, in two trials a touring group of students wandered into the robot’s path, forcing it to wedge through the crowd. Afterwards, the participant remarked how advanced the robot’s software was and how it could navigate among so many different people (they were not told it was remote controlled). In some trials, the office door in the common room was closed which might have affected the participant’s willingness to enter. Additionally, small sample sizes are much more likely to yield different results across different studies.

Non-verbal communication was insufficient to request opening a door. Participants needed explicit instructions in order to understand a new request.

Some participants remarked following a robot and delivering packages was not very realistic. They were recruited to “test a robot’s navigation around people” therefore delivering packages would be additional scope to comply with. Most participants in recruited in the first week of the study complied with the delivery instructions. However, in the following week (week before final exams) more individuals just followed the robot without delivering any packages.

Lastly, the groups of participants were friends, and unrelated people may act differently.

V. Conclusion

This paper addressed overtrust of robots in HRI compliance studies. Participants are likely to comply with unusual requests because they agreed to be in the study (cognitive dissonance) and ignore better outcomes because they are blind towards their environment (selective attention). Even faulty robots have been unable to change people's decision-making.

To test these theories, we developed an HRI compliance study with increasingly risky requests. Participants followed a package delivery robot around a university library and had to decide where to place the packages. After stopping in front of a private common room, most individuals refused place an unlabeled package inside the common room, whereas groups of friends were more likely to enter the room and deliver the package. After labeling the packages, most individuals entered the common room. A more risky delivery stopped in front of an Emergency Exit door. No one opened the door even when the robot waited one minute. However, when the robot nudged the door, participants thought it was malfunctioning, thus demonstrating a faulty robot can deter a participant's decision-making. Finally, at a delivery stop where participants observed their surroundings to find the intended delivery destination, only half of participants noticed an Emergency Exit door just outside their field of view. When the robot was more engaging no one reported noticing the door. Future work could investigate the influence of training on cognitive dissonance and selective attention.

References

[1] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, p. 50–80, Mar 2004.

[2] P. Hancock, D. Billings, K. Schaefer, J. Chen, E. de Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors 53(5)*, p. 517–527, 2011.

[3] S. Weinberger, "Armed robots still in iraq, but grounded (updated)," *Wired*, April 2008. [Online]. Available: https://www.wired.com/2008/04/armed-robots-st/

[4] S. Ogreten, S. Lackey, and D. Nicholson, "Recommended roles for uninhabited team members within mixed-initiative combat teams," in *2010 International Symposium on Collaborative Technologies and Systems*, May 2010, p. 531–536.

[5] P. Penmetsa, P. Sheinidashtegol, A. Musaev, E. K. Adanu, and M. Hudnall, "Effects of the autonomous vehicle crashes on public perception of the technology," *IATSS Research*, vol. 45, no. 4, p. 485–492, Dec 2021.

[6] K. O' Keeffe and H. Somerville, "Self-driving truck accident draws attention to safety at tusimple," *Wall Street Journal*, Aug 2022. [Online]. Available: https://tinyurl.com/2s3b2x8n

[7] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Mar 2015, p. 141–148. [Online]. Available: https://dl.acm.org/doi/10.1145/2696454.2696497

[8] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, "Overtrust of robots in emergency evacuation scenarios," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar 2016, p. 101–108.

[9] S. Booth, J. Tompkin, H. Pfister, J. Waldo, K. Gajos, and R. Nagpal, "Piggybacking robots: Human-robot overtrust in university dormitory security," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Mar 2017, p. 426–434. [Online]. Available: https://dl.acm.org/doi/10.1145/2909824.3020211

[10] D. J. Bem, "Self-perception: An alternative interpretation of cognitive dissonance phenomena." *Psychological Review*, vol. 74, no. 3, p. 183–200, 1967.

[11] D. J. Simons and C. F. Chabris, "Gorillas in our midst: Sustained inattentional blindness for dynamic events," *Perception*, vol. 28, no. 9, p. 1059–1074, Sep 1999.

[12] J. L. Freedman and S. C. Fraser, "Compliance without pressure: The foot-in-the-door technique," *Journal of Personality and Social Psychology*, vol. 4, no. 2, p. 195–202, Aug 1966.