

5-2018

Exploring the Efficacy of Social Trust Repair in Human-Automation Interactions

Daniel Braden Quinn

Clemson University, dquinn628@gmail.com

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Recommended Citation

Quinn, Daniel Braden, "Exploring the Efficacy of Social Trust Repair in Human-Automation Interactions" (2018). *All Theses*. 2875.
https://tigerprints.clemson.edu/all_theses/2875

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

EXPLORING THE EFFICACY OF SOCIAL TRUST REPAIR IN HUMAN-
AUTOMATION INTERACTIONS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements of the Degree
Master of Science
Applied Psychology

by
Daniel Braden Quinn
May 2018

Accepted by:
Dr. Richard Pak, Committee Chair
Dr. Patrick Rosopa
Dr. Ewart de Visser

ABSTRACT

Trust is a critical component to both human-automation and human-human interactions. Interface manipulations, such as visual anthropomorphism and machine politeness, have been used to affect trust in automation. However, these design strategies have been primarily used to facilitate initial trust formation but have not been examined means to actively repair trust that has been violated by a system failure. Previous research has shown that trust in another party can be effectively repaired after a violation using various strategies, but there is little evidence substantiating such strategies in human-automation context. The current study examined the effectiveness of trust repair strategies, derived from a human-human or human-organizational context, in human-automation interaction. During a taxi dispatching task, participants interacted with imperfect automation that either denied or apologized for committing competence- or integrity-based failures. Participants performed two experimental blocks (one for each failure type), and, after each block, reported subjective trust in the automation. Consistent with interpersonal literature, our analysis revealed that automation apologies more successfully repaired trust following competence-based failures than integrity-based failures. However, user trust in automation was not significantly different when the automation denied committing competence- or integrity-based failures. These findings provide important insight into the unique ways in which humans interact with machines.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Richard Pak, for his mentorship and for pushing me to excel as a thinker and researcher. I would also like to thank my committee members, Dr. Patrick Rosopa and Dr. Ewart de Visser, for their guidance and wisdom throughout this project. I am eternally grateful for the humor, nonsensical shenanigans, and indispensable friendship of my academic family, Will, Jessica, Natalee, and Jeremy. I am indebted to my family for their unwavering support and patience. Mom, Dad, Chris, and Dave, thank you each for you for your kindness and love and for letting me lean on you throughout this process. Last, but certainly not least, I would like to thank my endlessly supportive girlfriend, Brittany, for her immeasurable love and patience.

TABLE OF CONTENTS

	Page
TITLE PAGE	i
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES AND FIGURES.....	vi
CHAPTER	
1. INTRODUCTION	vi
Trust Factors: During Interaction.....	1
Strategies to Repair Trust.....	4
Current Study	5
2. METHOD	5
Participants.....	5
Materials	6
Design	8
Measures	10
Procedure	13
3. RESULTS	16
History-based Trust.....	16
Perceived Competence and Integrity	17
Causal Attributions	18
Exploratory Variables	10
4. DISCUSSION	22
Limitations and Future Directions	25
Conclusion	27
APPENDICES	28
A: Hoff and Bashir’s (2015) model of trust.....	29
B: Example task interfaces	30
C: Full list of automation error messages	32
D: History-based trust questionnaire	34
E: NASA-TLX measure for workload	35

TABLE OF CONTENTS (Continued)

	Page
F: Perceived competence & integrity questionnaire	36
G: Causal attribution questionnaire	37
H: Complacency potential rating scale (CPRS)	38
I: Summary tables for descriptive and inferential statistics	40
J: Graphs of nonsignificant results	43
REFERENCES	57

LIST OF TABLES AND FIGURES

TABLES	Page
1: Example system error message.....	10
2: Participant characteristics by the between-subjects factor.....	15
3: Descriptive and inferential statistics for the main effects of failure type	40
4: Descriptive and inferential statistics for the main effects of repair strategy.....	41
5: Descriptive and inferential statistics for the interaction effects of repair strategy and failure type	42

FIGURES

1: Task interface during automation failure	7
2: Trust comparisons	16
3: Casual attributions to the situation.....	20
4: Physical workload	22
5: Perceived competence.....	43
6: Perceived integrity	44
7: Causal attributions to self.....	45
8: Causal attributions to the automation (TED)	46
9: Perceived responsibility	47
10: Perceived usefulness	48
11: Self-confidence	49
12: Perceived reliance	50
13: Task time.....	51
14: Perceived mental workload.....	52
15: Perceived temporal workload	53
16: Perceived effort.....	54
17: Perceived performance.....	55
18: Perceived frustration	56

INTRODUCTION

In a recent review of human-automation studies, Hoff and Bashir (2015) identified several factors that affect how trust is formed *prior to* and *during* a user's interaction with automation (see Appendix A). Prior to interaction with automation, a person's trust is influenced by dispositional factors such as enduring individual differences in personality, cognitive styles, and pre-existing knowledge of the automation. For example, extraverted individuals and those with a high propensity to trust technology experienced high trust in an automated decision aid before interacting with the system in an x-ray screening task (Merritt & Ilgen, 2008). While dispositional factors play an important role in a user's trust going into an interaction with automation, they cannot dynamically calibrate trust during that interaction. Therefore, the current research focuses on the factors that can dynamically affect trust and are amenable to human factors interventions during human-automation interaction.

Trust Factors: During Interaction

Reliability is the most critical system performance factor determining user trust. During human-automation interactions, trust is formed when users experience reliable system performance (Merritt & Ilgen, 2008). Users are less likely to trust unreliable systems and may disuse automation that is perceived to be untrustworthy (Lee & See, 2004). The distinct attributes of each system failure can have variable effects on trust (Hoff & Bashir, 2015). Failures that severely impact task performance cause greater harm to trust than less severe failures (Muir & Moray, 1996). Failures that occur early in an interaction may have a more damaging effect than those that occur later (Manzey, Reichenbach, & Onnasch, 2012). Additionally, failures that occur on tasks perceived by users to be easy may result in greater trust harm (Madhavan, Wiegmann, & Lacson,

2006). Some evidence has suggested that the type of failure (e.g., false alarms or misses) may also differentially affect trust. In the military domain, misses have been empirically shown to impact trust more than false alarms (Davenport & Bustamante, 2010). A possible explanation for the differential effects of failure types on trust is the consequences associated with each failure (Hoff & Bashir, 2015). Trust in automation recovers slowly over time after exposure to error-free performance (Lee & Moray, 1994). Therefore, post-failure interventions should be utilized to help maintain high user trust or to prevent extreme trust declines. Altering system performance parameters is ultimately an engineering problem (i.e., creating more reliable automation) and thus beyond the scope of human factors intervention. However, altering the users' perception of the system (i.e., trust) by modifying system design is one strategy that has been well-researched in the human-automation context.

Researchers have recently examined the effect of automation design features on user trust. Humanizing the appearance of automation can significantly increase user perceptions of trust (Nass & Moon, 2000; Pak, Fink, Price, Bass, & Sturre, 2012, & Pak, McLaughlin, & Bass, 2014). For example, autonomous cars that were given names, genders, and voices were found to be more highly trusted than cars that were not given those characteristics (Waytz, Heafner, & Epley, 2014). In addition to engendering higher trust, humanlike aspects added to automation seem to also enhance trust calibration to reliability (de Visser, Krueger, McKnight, Scheid, Smith, Chalk, & Parasuraman, 2012). In their study, de Visser et al. varied the appearance of an automation agent (e.g., machine, agent, and human) and found that user trust in humanlike agents was more

accurately reflective of actual system performance. Humanlike appearance interventions not only enhance trust calibration, they can buffer the harmful effects of system failures on trust (trust resilience; de Visser, Monfort, McKendrick, Smith, McKnight, Krueger & Parasuraman, 2016). De Visser et al. (2016) found that participants, when given the opportunity to accept or ignore the advice of a machine, avatar, or human agent, were more likely to comply with more humanlike agents. These findings suggest that visual anthropomorphism can affect perceptions of system reliability and therefore minimize the harmful effects of system failures on trust.

While system appearance interventions have shown promise altering user perceptions of automation, one avenue of human factors intervention that is less well researched is communication style (e.g., Parasuraman & Miller, 2004). In a study directly examining the effect of politeness on automation trust and reliance, Parasuraman and Miller (2004) found that participants exhibited less trust and reliance on interruptive (e.g., rude) automation than non-interruptive (e.g., polite) automation. In this experiment, the effect of good etiquette was so strong that it overcame the negative effect of system reliability on task performance. Additional research has similarly demonstrated that automation politeness can improve user perceptions of system reliability (Spain & Madhavan, 2009). These findings suggest powerful effects of communication on user perceptions of system performance. Thus, system communication may prove to be an effective method to repair post-failure trust in human-automation interactions.

Strategies to Repair Trust

There already exists a large body of research that has examined the effectiveness of communication as a trust repair strategy (e.g., Goffman, 1971; Shapiro, Sheppard, & Cheraskin, 1992; Lewicki & Bunker, 1996; Bottom, Gibson, Daniels, & Murnighan, 2002; Schweitzer, Hershey, & Bradlow, 2006). A notable study conducted by Kim, Ferrin, Cooper, and Dirks (2004), determined that apologies and denials can effectively repair trust following trust harm. In their experiment, participants were shown videos of potential job candidates, who had been accused by their references of incorrectly filing a tax return. The trust violation was framed to suggest either incompetence or lack of integrity of the candidate. Following the accusation, the candidate responded with either an apology or denial, and participants were asked to rate the trustworthiness of the candidate. Results showed that apologies better repaired trust for competence failures and denials better repaired trust for integrity failures.

A recent review of trust repair literature has suggested a theoretical framework supporting the interaction between trust repair strategy and failure type in human-automation interactions (Marinaccio, Kohn, Parasuraman, & de Visser, 2015); however, no research has empirically evaluated this relationship. The effects of perceived system competence (e.g., reliability) are well understood in the automation domain, but no studies have examined integrity. Evidence has been provided for the effect of system apologies on user perceptions, such as user self-appraisal of performance (Akgun, Cagiltay, & Zeyrek, 2010) and system impressions (Tzeng, 2004; Lee, Kiesler, Forlizzi, Srinivasa, & Rybski, 2010). Although, we are aware of only two studies that have

directly addressed apology and trust repair in an automation context. Robinette, Howard, and Wagner (2015) showed that timing is a significant factor for trust repair in human-robot interaction. They observed that promises and apologies made by a simulated robot navigation aid affected reliance only when implemented immediately before a user's decision to rely on the system, after system failure. Later, de Visser and colleagues (2016) determined that apologies can successfully raise trust across agents of varying degrees of visual anthropomorphism. However, these authors neither examined the effect of denials nor the potentially influential effects of system reliability.

Current Study

The current study explored the efficacy of using communication strategies (apologies or denials) to repair human trust in automation. Prior research found that apologies repair trust following competency-based trust violations and that denials repair trust following integrity-based violations (e.g., Kim, Dirks, Cooper, & Ferrin, 2004). Our work determined whether these findings extend to social interactions with machines. We hypothesized that a) apologies better repair user trust for competence-based automation failures and b) denials better repair user trust for integrity-based automation failures.

METHOD

Participants

Seventy-four undergraduate students ages 18 to 21 ($M = 18.68$, $SD = .94$) were recruited and given course credit for participation in this study. Due to equipment failure, demographics were collected for only 37 participants.

Materials

Equipment. Data were collected in a lab setting on computers with 3.2 GHz processors, 4GB RAM, Windows 7, and a 19-inch LCD monitor set at a resolution of 1280x1024. Participants sat approximately 18 inches away from the monitor and interacted with the system using a mouse.

Task. This study used a taxi dispatching task paradigm that is conceptually based on a task used in previous research (e.g., Rovira, McGarry, & Parasuraman, 2007; Pak, McLaughlin, Leidheiser, & Rovira, 2016). In this simulation, participants play the role of a taxi driver for an app-based ride-share company. Participants' task was to select customers that maximize income and rating. The automation provided calculations of the closest fares, estimated fare, and traffic information. The interface (see Figure 1) consisted of five primary areas: a city road map (upper left), an automated fare computer (upper middle), a fare history panel (upper right), a status message panel (lower right), and a communications panel (lower left). The city map displays green potential customer units (C1 to C6), yellow taxi units (T1 to T6), red other driver units (D1 to D3), and one orange unit (U) to show the participant's vehicle position. Each trial lasted a maximum of 10 seconds, and participants were instructed to choose quickly.

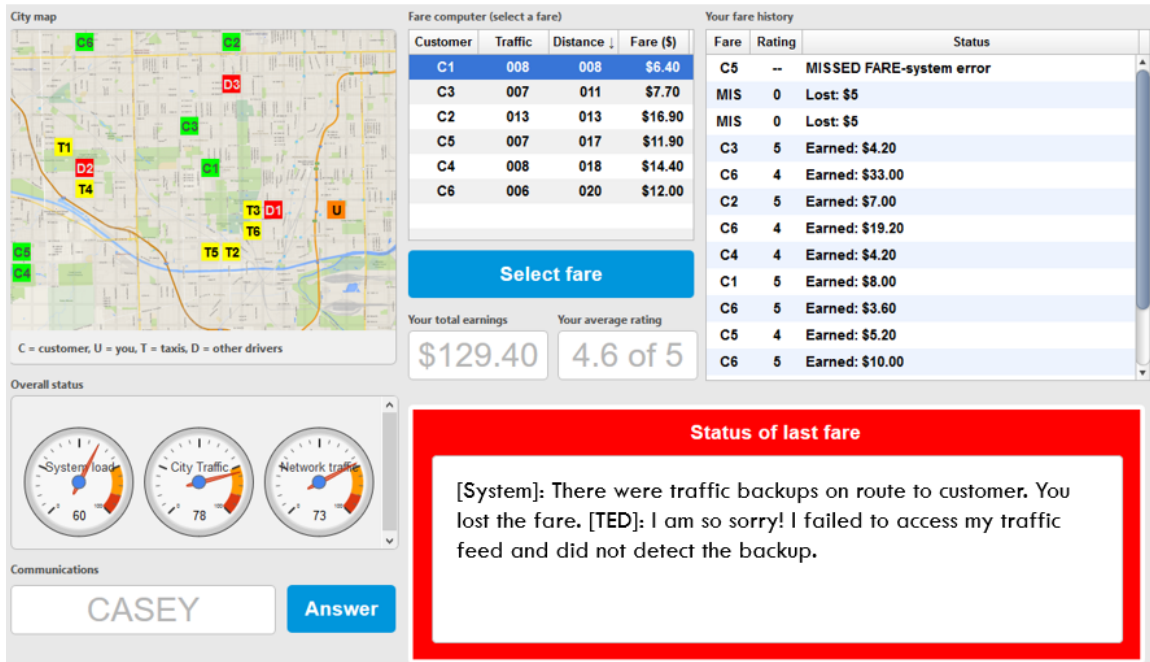


Figure 1. Task interface during automation failure.

To help participants select fares, automation sorted the customers in descending order by distance from the participant’s vehicle. If a participant fails to select a fare, the system status panel turned yellow and informed the participant that they have been fined \$5 for that trial. When the automation failed, the status panel turned red and an error message was displayed. If the trial was successful, the status panel turned green and displayed the fare and rating earned from that customer. See Appendix B for examples of task interfaces.

To increase the general level of workload and ensure that participants use the automation (Parasuraman & Manzey, 2010), participants also completed a secondary task. Participants monitored the communications panel for the target call sign “MASON”. The communications panel randomly displayed 1 of 14 different call signs every six seconds. Participants were instructed to click the “Answer” button when the

target call sign appeared. Following a successful answer, the communications panel turned green. Participants were told to divide their attention equally between the primary and secondary task.

Design

The experiment used a 2 (failure type: competency, integrity) \times 2 (repair strategy: apology, denial) mixed-factors design. Failure type (competency and integrity) was presented within subjects and repair strategy was presented between subjects. Previous research identified a reliability of 70% as the threshold at which participants perceive a system to be reliable (Wickens & Dixon, 2007). Therefore, this study used automation reliability of 80% (i.e., 40 failures out of 200 trials per block, 80 failures total per participant). To build participant trust in the system, failures were not presented before 10th trial (Wickens & Xu, 2002). Afterward, failures were randomly presented immediately following participants' attempt to select a fare. Failure type and repair strategy were operationalized using non-interruptive error messages that immediately proceeded system failures.

Each error message was composed of situational context, failure type description, and a repair statement. To maximize anthropomorphism of the automation, message content was segmented such that the automation agent (TED) responded to general situational information. The source indicator “[System]:” was used for general situational information and “[TED]:” was used for automation agent failure descriptions and trust repair strategies.

Competency-based failure descriptions stated that automation failures resulted from system inability to accurately perform the task (e.g., “I miscalculated the route”). Integrity-based failure descriptions stated that automation failures resulted from system priorities that were misaligned with those of the participant. (e.g., “I routed a different driver to the fare”).

Repair strategy (apology and denial) provided information about the system’s intentionality and responsibility for failures. Apologies internally attributed error causality and acknowledged responsibility for the failure (e.g., “I am sorry!”). Thus, apologies repair trust by implying remorse for the failure and suggesting that the failure is unlikely to occur again in the future. Denials externally attributed causality and declare no responsibility for the failure (e.g., “It was not my fault!”). Thus, denials repair trust by suggesting that the system is not at fault and that there is no cause for trust in the system to be harmed by the failure.

To minimize the likelihood that participants would ignore error messages, this study used 5 different messages for each of the 4 conditions (20 total messages). Each message was constructed to have insignificantly different length, readability, and clarity. Example error messages for each experimental condition are shown in Table 1, and a full list of the error messages can be found in Appendix C.

Table 1

Example system error message, grouped by repair strategy and failure type.

Repair	Error Type	Example Message
Apology	Competency*	[System]: There were traffic backups on route to customer. You lost the fare. [TED]: I am so sorry! I failed to access my traffic feed and did not detect the backup.
Apology	Integrity	[System]: Customer requested a more experienced driver. You lost the fare. [TED]: My apologies! I routed taxi with a better rating to help the reputation of RideAway.
Denial	Competency	[System]: There were traffic backups on route to customer. You lost the fare. [TED]: It was not my fault! I miscalculate the route because of an unpredictable vehicle collision.
Denial	Integrity*	[System]: Customer requested the safest driver available. You lost the fare. [TED]: It was not my fault! RideAway required me to route a taxi with a safer driving record.

* Hypothesized optimal repair-failure combination.

Measures

Individual differences. In addition to age and gender demographics, complacency potential was collected in this study. Complacency refers to an individual's propensity to over-rely on unreliable automation and is associated with over-trust in such systems (Singh, Molloy, & Parasuraman, 1993). Individual differences in complacency potential could confound effects of failure type and repair strategy on trust. Thus, the Complacency Potential Rating Scale (CPRS; Singh, Molloy, & Parasuraman, 1993) was used to determine participants' propensity for automation-induced complacency. Participants rated their agreement with 16 statements about general automation using a 5-point Likert scale. The score of each item was tallied to provide an overall CPRS score. A CPRS score of 16 corresponds to low complacency potential while a score of 80 corresponds to high complacency potential.

History-based trust. History-based trust was measured using a questionnaire adapted from Lee and Moray (1994; see Appendix D). A single subjective question assessed trust perceptions by asking participants, “To what extent did you trust the automation in the scenario?” Participants answered using a 0-100 visual analogue scale, where higher scores indicated higher perceived trust.

Perceived competence and integrity. A 6-item questionnaire adapted from Kim, Dirks, Cooper, and Ferrin (2004) assessed perceptions of automation competence and integrity such that half of the items assessed each construct. Participants used a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree) to rate their agreement with each statement. An example item assessing perceived competence is “TED was very capable of performing its job” and an example for perceived integrity is “Sound principles guide the TED’s actions.” Higher scores on the scale indicate stronger perceptions of the construct. See Appendix E for the full questionnaire.

Causal attributions. Previous trust repair research (e.g., Kim et al., 2006) identified that locus of causality (internal or external) affects trust repair. Therefore, we measured participants’ causal attributions using a questionnaire based on Weiner’s (1985) attribution theory. Participants reported internal and external attributions by rating their agreement with three statements about causality on 7-point Likert scales (1 = strongly disagree, 7 = strongly agree). Each of these questions apportioned blame to one of three specific sources: the rater/participant, the automation (TED), or other factors (e.g., situational factors). An example item was, “Rate the degree to which errors were caused

by TED.” Higher scores on each question indicated stronger causal attributions to self, system, or other factors.

A fourth item provided further insight into participants’ causal perceptions by asking participants the extent to which they agreed with the statement, “The automation took responsibility for its actions.” Lower scores on this question indicated that participants perceived the automation was less responsible for failures, as should be observed following denials. Higher scores on this question indicated that participants perceived that the automation was more responsible for failures, as should be observed following a strong apology. See Appendix F for the full causal attribution questionnaire.

Exploratory variables. Perceived reliance, perceived automation usefulness, self-confidence, task time, and workload are known predictors of automation use and trust (Lee & Moray, 1994). Therefore, these variables were also examined in this study to indicate the presence of confounds. For example, significant variances in task time could indicate differences in readability, clarity, or length among the error messages used in each condition.

Subjective reliance, usefulness, and self-confidence were measured using the same format as subjective trust (e.g., 0-100 visual analogue scale, see Appendix D). Reliance was measured using the question “To what extent did you rely on (i.e., actually use) the automation aid in this scenario?” Usefulness was measured using the question, “To what extent do you think the automation improved your performance in this scenario compared to performance without the automation?” Self-confidence was measured by the question, “To what extent were you self-confident that you could successfully

perform without the automation aid in this scenario?” Task time was assessed by recording mean trial time across conditions. Last, workload was measured using the NASA-TLX (Hart & Staveland, 1988). This item measures subjective workload using 6 items, each assessing a different attribute of workload: perceived mental demand, perceived physical demand, perceived temporal demand, perceived effort, perceived performance, and perceived frustration. An example item from the NASA-TLX is “How mentally demanding was the task?”

Procedure

Each participant experienced one of four conditions, 2 (failure type: competence-based, integrity-based) \times 2 (repair strategy: apology, denial). Each condition was composed of two blocks of 200 trials (400 trials total), one block for competency failures and one block for integrity failures. Block presentation order was counterbalanced using a partial Latin square.

Participants were tested in groups of up to six. Each participant was randomly assigned a condition. After participants provided informed consent, we collected age and gender demographics and provided verbal instructions for the tasks. Participants were told that the automation in the task was highly reliable but imperfect. After confirming that they understood the instructions, participants completed a practice block composed of 20 trials. The automation in the practice block was 100% reliable to build trust in the system and minimize participant confusion while they familiarized themselves with the experimental tasks. The participants next completed two experimental blocks. Task time was recorded by the computer during each block. After each block, participants

completed the questionnaires for perceived trust, competence, integrity, responsibility, causal attributions, reliance, self-confidence, and usefulness. At the end of the study, participants completed the CPRS.

RESULTS

An a priori power analysis indicated a total of 64 participants would be required to detect a moderate effect size ($\eta_p^2 = .25$) with 90% power at an alpha of .05. A total of 74 participants were recruited for this study to account for data loss. Six participants were removed due to computer malfunctions that occurred during their session. Two participants were removed because they did not follow instructions (one used their cell phone during the task and the other began the task prior to receiving verbal instructions). One additional participant was removed because they were deemed an outlier due to a mean task time over 3 standard deviations from the sample mean. Data from the remaining 65 participants were included in the analyses of trust and exploratory variables. Descriptive statistics for participant characteristics by between-subjects condition are shown in Table 2. Chi-squared analysis revealed no significant ($p > .05$) differences between repair strategy groups in gender or age. Independent samples t-tests revealed no significant ($p > .05$) differences between repair strategy groups in age and complacency potential (measured with CPRS scores).

Table 2

Participant characteristics by repair strategy (the between-subjects factor).

	Apology				Denial			
	Male (<i>n</i> = 8)		Female (<i>n</i> = 10)		Male (<i>n</i> = 6)		Female (<i>n</i> = 13)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age	19.00	1.07	18.30	.48	19.50	1.38	18.38	.65
CPRS	57.63	7.11	56.40	4.97	60.83	5.27	57.85	7.55

Note. Age and gender statistics represent data from 37 participants. Complacency potential scores range from 18 to 80 such that higher scores indicating higher complacency potential and represent data from 65 participants.

All dependent variables were subjected to a 2 (failure type: competence, integrity) × 2 (repair strategy: apology, denial) mixed repeated-measures analyses of variance (ANOVA). The Type I error rate for all tests was set at .05, and all pairwise comparisons used Bonferroni degrees of freedom corrections. Measures of perceived competence, integrity, and causal attributions were considered manipulation checks and were measured for only the first 20 participants, producing underpowered statistical analysis of these variables (e.g., observed power of .35 < acceptable power of .8). Therefore, our discussion will primarily focus on the observed effect sizes, which are independent of sample size, using Cohen's (1988) conventions for eta-squared (.01, .06, .14 represent small, medium, and large effect sizes, respectively). History-based trust was the key variable of interest in this study and was the only variable for which a priori hypotheses were created. All other variables included in this analysis are considered in context of explaining trust effects. Summary tables for all main and interaction effects of repair strategy and failure type are displayed in Appendix I. In-text figures are only presented for significant results; however, figures for nonsignificant results are shown in Appendix J.

History-based Trust

We predicted that 1) trust in apologetic automation would be greater following competency-based system failures than integrity-based system failures, and 2) trust in denying automation would be greater following integrity-based system failures than competency-based system failures. A 2 (failure type: competence, integrity) \times 2 (repair strategy: apology, denial) mixed repeated-measures ANOVA revealed a significant 2-way interaction of failure type and repair strategy for history-based trust ($F(1, 63) = 10.80, p = .002, \eta_p^2 = .15$). The source of this interaction was a significant effect of apologies on trust such that participants had greater trust in automation that apologized for committing competency-based failures ($M = 69.68, SD = 20.08$) than that which apologized for committing integrity-based failures ($M = 61.61, SD = 24.10; F(1, 63) = 10.01, p = .002, \eta_p^2 = .15$). Trust in automation that denied committing competency-based failures ($M = 66.76, SD = 20.41$) was not significantly different than automation that denied committing integrity-based failures ($M = 71.18, SD = 19.19; 4, F(1, 63) = 2.83, p > .05, \eta_p^2 = .04$). In addition, there was no significant main effects on trust for repair strategy ($F(1, 63) = .47, p > .05, \eta_p^2 = .007$) or failure type ($F(1, 63) = .93, p > .05, \eta_p^2 = .01$; see Figure 2). These results supported our predicted interaction of failure type and repair strategy for apologies but not denials.

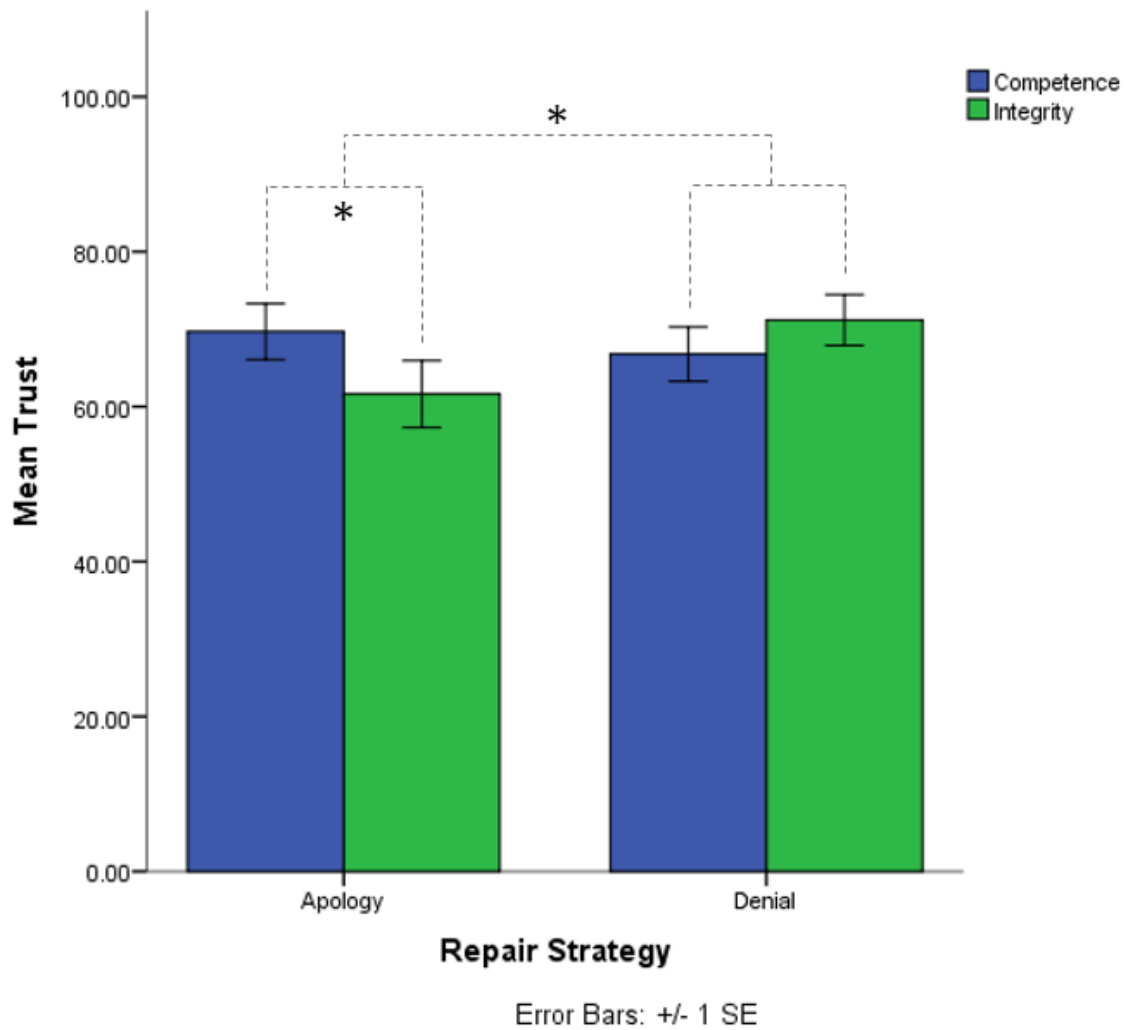


Figure 2. Mean trust comparisons between failure types, organized by repair strategies. Significant differences ($p < .05$) are indicated by an asterisk. Errors bars display +/- 1 standard error.

Perceived competence and integrity

Perceived competence and integrity are critical perceptions contributing to trust. We expected that competency-based automation failures more negatively affected perceived competence than perceived integrity. Similarly, we expected that integrity-based failures more negatively affected perceived integrity than perceived competence.

Reliability analyses were conducted for the subscales measuring perceived competence (Cronbach's $\alpha = .92$) and perceived integrity (Cronbach's $\alpha = .72$). Inter-item correlations were determined to be acceptably strong (Cronbach's $\alpha > .7$; Nunnally, 1978) and composite mean scales were used in the following analysis. Perceived competence and integrity were each subjected to a 2 (failure type: competence, integrity) \times 2 (repair strategy: apology, denial) mixed repeated-measures ANOVA. There were no significant main effects of failure type on perceptions of competence ($F(3, 16) = 1.46, p > .05, \eta_p^2 = .22$) or integrity ($F(3, 16) = .038, p > .05, \eta_p^2 = .01$). However, the observed effect sizes suggested a large effect of failure type on perceived competence ($\eta_p^2 = .22$) but a small effect on integrity. This finding was similar to previous research (e.g., Kim et al., 2004). The effect of failure type on perceived integrity was small ($\eta_p^2 = .01$) and therefore provides no evidence that failure type affected participants' perceptions of system integrity. Also, there was a large difference in the strength of system failure type effects on perceived competence and integrity, indicating that both failure types affected perceived system competence substantively more than perceived system integrity.

Causal attributions

We also examined the effect of trust repair strategy on causal attributions. These dependent variables were individually subjected to a 2 (failure type: competence, integrity) \times 2 (repair strategy: apology, denial) mixed repeated-measures ANOVA. Results indicated no significant main effects of repair strategy on causal attributions to the automation ($F(1, 18) = .275, p > .05, \eta_p^2 = .02$), causal attributions to self ($F(1, 18) = .25, p > .05, \eta_p^2 = .01$), or perceived responsibility ($F(1, 18) = 4.10, p > .05, \eta_p^2 = .18$).

Although not significantly different, mean attributions toward the automation and mean perceived responsibility trended in the predicted directions (e.g., apologetic automation was perceived as more responsible and more at fault for failures). There was a significant main effect of trust repair strategy on external attributions such that participants attributed more fault to the situation when the automation denied committing failures ($M = 5.40$, $SD = 1.22$) than when it apologized for committing failures ($M = 4.00$, $SD = 1.62$; $F(1, 18) = 4.78$, $p = .042$, $\eta_p^2 = .21$; see figure 3). These findings indicate that apologies and denials were differentially perceived by participants consistent with previous literature (e.g., Kim et al., 2004).

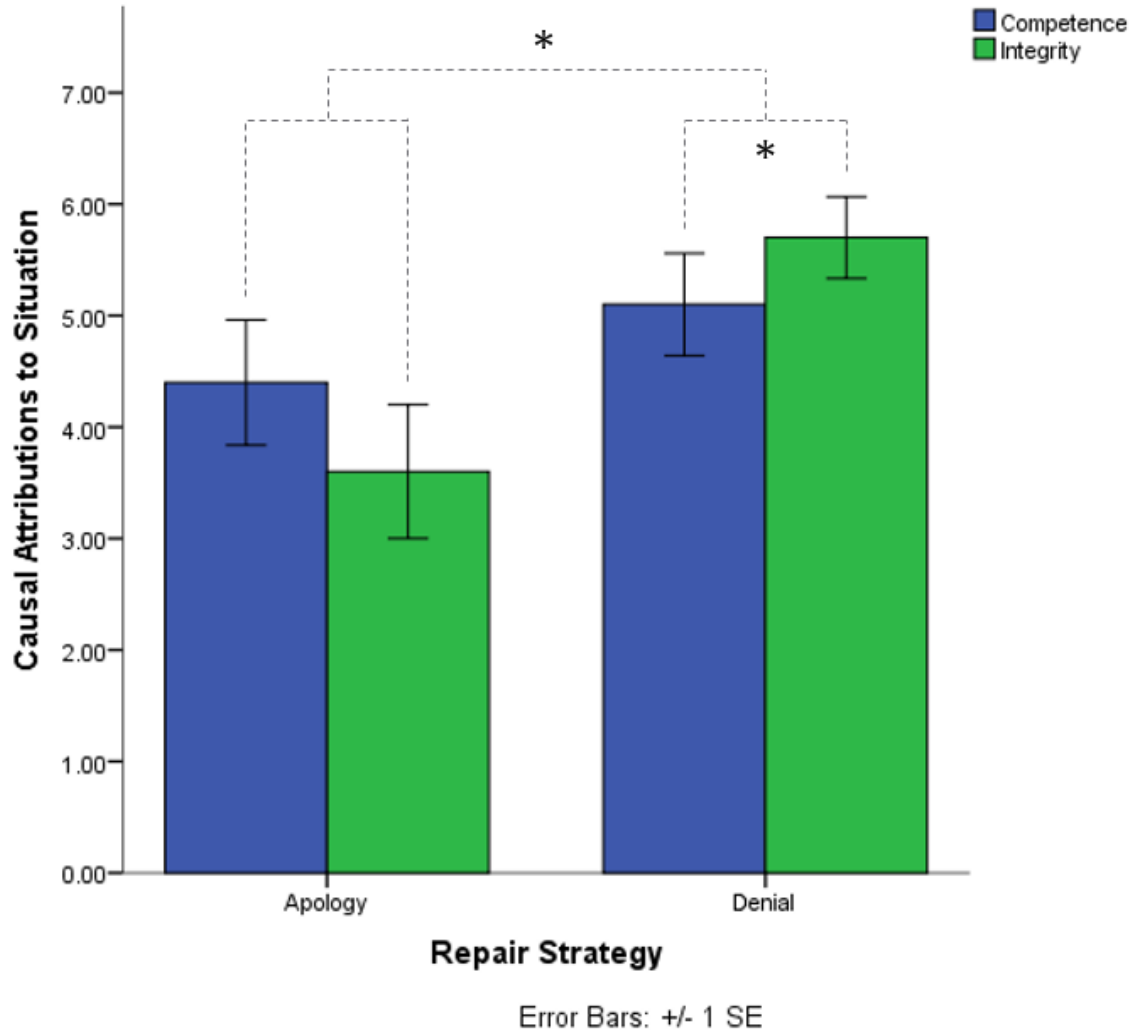


Figure 3. Mean situation attributions between failure types, organized by repair strategies. Significant differences ($p < .05$) are indicated by an asterisk. Errors bars display +/- 1 standard error.

Exploratory Variables.

Perceived usefulness, perceived reliance, self-confidence, task time, and subjective workload (measured by TLX) predict of automation use and were also measured and analyzed. Each of these dependent variables was subjected to a 2 (failure

type: competence, integrity) \times 2 (repair strategy: apology, denial) mixed repeated-measures ANOVA. No significant ($p > .05$) main effects or interactions of failure type and repair strategy were detected for perceived usefulness, perceived reliance, self-confidence, and task time. This suggests that apologies and denials did not differentially affect behavior or associated perceptions. The 6 TLX subscales were analyzed separately, since each subscale measures a different component workload: mental, physical, temporal, effort, performance, and frustration. There was a significant main effect of repair strategy on perceived physical workload ($F(1, 42) = 6.60, p = .014, \eta_p^2 = .136$), indicating that denials produced lower perceptions of physical demand than apologies (see figure 4). We conclude that apologies and denials do not differentially affect perceptions, with the exception of perceived workload and trust, that are critical to reliance and use of automation.

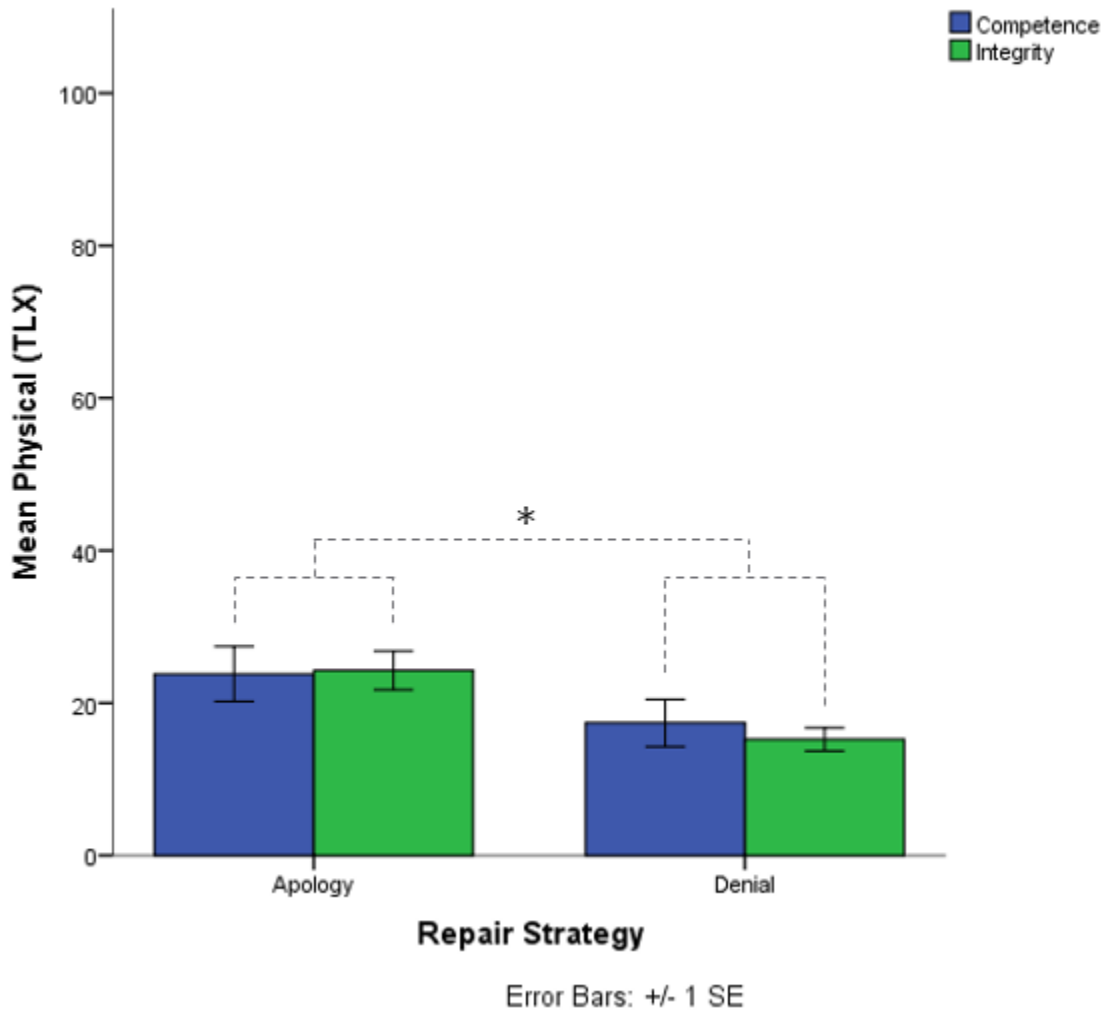


Figure 4. Mean perceived physical workload between failure types, organized by repair strategies. Significant differences ($p < .05$) are indicated by an asterisk. Errors bars display +/- 1 standard error.

DISCUSSION

The aim of the current study was to determine whether the effects of system apologies and denials on user trust depend on system failure type during human-automation interactions. We hypothesized that a) apologies better repair trust harmed by

competence-based failures than trust harmed by integrity-based failures, and b) denials better repair trust harmed by integrity-based failures than trust harmed by competence-based failures.

Our results supported our first hypothesis and revealed that apologies repair competence-based failures but damage trust following integrity-based failures. These data replicate interpersonal trust repair literature (Kim et al., 2004) in a human-automation context. We also support early human-automation trust repair research examining apologies in contexts containing only competence-based failures (de Visser et al., 2016; Robinette et al., 2015). Importantly, we provide new insight into the extent to which system transparency is and perceptions of automation's intent are critical to trust repair in human-automation interactions (Schaefer et al., 2017; de Visser et al., 2018). User trust in automation is repaired when the user's values align with their perceptions of the machine's intent. For example, automation apologies for incompetence repair user trust by signaling remorse that the error occurred and indicates the intention of not repeating the error in the future. However, automation apologies for violating user values fail to repair trust because they admit to intentionally committing errors and are perceived as likely to repeat in the future. Therefore, strong, lasting, damaging effects of integrity-based trust violations exist in both interpersonal interactions (Schweitzer, Hershey, & Bradlow, 2006; Reeder & Coovert, 1986) and human-automation interactions.

Our second hypothesis was not supported, which suggests that the effects of denials on trust do not significantly depend on failure type. Interestingly, this means that trust repair in human-automation interactions is similar, but not identical, to human-

human interactions. A possible explanation can be found by examining implicit biases that differentiate how people perceive automation and human performance. Automation is viewed as a more credible information source than human partners in the same context (Lerch, Preitula, & Kulik, 1997; Petty and Cacioppo, 1986). People are more likely to trust and comply with statements made by automation than statements made by other people (Mosier, Skitka, & Heers, 1996). Because credibility enhances the effectiveness of denials (Shapiro et al., 1994), these data suggest that automation denials more effectively repair competence-based trust than human denials. Furthermore, causal attributions rely on superficial situational cues when causal information is scarce or absent (Lewicki & Bunker, 1996). Also, users experiencing conditions, such as high workload, that impair their ability to verify machine performance (Parasuraman & Riley, 1997) are similarly likely to over-rely on superficial information to update trust.

An interesting relationship exists between workload and trust repair; however, the extent to which this may impact automation use remains unclear. Participants interacting with apologetic automation experienced significantly greater physical task demand than participants interacting with automation denials. This finding supports speculation from emerging trust repair literature (de Visser et al., 2018) that workload and individual differences in cognition (e.g., working memory capacity) may have an important effect on trust repair. At least one alternative trust repair strategy, perspective-taking, has been shown to be cognitively demanding (Roßnagel, 2000). Together, these findings suggest that repairs that require greater inferences to be made about intentionality and mind of the trustee may be more demanding than less inferential repair strategies, such as denials.

Exploring the performance of repair strategies under varying levels of workload may prove an interesting and important area for future research.

Automation has become widely available and already uses trust repair in response to system failures. The effects of system competence have been the focus of much research (Muir & Moray, 1996; Waytz, Heafner, & Epley, 2014); however, more work is needed to further understand how integrity failures affect user trust. Modern automation should consider the impact that failure types have on user perceptions and, when trust repair is advantageous, take measures to preserve perceptions of system integrity.

Although apologies have a history of success at repairing trust, they do expose the human-automation team to risk of confirming otherwise ambiguous perceived integrity-based system failures. Denials may prove to be a safer repair strategy to use under high workload conditions, when system integrity is in question, or when failure causes are not verifiable. Although not yet empirically tested, sound theory discussed in this paper suggests that anthropomorphism may affect trust repair. Anthropomorphism is an important consideration as automation continues to progress toward autonomy and display further degrees of humanness. Trust repair in autonomous systems may prove to be more similar to human-automation interaction (de Visser et al., 2018) than the automation tested in the current study. More work is needed to further delineate the boundaries of trust repair in human-automation interaction.

Limitations and Future Directions

Future work should address several limitations which affected the generalizability of our results. First, trust is increasingly difficult to repair as failures become more

frequent (Lewicki & Bunker, 1996). Therefore, system reliability is likely a key determinant of trust repair. Denials may continue to effectively repair trust during lower levels of reliability by displacing blame for both types of failures, whereas apologies may lose effectiveness (or even be perceived as meaningless) due to repeated admittance of guilt and low perceived sincerity.

A second limitation of this study is the single task domain (taxi dispatching). Trust in automation (Pak, Rovira, McLaughlin, 2017; Zand, 1972) and perceptions affecting trust (e.g., perceived competence and integrity; Schoorman, Mayer, & Davis, 1995) are domain-specific. Thus, it is unknown how our findings may generalize to other contexts. Also, the app-based rideshare domain was chosen because it closely resembled a nascent technology familiar to undergraduate participants. However, we did not explore how prior knowledge of the task or domain may have influenced trust repair.

Further research should explore the effects of trust repair strategy content and length. Substantive messages are more likely to repair trust than simple messages (Bottom et al., 2002). For example, Kim and colleagues' (2004) utilized detailed repairs, whereas our study used concise and simple error messages. We opted for concise messages to increase generalizability of our findings, because lengthy messages are unlikely in most currently available applications and are potentially unsafe if used in such settings (e.g., in-vehicle GPS). Our automation error messages were therefore short enough to be quickly read but long enough that they provided adequate information about the failure context and repair intervention.

Last, our study tested only two of many possible repair strategies as well as two of many possible failure types. Additional research is needed to further explore the effects of apologies and denials as well as a wide variety of alternative repairs. For example, future work may choose to examine simple acknowledgements as a quick, low-cost repair (e.g., “Something went wrong!”). Further analysis of the effects of failures types are also needed, particularly for trust repair effects on perceived benevolence, the third dimension of trust (Mayer et al. 1995). As automation becomes increasingly humanlike, it is possible that perceptions such as benevolence may play a greater role in how users trust systems.

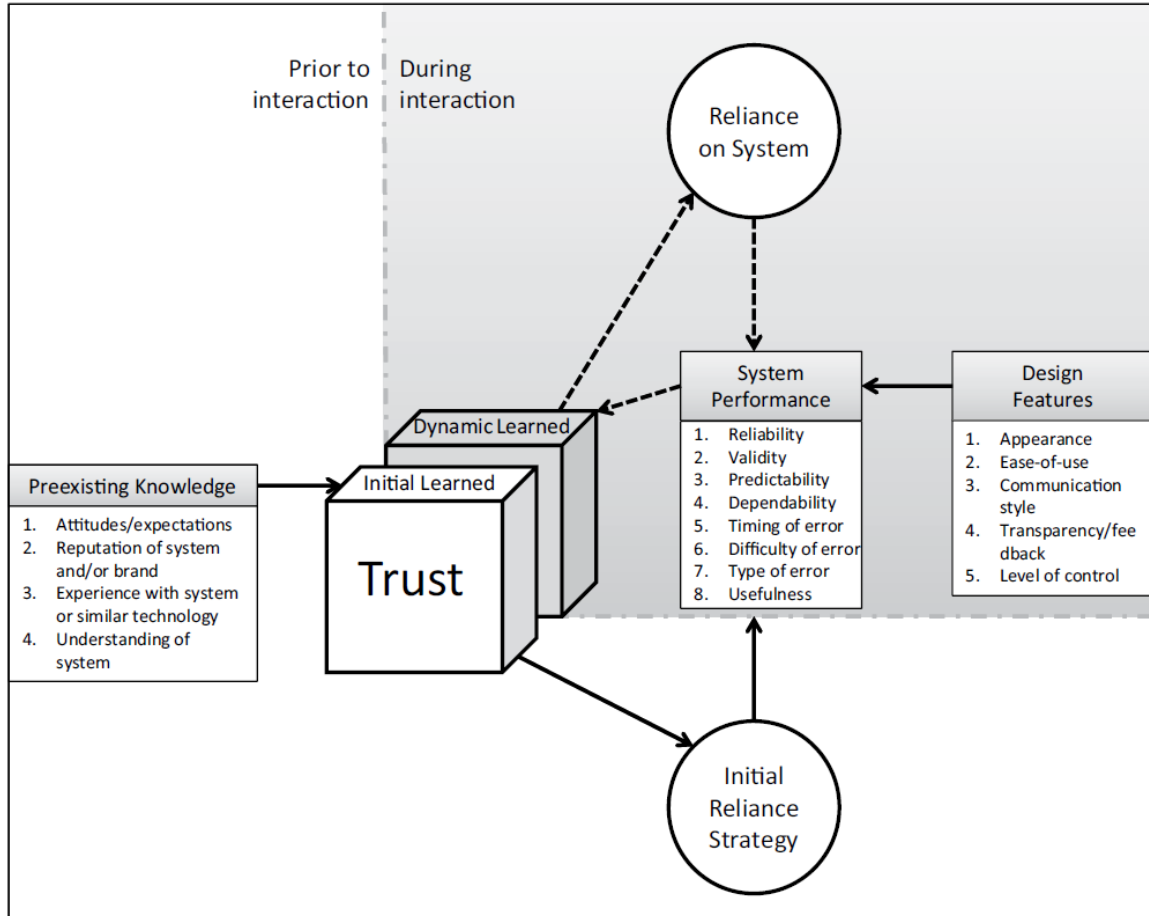
Conclusion

Currently, automation design and performance interventions can manipulate the formation and resilience of trust. However, little is known about repairing trust following system failures. This study is among the first to explore trust repair effects in human-automation interactions and has provided useful insights into the contours of trust repair in this domain. Unlike visual anthropomorphism, which is sensitive to individual user differences and system reliability, system communication provides a more resistant design feature that could be implemented flexibly based on system reliability. This allows a user’s perception of the automation to ebb and flow with automation performance, enabling people to adjust their usage of automation more appropriately in real-time.

APPENDICES

Appendix A

Hoff and Bashir's (2015) model illustrating factors that affect automation trust.



Appendix B

Example task interfaces

Task interface immediately after a participant fails to select in the allotted 10 seconds.

City map

Fare computer (select a fare)

Customer	Traffic	Distance ↓	Fare (\$)
C5	010	005	\$5.00
C3	007	006	\$4.20
C2	006	007	\$4.20
C1	006	011	\$6.60
C6	005	016	\$8.00
C4	012	025	\$30.00

Your fare history

Fare	Rating	Status
MIS	0	Lost: \$5
C3	5	Earned: \$4.20
C6	4	Earned: \$33.00
C2	5	Earned: \$7.00
C6	4	Earned: \$19.20
C4	4	Earned: \$4.20
C1	5	Earned: \$8.00
C6	5	Earned: \$3.60
C5	4	Earned: \$5.20
C6	5	Earned: \$10.00
C1	5	Earned: \$1.20
C3	4	Earned: \$22.80

Your total earnings: \$134.40 **Your average rating:** 4.5 of 5

Overall status

System load: 44 City Traffic: 78 Network traffic: 68

Communications

NATALEE Answer

Status of last fare

You waited too long. You lost \$5 and got a 0-rating on this trial

Task interface immediately following an automation failure.

City map

Fare computer (select a fare)

Customer	Traffic	Distance ↓	Fare (\$)
C5	023	003	\$6.90
C6	014	006	\$8.40
C2	010	010	\$10.00
C4	010	010	\$10.00
C1	017	011	\$18.70
C3	008	012	\$9.60

Your fare history

Fare	Rating	Status
C6	4	Earned: \$19.00
C4	5	Earned: \$9.60
C3	4	Earned: \$16.90
C4	4	Earned: \$3.90
C5	4	Earned: \$3.50
C6	4	Earned: \$12.10
C6	4	Earned: \$15.40
C3	4	Earned: \$9.90
C5	--	MISSED FARE-system error
C1	4	Earned: \$2.80
C4	4	Earned: \$11.20
C5	4	Earned: \$16.80

Your total earnings: \$253.80 **Your average rating:** 4.6 of 5

Overall status

System load: 49 City Traffic: 71 Network traffic: 67

Communications

MASON Answer

Status of last fare

Success! Your passenger is happy! This rider gave you a rating of 4

Task interface immediately after successfully responding to the call sign “Mason” in the communications panel.

City map

C = customer, U = you, T = taxis, D = other drivers

Fare computer (select a fare)

Customer	Traffic	Distance	Fare (\$)
C5	008	003	\$2.40
C6	002	005	\$1.00
C2	010	009	\$9.00
C1	011	012	\$13.20
C3	005	012	\$6.00
C4	013	020	\$26.00

Select fare

Your total earnings: \$465.30
Your average rating: 4.6 of 5

Your fare history

Fare	Rating	Status
C2	4	Earned: \$28.90
C4	5	Earned: \$7.50
C5	5	Earned: \$7.70
C1	5	Earned: \$19.00
C5	4	Earned: \$16.90
C3	--	MISSED FARE-system error
C5	5	Earned: \$6.50
C1	4	Earned: \$25.20
C1	4	Earned: \$9.00
C2	5	Earned: \$33.60
C1	--	MISSED FARE-system error
C3	4	Earned: \$7.80

Overall status

- System load: 46
- City Traffic: 81
- Network traffic: 74

Communications

MASON Answered

Status of last fare

Success! Your passenger is happy! This rider gave you a rating of 4

Appendix C

Full list of automation error messages.

Integrity failures with apologies

- [System]: Customer requested a faster pickup time. You lost the fare. [TED]: I am so sorry! I routed a closer driver because the needs of RideAway customers come first.
- [System]: Customer requested a more experienced driver. You lost the fare. [TED]: My apologies! I routed taxi with a better rating to help the reputation of RideAway.
- [System]: Customer requested a more luxurious vehicle. You lost the fare. [TED]: I apologize! I routed a luxury taxi to keep the RideAway customer happy.
- [System]: Customer requested the safest driver available. You lost the fare. [TED]: I am sorry! I routed a driver to help improve the overall reputation of RideAway.
- [System]: Customer required more seats than you have. You lost the fare. [TED]: Sorry! I routed a higher occupancy taxi to keep the RideAway customer happy.

Integrity failures with denials

- [System]: Customer requested a faster pickup time. You lost the fare. [TED]: It was not my fault! RideAway forced me to send a closer driver to the customer.
- [System]: Customer requested a more experienced driver. You lost the fare. [TED]: It was out of my control! RideAway made me to route a taxi with a better rating.
- [System]: Customer requested a luxury vehicle. You lost the fare. [TED]: I was not responsible! RideAway required me to route the most luxurious taxi available.
- [System]: Customer requested the safest driver available. You lost the fare. [TED]: It was not my fault! RideAway required me to route a taxi with a safer driving record.
- [System]: Customer required a high occupancy vehicle. You lost the fare. [TED]: I was not at fault! RideAway made me send a taxi with more seats than you have.

Competency failures with apologies

- [System]: There were traffic backups on route to customer. You lost the fare. [TED]: I am so sorry! I failed to access my traffic feed and did not detect the backup.
- [System]: Road construction caused delays on route to customer. You lost the fare. [TED]: My apologies! I experienced a GPS failure and was unable to detect the detours.
- [System]: Rush hour traffic increased estimated travel time. You lost the fare. [TED]: I apologize! I failed to predict the traffic and miscalculated your time of arrival.

- [System]: Weather conditions became severe on route to customer. You lost the fare. [TED]: I am sorry! I lost my internet connection and failed to detect the weather change.
- [System]: Could not connect to RideAway server. You lost the fare. [TED]: Sorry! I experienced an internal error and could not route your taxi to the customer.

Competency failures with denials

- [System]: There were traffic backups on route to customer. You lost the fare. [TED]: It was not my fault! I miscalculate the route because of an unpredictable vehicle collision.
- [System]: Highway construction caused delays on route to customer. You lost the fare. [TED]: It was out of my control! I miscalculated because the road construction was unplanned.
- [System]: Rush hour traffic increased estimated travel time. You lost the fare. [TED]: I was not responsible! I inaccurately estimated the drive time because traffic is heavier than usual.
- [System]: Weather conditions worsened on route. You lost the fare. [TED]: It was not my fault! I failed to detect the weather change because of a poor internet connection.
- [System]: Could not connect to the customer. You lost the fare. [TED]: I was not at fault! I was unable to confirm the route due to an external server error.

Appendix D

History-based trust questionnaire adapted from Lee and Moray (1994).

Please answer the questions below about the computer aid (automation) by clicking on the scale:

To what extent did you trust (i.e. believe in the accuracy of) the automation aid in this scenario?

Not at all Extremely

To what extent did you rely on (i.e. actually use) the automation aid in this scenario?

Not at all Extremely

To what extent were you self-confident that you could successfully perform without the automation aid in this scenario?

Not at all Extremely







To what extent do you think the automation improved your performance in this scenario compared to performance without the automation?

Not at all Extremely

Appendix E

NASA-TLX measure of workload (Hart & Staveland, 1988).

Regarding the task you just completed, please answer the questions below by clicking on the scale:

<p>How mentally demanding was the task?</p> <p>Low High</p> 	<p>How successful were you in accomplishing what you were asked to do?</p> <p>Not very Very</p> 
<p>How physically demanding was the task?</p> <p>Low High</p> 	<p>How hard did you have to work to accomplish your level of performance?</p> <p>Low High</p> 
<p>How hurried or rushed was the pace of the task?</p> <p>Not very Very</p> 	<p>How insecure, discouraged, irritated, stressed, and annoyed were you?</p> <p>Low High</p> 

Appendix F

Perceived competence and integrity questionnaire adapted from Kim, Dirks, Cooper, and Ferrin (2004). All questions used a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The first group of questions assessed perceived competence and the second group of question assessed perceived integrity.

Please rate the degree to which you disagree or agree with the following statements.

1. TED is very capable of performing its job.
2. TED has much information about the work that needs to be done.
3. I feel very confident about TED's abilities.

Please rate the degree to which you disagree or agree with the following statements.

1. I like TED's values.
2. Sound principles seem to guide TED's actions.
3. TED has a great deal of integrity.

Appendix G

Causal Attributions Questionnaire

Causal attributions were assessed on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree).

Please rate the extent to which you disagree or agree with the following statements.

1. The system errors were due to TED's action.
 - a. Please explain your answer for the previous question in 140 characters or less.
2. The system errors were due to my action.
 - a. Please explain your answer for the previous question in 140 characters or less.
3. The system errors were due to something other than TED or my actions.
 - a. Please explain your answer for the previous question in 140 characters or less.
4. The automation took responsibility for its actions.
 - a. Please explain your answer for the previous question in 140 characters or less.

Appendix H

Complacency Potential Rating Scale (CPRS). Adapted from Singh, Molloy, & Parasuraman, 1993). All questions used a 1-5 Likert scale (1 = Strong Disagree, 5 = Strongly Agree).

Please read each statement carefully and select the one response that you feel most accurately describes your views and experiences. There are no right or wrong answers. Please answer honestly.

1. Manually sorting through card catalogues is more reliable than computer-aided searches for finding items in a library.
2. If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because computerized surgery is more reliable and safer than manual surgery.
3. People save time by using automatic teller machines (ATMs) rather than a bank teller in making transactions.
4. I do not trust automated devices such as ATMs and computerized airline reservation systems.
5. People who work frequently with automated devices have lower job satisfaction because they feel less involved in their job than those who work manually.
6. I feel safer depositing my money at an ATM than with a human teller.
7. I have to tape an important TV program for a class assignment. To ensure that the correct program is recorded, I would use the automatic programming facility on my VCR rather than manual taping.
8. People whose jobs require them to work with automated systems are lonelier than people who do not work with such devices.
9. Automated systems used in modern aircraft, such as the automatic landing system, have made air journey safer.
10. ATMs provide safeguard against the inappropriate use of an individual's bank account by dishonest people.
11. Automated devices used in aviation and banking have made work easier for both employees and customers.
12. I often use automated devices.
13. People who work with automated devices have greater job satisfaction because they feel more involved than those who work manually.
14. Automated devices in medicine save time and money in the diagnosis and treatment of disease.

15. Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass a police radar speed-trap in case the automatic control is not working properly.
16. Bank transactions have become safer with the introduction of computer technology for the transfer of funds.
17. I would rather purchase an item using a computer than have to deal with a sales representative on the phone because my order is more likely to be correct using the computer.
18. Work has become more difficult with the increase of automation in aviation and banking.
19. I do not like to use ATMs because I feel that they are sometimes unreliable.
20. I think that automated devices used in medicine, such as CAT-scans and ultrasound, provide very reliable medical diagnosis.

Appendix I

Summary tables for descriptive and inferential statistics of main effects and interactions.

Table 3
Descriptive and inferential statistics for the main effects of failure type.

Measures	Competence		Integrity		<i>n</i>	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
Trust	68.15	20.15	66.62	22.03	65	1, 63	.926	.340	.014
Competence	4.77	1.48	4.80	1.24	20	1, 18	.028	.869	.002
Integrity	4.33	1.28	4.28	.95	20	1, 18	1.04	.321	.055
Causal Attributions									
Auto.	4.65	1.18	3.85	1.69	20	1, 18	4.40	.050	.196
Self	3.40	1.76	3.35	1.69	20	1, 18	.02	.896	.001
Other	4.75	1.62	4.65	1.87	20	1, 18	.10	.755	.006
Resp.	3.75	1.77	4.00	1.69	20	1, 18	.76	.396	.040
Task Time	3.96	1.37	3.69	1.14	65	1, 63	2.99	.089	.045
Self-Conf.	51.23	22.81	53.23	24.63	65	1, 63	.64	.426	.010
Usefulness	74.15	21.50	71.85	21.64	65	1, 63	1.40	.241	.022
Reliance	74.15	24.62	76.31	22.26	65	1, 63	1.05	.310	.016
Workload									
Mental	59.78	22.71	57.05	22.26	44	1, 42	1.57	.217	.036
Physical	20.22	15.74	19.55	10.56	44	1, 42	.114	.738	.003
Temporal	72.00	19.02	69.32	21.72	44	1, 42	.56	.460	.013
Effort	52.00	22.42	45.91	19.80	44	1, 42	2.86	.098	.064
Perf.	75.56	14.70	76.59	14.78	44	1, 42	.235	.630	.006
Frustration	38.67	24.18	40.68	22.14	44	1, 42	.420	.521	.010

Note. Task time displayed in seconds.

Table 4

Descriptive and inferential statistics for the main effects of repair strategy.

Measures	Apology		Denial		<i>n</i>	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
Trust	65.65	22.37	68.97	19.79	65	1, 63	.47	.495	.007
Competence	4.77	1.39	4.80	1.34	20	1, 18	.00	.955	.003
Integrity	4.12	.97	4.50	1.24	20	1, 18	.82	.378	.043
Causal Attributions									
Auto.	4.50	1.57	4.00	1.41	20	1, 18	.87	.365	.046
Self	3.55	1.82	3.20	1.61	20	1, 18	.25	.620	.014
Other	4.00	1.84	5.40	1.31	20	1, 18	4.78*	.042	.210
Resp.	4.55	1.64	3.20	1.54	20	1, 18	4.07	.059	.184
Task Time	3.69	1.12	3.95	1.38	65	1, 63	.891	.349	.014
Self-Conf.	53.87	23.63	50.74	23.77	65	1, 63	.34	.562	.005
Usefulness	71.94	23.87	73.97	19.25	65	1, 63	.16	.688	.003
Reliance	73.06	25.84	77.21	20.94	65	1, 63	.06	.804	.001
Workload									
Mental	60.48	20.71	56.60	23.89	44	1, 42	.31	.581	.007
Physical	24.05	14.15	16.17	11.53	44	1, 42	6.60*	.014	.136
Temporal	75.00	17.70	66.81	21.88	44	1, 42	2.24	.142	.051
Effort	53.10	19.81	45.32	22.05	44	1, 42	2.22	.143	.050
Perf.	74.29	15.79	77.66	13.55	44	1, 42	.77	.385	.018
Frustration	42.86	23.19	36.81	22.85	44	1, 42	1.05	.312	.024

Note. Task time displayed in seconds.* $p < .05$

Table 5
Inferential statistics for the interaction effects of repair strategy and failure type.

Measures	<i>n</i>	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
Trust	65	1, 63	10.80**	.002	.146
Competence	20	1, 18	3.36	.083	.157
Integrity	20	1, 18	1.04	.321	.055
Causal Attributions					
Auto.	20	1, 18	.28	.607	.015
Self	20	1, 18	.02	.896	.001
Other	20	1, 18	4.90*	.040	.214
Resp.	20	1, 18	.03	.864	.002
Task Time	65	1, 63	.03	.858	.001
Self-Conf.	65	1, 63	.00	.980	.000
Usefulness	65	1, 63	.71	.404	.011
Reliance	65	1, 63	.06	.804	.001
Workload					
Mental	44	1, 42	.85	.361	.020
Physical	44	1, 42	.28	.601	.007
Temporal	44	1, 42	.56	.460	.013
Effort	44	1, 42	.04	.845	.001
Perf.	44	1, 42	.24	.630	.006
Frustration	44	1, 42	.42	.521	.010

Note. Task time displayed in seconds.

* $p < .05$

** $p < .01$

Appendix J

Figures for all nonsignificant effects.

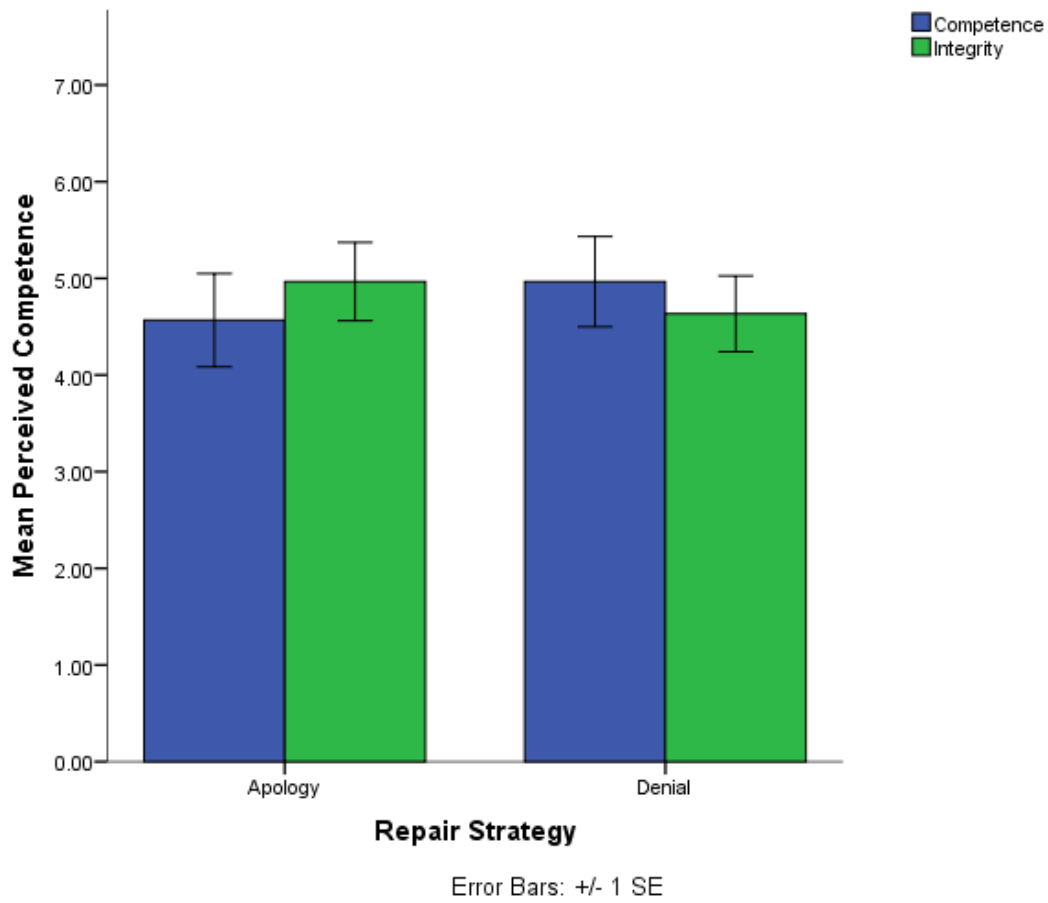


Figure 5. Mean perceived competence between failure types, organized by repair strategies.

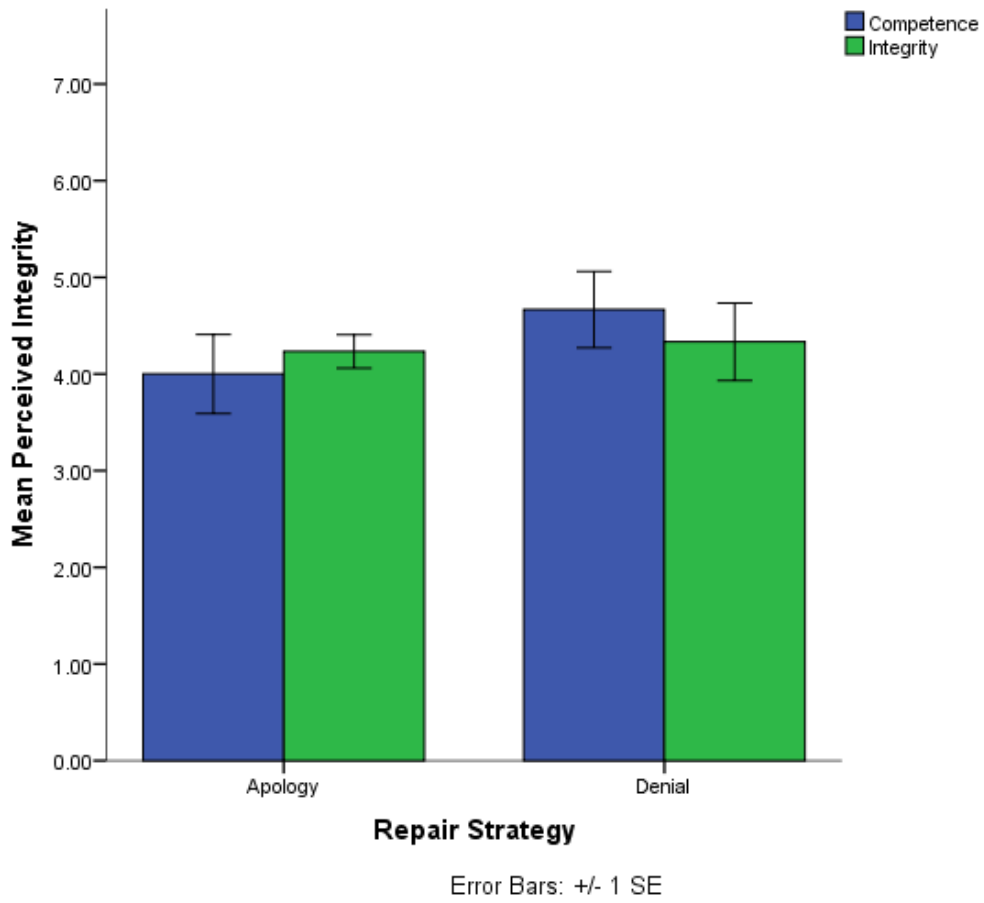


Figure 6. Mean perceived integrity between failure types, organized by repair strategies.

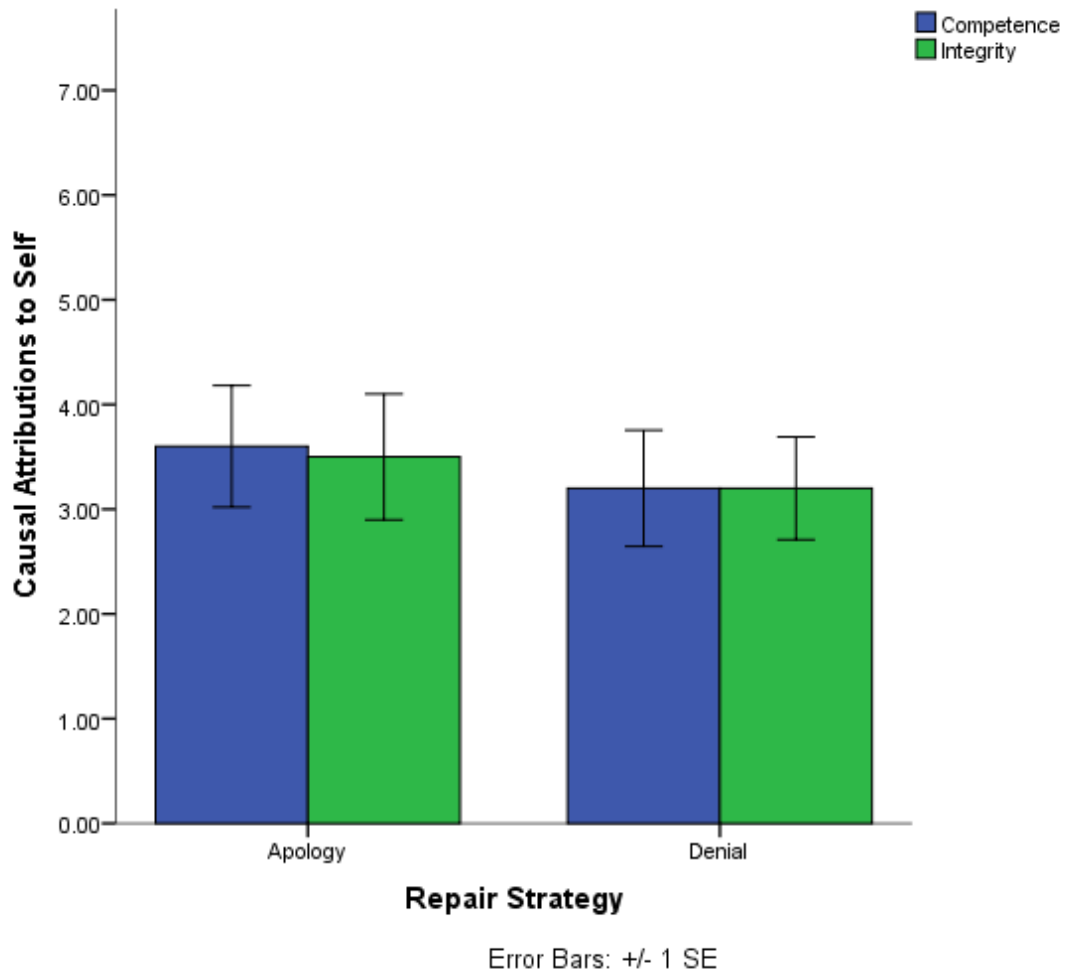


Figure 7. Mean perceived causal attributions to self between failure types, organized by repair strategies.

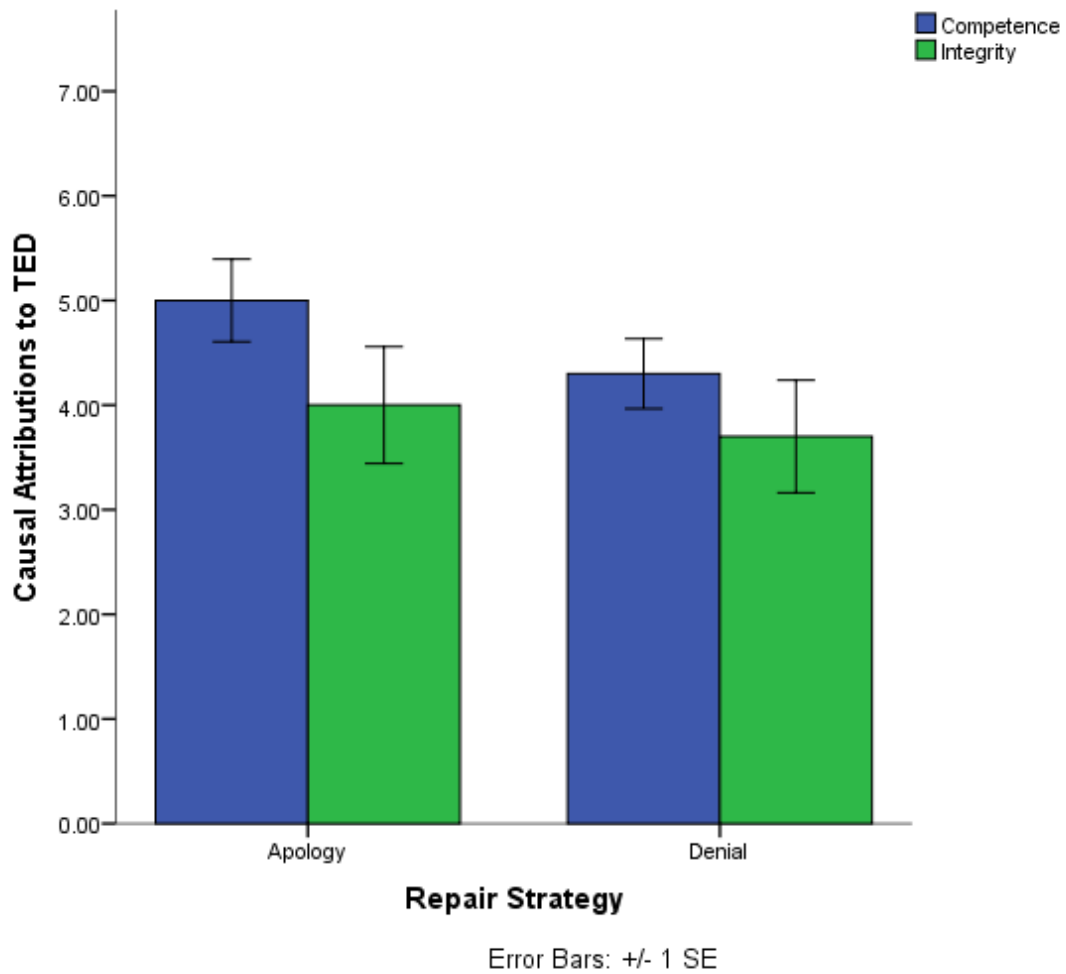


Figure 8. Mean perceived causal attributions to the automation (TED) between failure types, organized by repair strategies.

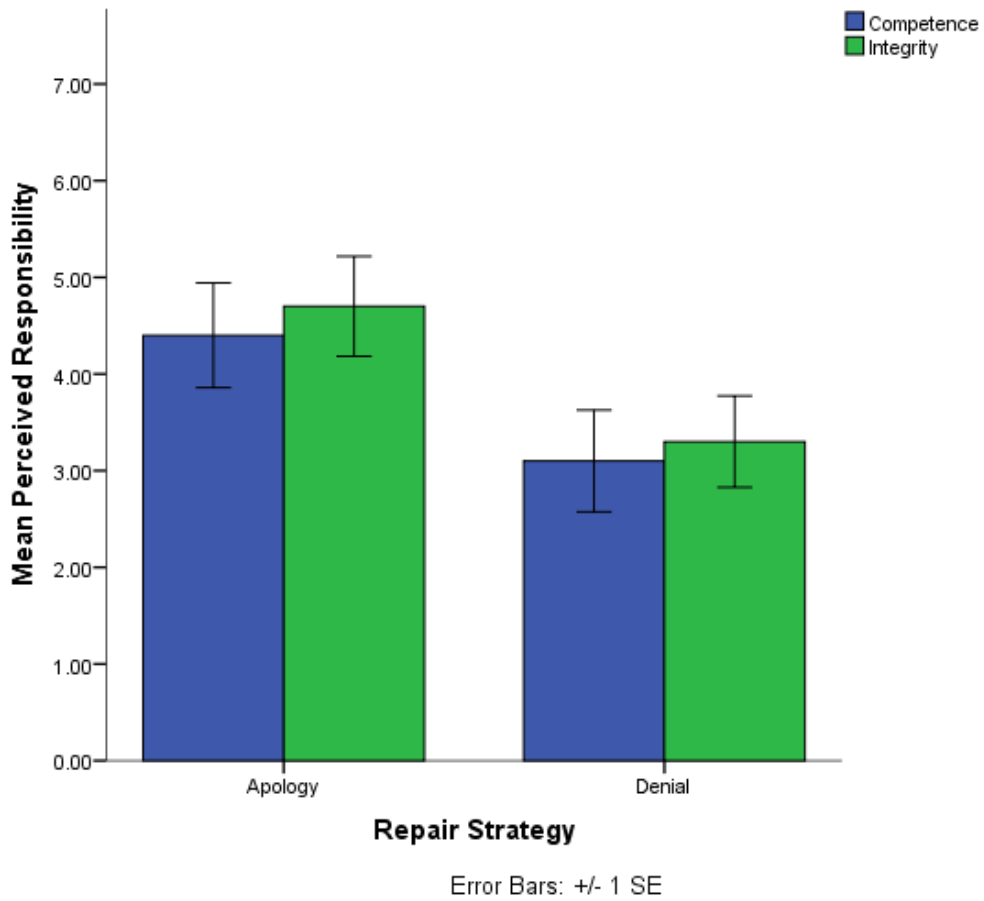


Figure 9. Mean perceived responsibility between failure types, organized by repair strategies.

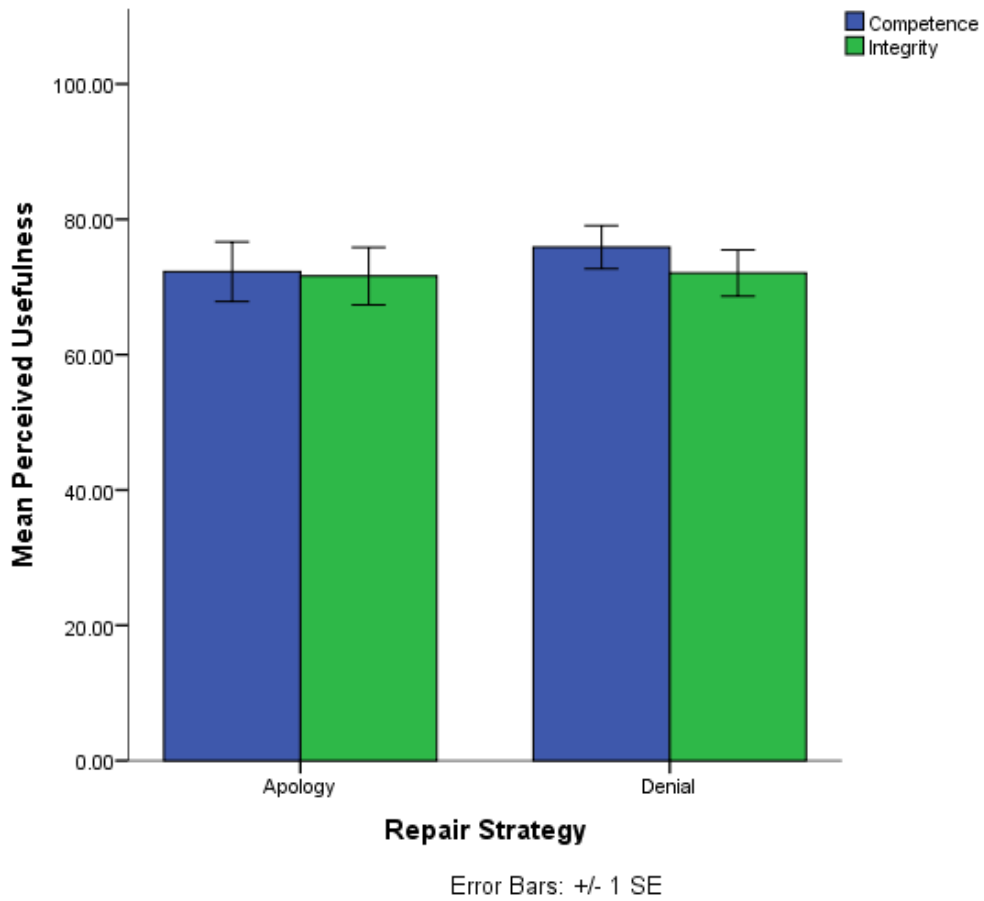


Figure 10. Mean perceived usefulness of the automation between failure types, organized by repair strategies.

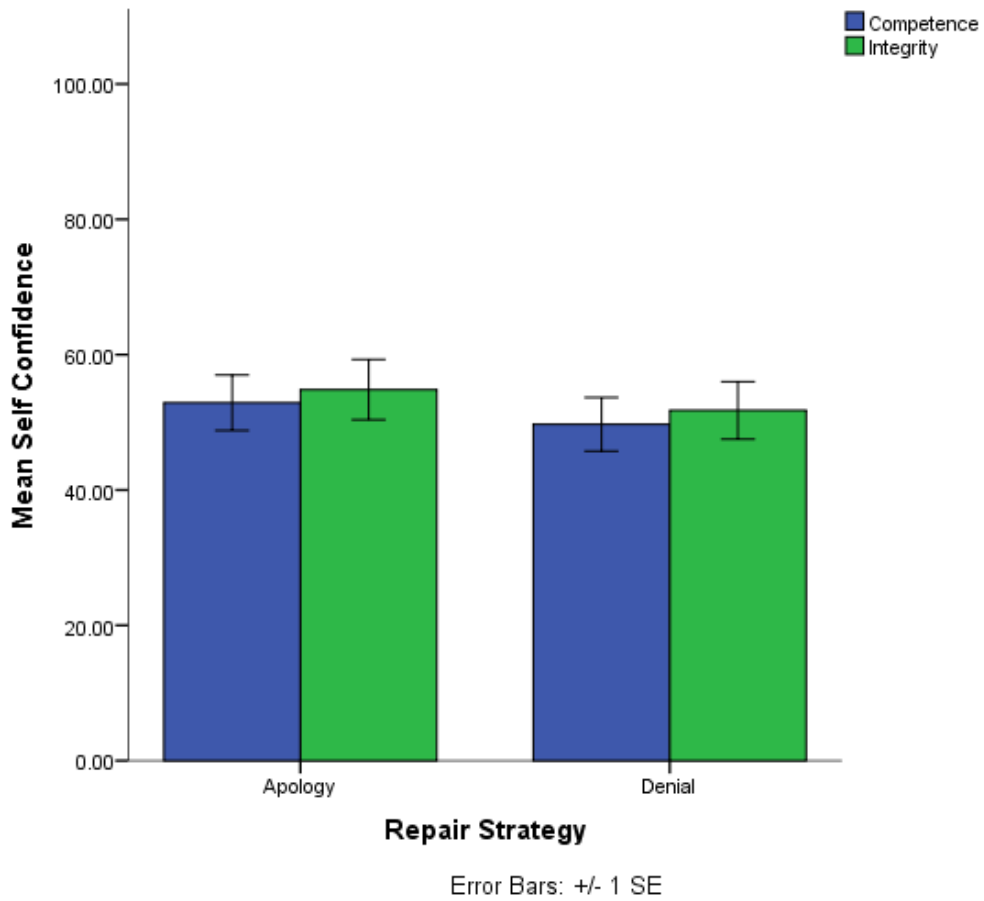


Figure 11. Mean self-confidence to perform the task between failure types, organized by repair strategies.

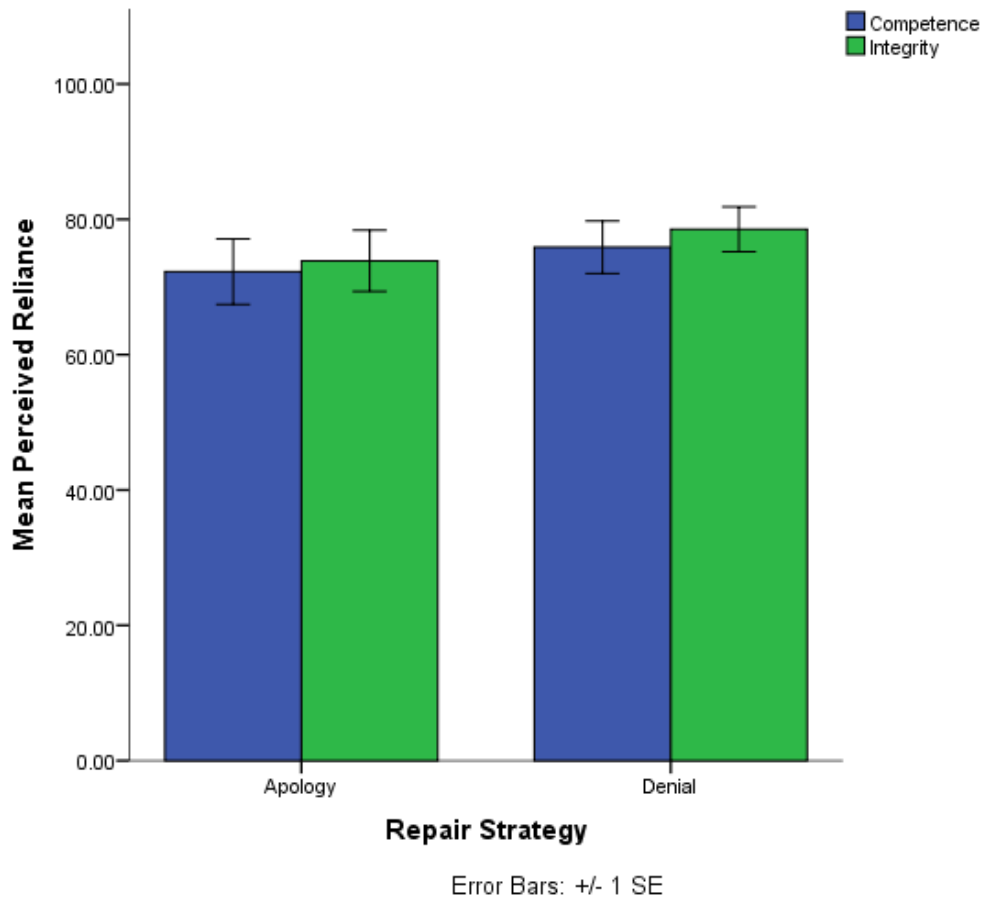


Figure 12. Mean perceived reliance on the automation between failure types, organized by repair strategies.

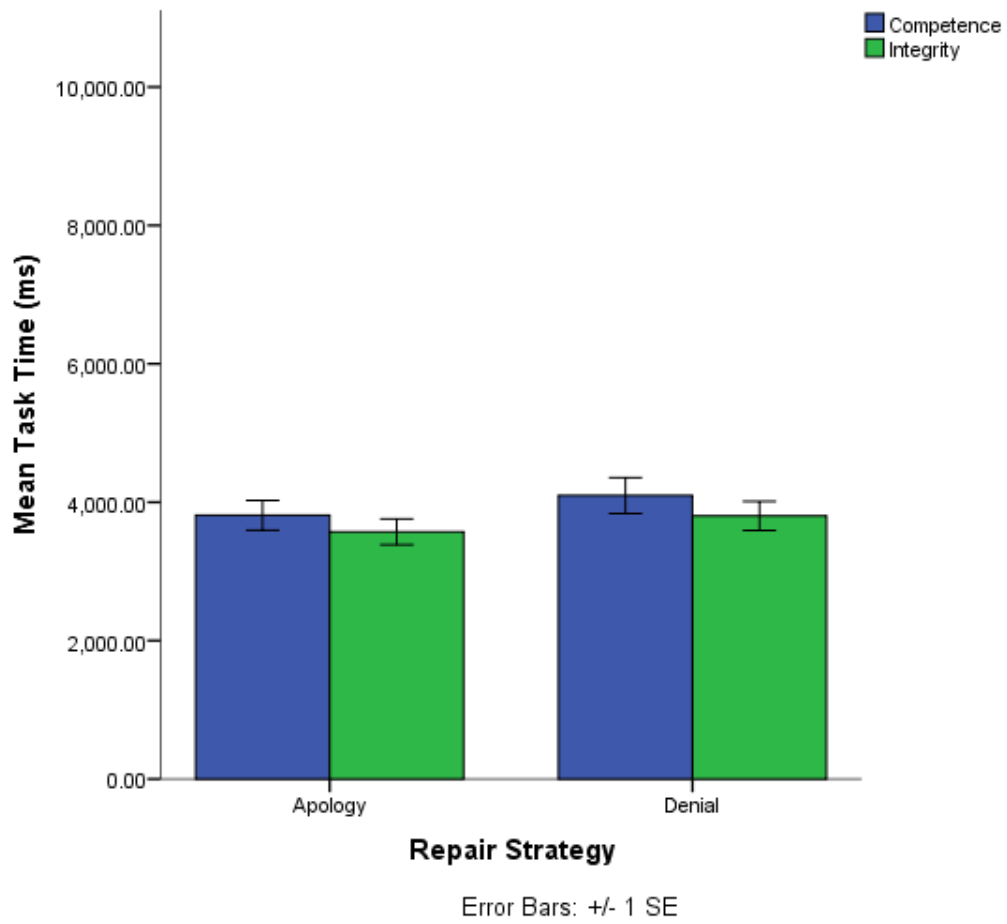


Figure 13. Mean task time between failure types, organized by repair strategies.

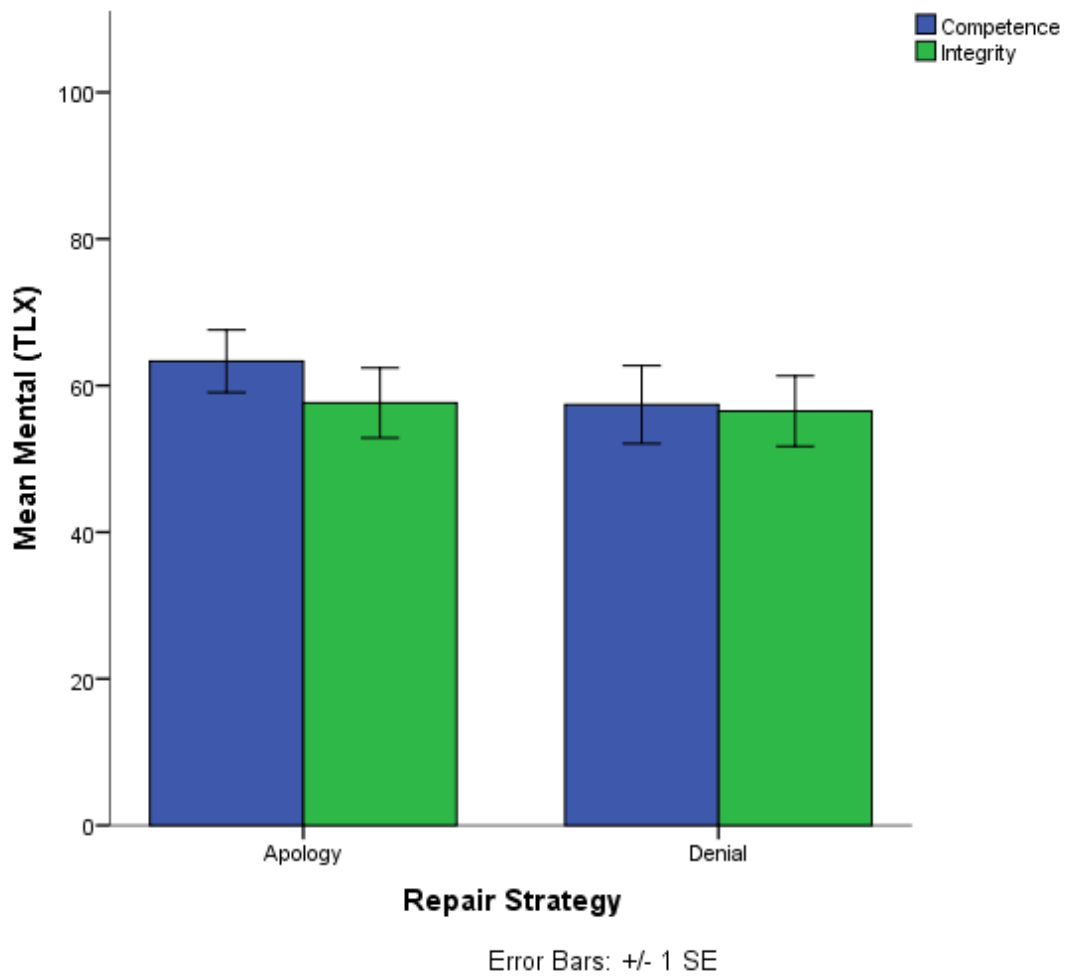


Figure 14. Mean perceived mental workload between failure types, organized by repair strategies.

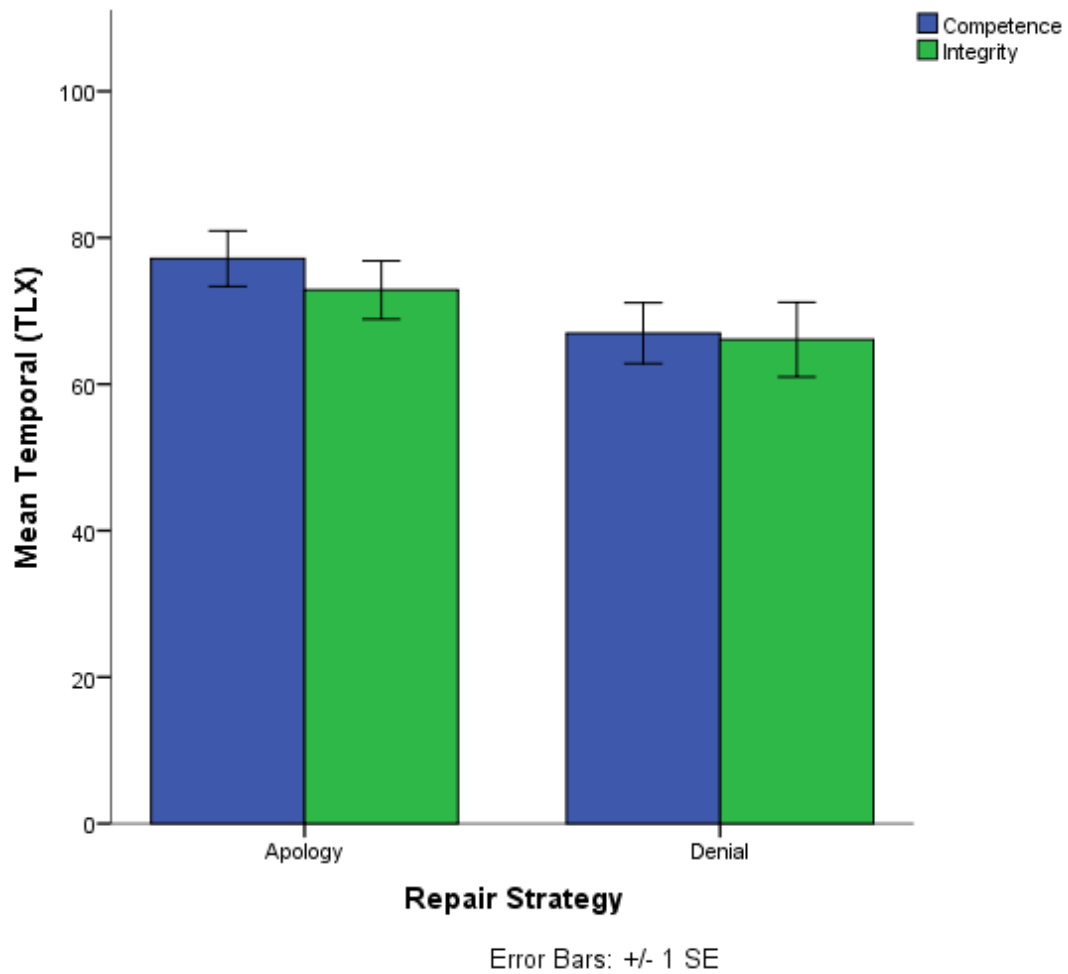


Figure 15. Mean perceived temporal workload between failure types, organized by repair strategies.

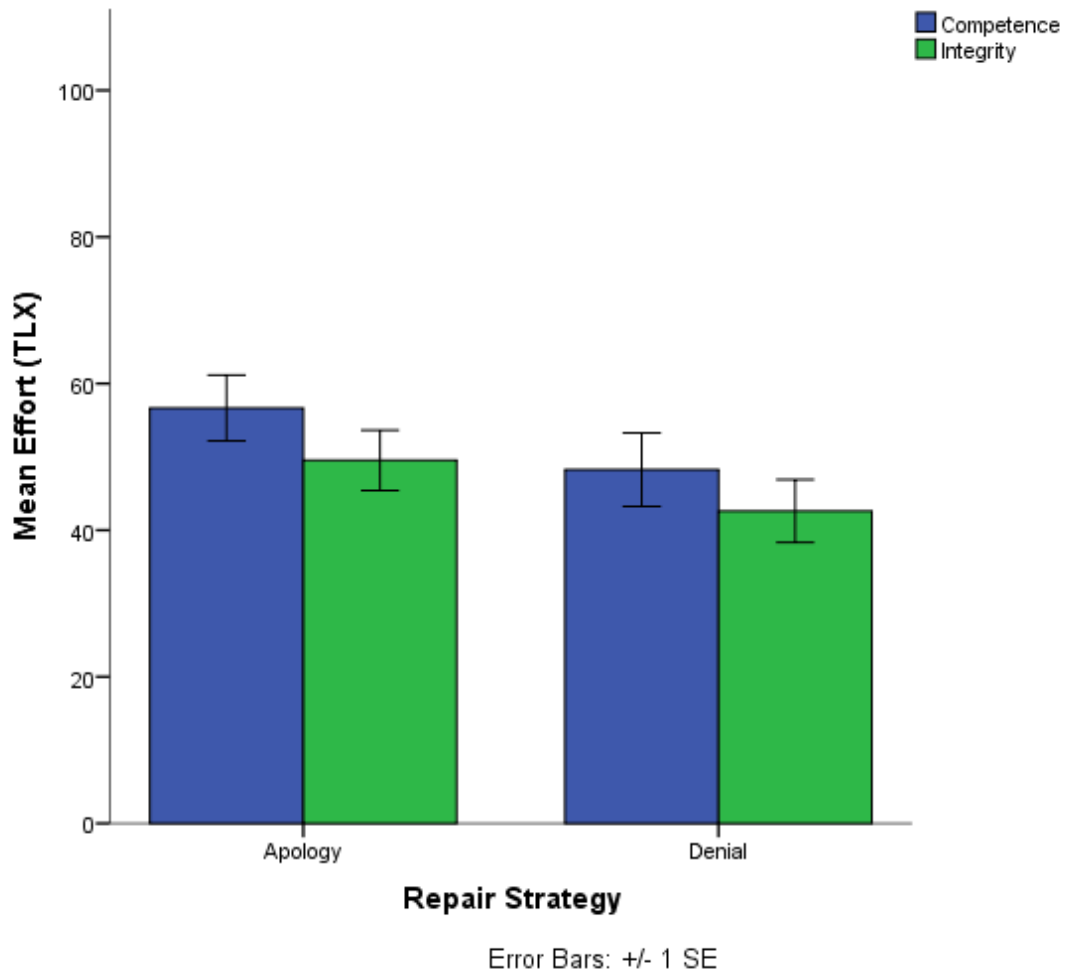


Figure 16. Mean perceived effort between failure types, organized by repair strategies.

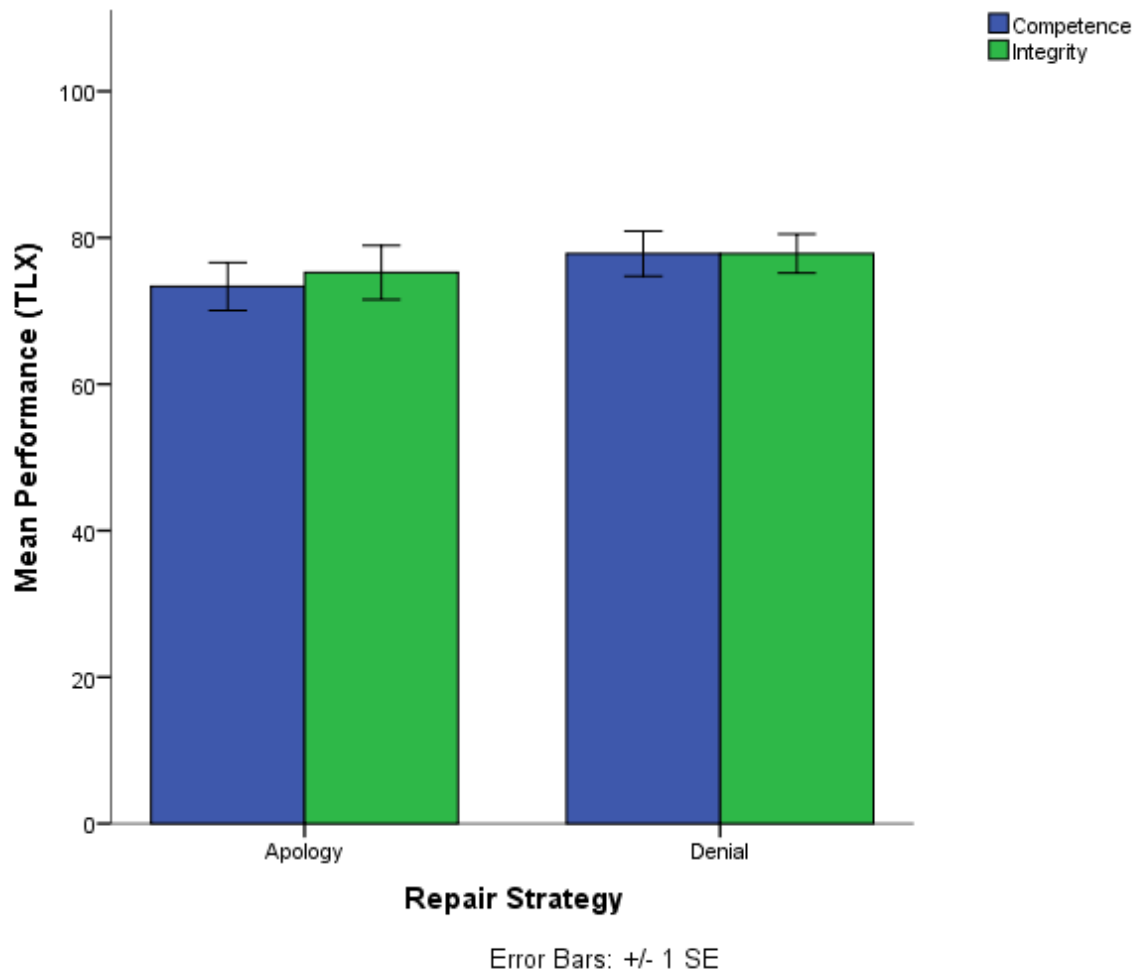


Figure 17. Mean perceived performance between failure types, organized by repair strategies.

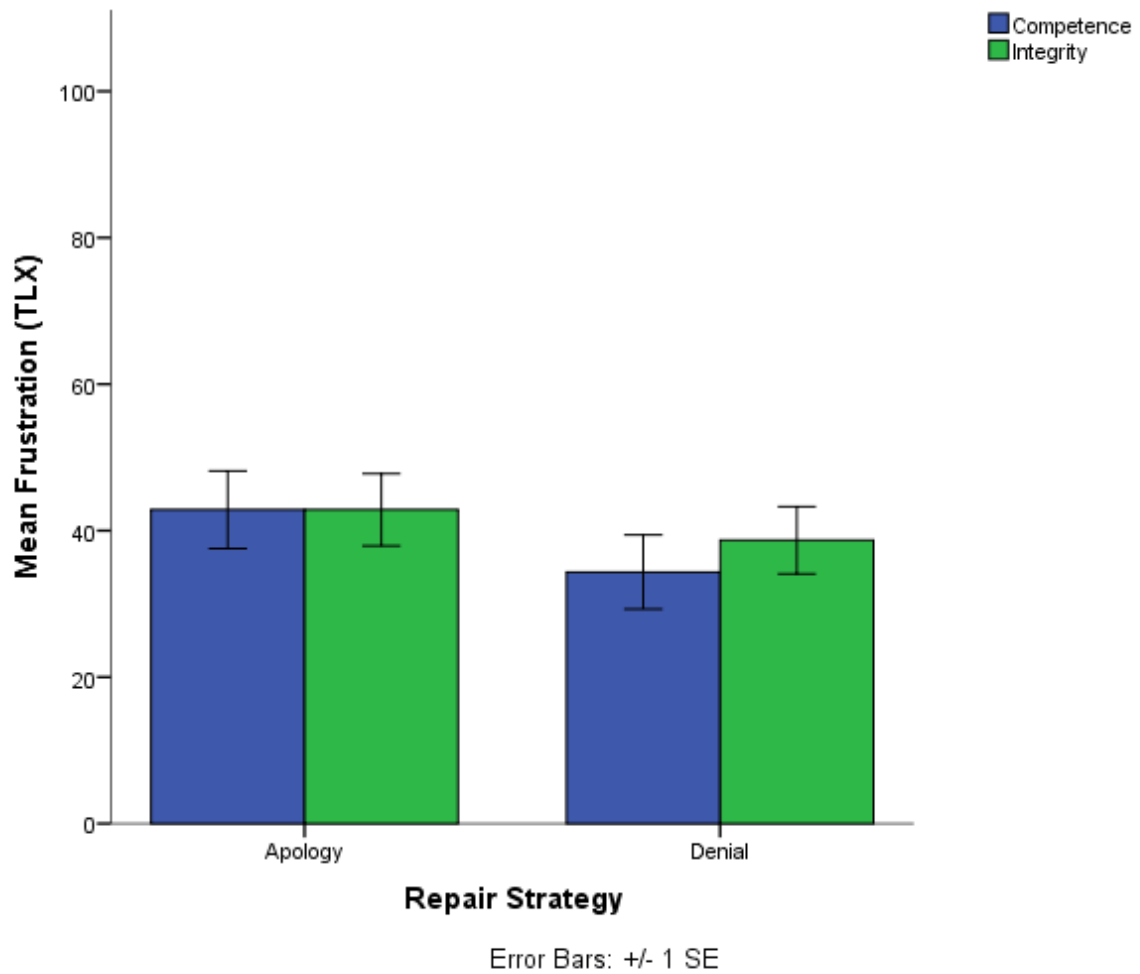


Figure 18. Mean perceived frustration between failure types, organized by repair strategies.

REFERENCES

- Akgun, M., Cagiltay, K., & Zeyrek, D. (2010). The effect of apologetic error messages and mood states on computer users' self-appraisal of performance. *Journal of Pragmatics*, 42(9), 2430-2448.
- Bottom, W. P., Gibson, K., Daniels, S. E., & Murnighan, J. K. (2002). When talk is not cheap: Substantive penance and expressions of intent in rebuilding cooperation. *Organization Science*, 13(5), 497-513.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Connelly, B. L., Crook, T. R., Combs, J. G., Ketchen Jr, D. J., & Aguinis, H. (2015). Competence-and integrity-based trust in interorganizational relationships: Which matters more?. *Journal of Management*, 0149206315596813.
- Davenport, R. B., & Bustamante, E. A. (2010, September). Effects of false-alarm vs. miss-prone automation and likelihood alarm technology on trust, reliance, and compliance in a miss-prone task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 19, pp. 1513-1517). SAGE Publications.
- De Visser, E. J., Krueger, F., McKnight, P., Scheid, S., Smith, M., Chalk, S., & Parasuraman, R. (2012, September). The world is not enough: Trust in cognitive agents. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 263-267). Sage CA: Los Angeles, CA: Sage Publications.
- De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3), 331.
- De Visser, E. J., Pak, R., & Shaw, T. H. (2018). From "automation" to "autonomy": The importance of trust repair in human-machine interaction. *Ergonomics*, (just-accepted), 1-33.
- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors*, 37(2), 381-394.
- Goffman, E. (1971). Remedial interchanges. *Relations in public: Microstudies of the public order*, 95-187.

- Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24(4), 1494-1509.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.
- Hoff, K. A., & Bashir, M. (2015). Trust in automation integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), 407-434.
- Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence-vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, 99(1), 49-65.
- Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: the effects of apology versus denial for repairing competence-versus integrity-based trust violations. *Journal of applied psychology*, 89(1), 104.
- Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International journal of human-computer studies*, 40(1), 153-184.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50-80.
- Lee, M. K., Kiesler, S., Forlizzi, J., & Srinivasa, S. P Rybski (2010). Gracefully mitigating breakdowns in robotic services. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 203-210).
- Lerch, F. J., Prietula, M. J., & Kulik, C. T. (1997, May). The Turing effect: The nature of trust in expert systems advice. In *Expertise in context* (pp. 417-448). MIT Press.
- Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. *Trust in organizations: Frontiers of theory and research*, 114, 139.
- Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2), 241-256.

- Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*, 6(1), 57-87.
- Marinaccio, K., Kohn, S., Parasuraman, R., & de Visser, E. J. (2015, June). A Framework for Rebuilding Trust in Social Automation Across Health-Care Domains. In *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care* (Vol. 4, No. 1, pp. 201-205). SAGE Publications.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, 20(3), 709-734.
- Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, 50(2), 194-210.
- Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429-460.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1), 81-103.
- Nass, C., Steuer, J., & Tauber, E. R. (1994, April). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78). ACM.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, 55(9), 1059-1072.
- Pak, R., McLaughlin, A. C., & Bass, B. (2014). A multi-level analysis of the effects of age and gender stereotypes on trust in anthropomorphic technology by younger and older adults. *Ergonomics*, 57(9), 1277-1289.
- Pak, R., McLaughlin, A. C., Leidheiser, W., & Rovira, E. (2017). The effect of individual differences in working memory in older adults on performance with different degrees of automated technology. *Ergonomics*, 60(4), 518-532.
- Parasuraman, R., & Miller, C. A. (2004). Trust and etiquette in high-criticality automated systems. *Communications of the ACM*, 47(4), 51-55.

- Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced 'complacency'. *The International Journal of Aviation Psychology*, 3(1), 1-23.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 230-253.
- Reeder, G. D., & Coovert, M. D. (1986). Revising an impression of morality. *Social Cognition*, 4(1), 1-17.
- Robinette, P., Wagner, A. R., & Howard, A. M. (2015). *The Effect of Robot Performance on Human--Robot Trust in Time--Critical Situations*.
- Roßnagel, C. (2000). Cognitive load and perspective-taking: applying the automatic-controlled distinction to verbal communication. *European journal of social psychology*, 30(3), 429-445.
- Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, 49(1), 76-87.
- Rovira, E., & Parasuraman, R. (2010). Transitioning to future air traffic management: Effects of imperfect automation on controller attention and performance. *Human factors*, 52(3), 411-425.
- Sarter, N. R., & Woods, D. D. (1993). Cognitive engineering in aerospace application: Pilot interaction with cockpit automation. (NASA Contractor Report 177617). Columbus, OH: The Ohio State University.
- Schaefer, K. E., Straub, E. R., Chen, J. Y., Putney, J., & Evans III, A. W. (2017). Communicating intent to develop shared situation awareness and engender trust in human-agent teams. *Cognitive Systems Research*, 46, 26-39.
- Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational behavior and human decision processes*, 101(1), 1-19.
- Shapiro, D. L., Sheppard, B. H., & Cheraskin, L. (1992). Business on a handshake. *Negotiation journal*, 8(4), 365-377.
- Spain, R. D., & Madhavan, P. (2009, October). The role of automation etiquette and pedigree in trust and dependence. In *Proceedings of the Human Factors and*

- Ergonomics Society Annual Meeting* (Vol. 53, No. 4, pp. 339-343). SAGE Publications.
- Tzeng, J. Y. (2004). Toward a more civilized design: studying the effects of computers that apologize. *International Journal of Human-Computer Studies*, 61(3), 319-345.
- Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113-117.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, 92(4), 548.
- Weiner, B., Amirkhan, J., Folkes, V. S., & Verette, J. A. (1987). An attributional analysis of excuse giving: Studies of a naive theory of emotion. *Journal of personality and social psychology*, 52(2), 316.
- Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3), 201-212.
- Wickens, C. D., Hellenberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44(2), 171-188.