Toward quantifying trust dynamics: How people adjust their trust after moment-to-moment interaction with automation

X. Jessie Yang

Industrial and Operations Engineering, University of Michigan

Christopher Schemanske

Industrial and Operations Engineering, University of Michigan

Christine Searle

Robotics Institute, University of Michigan

Accepted to be published in Human Factors 06/29/2021

Manuscript type: *Research Article*

Running head: *How people adjust their trust in automation*

Word count: 4600

Corresponding author: X. Jessie Yang, Industrial and Operations Engineering, University of Michigan, Ann Arbor. Email: xijyang@umich.edu

Acknowledgment: This material is based upon work supported in part by the National Science Foundation under Grant No. 2045009.

Précis: We examine how human operators adjust their trust in automation as a result of moment-to-moment interaction with automation. Results indicate that operators' trust adjustments are significantly influenced by decision-making heuristics/biases including the outcome bias and the contrast effect. Additionally, automation failures engender a larger effect on trust adjustment than successes.

Topic: Automation, Expert Systems

ABSTRACT

Objective: We examine how human operators adjust their trust in automation as a result of their moment-to-moment interaction with automation.

Background: Most existing studies measured trust by administering questionnaires at the end of an experiment. Only a limited number of studies viewed trust as a dynamic variable that can strengthen or decay over time.

Method: Seventy-five participants took part in an aided memory recognition task. In the task, participants viewed a series of images and later on performed 40 trials of the recognition task to identify a target image when it was presented with a distractor. In each trial, participants performed the initial recognition by themselves, received a recommendation from an automated decision aid, and performed the final recognition. After each trial, participants reported their trust on a visual analog scale.

Results: Outcome bias and contrast effect significantly influence human operators' trust adjustments. An automation failure leads to a larger trust decrement if the final outcome is undesirable, and a marginally larger trust decrement if the human operator succeeds the task by him-/her-self. An automation success engenders a greater trust increment if the human operator fails the task. Additionally, automation failures have a larger effect on trust adjustment than automation successes.

Conclusion: Human operators adjust their trust in automation as a result of their moment-to-moment interaction with automation. Their trust adjustments are significantly influenced by decision-making heuristics/biases.

Application: Understanding the trust adjustment process enables accurate prediction of the operators' moment-to-moment trust in automation and informs the design of trust-aware adaptive automation.

Keywords: Human-automation interaction, human-autonomy interaction, heuristics and biases, decision aid

1. INTRODUCTION

Consider the following hypothetical scenarios:

> Scenario A: Assume Mark and Brian were identical from the medical perspective. Both of them spotted blood in their stools. A clinical decision system was used to decide whether they were at risk of colon cancer and if follow-up colonoscopy examinations were necessary. The decision system decided that both patients were at very low risk of developing colon cancer and colonoscopy examinations were unnecessary. Mark, a hypochondriac fully covered by medical insurance, still requested a follow-up colonoscopy examination, which turned out to reveal cancerous polyps in his colon. A polypectomy afterward removed those polyps and saved his life. In contrast, Brian, constrained by his financial status, did not request a colonoscopy. Shortly after, unfortunately, he was diagnosed with colon cancer.

In Scenario A, the clinical decision system made wrong diagnoses for both patients. Therefore, we would expect a decrement of trust toward the system in both cases, but would the levels of trust drop be equal?

> Scenario B: Assume Amy and Marina were identical from the medical perspective. Both Amy and Marina spotted a painful lesion on their skins. Their symptoms were analyzed by a medical doctor and a clinical decision system. Based on the medical doctor's diagnoses, only Amy has developed skin cancer. However, the clinical decision system concluded that both patients had skin cancer. A tandem expert review confirmed that the decisions by the clinical system were correct.

In Scenario B, the clinical decision system made correct diagnoses for both patients. Therefore we expect an increment of trust toward the system in both cases, but would the levels of trust increment be equal?

The present study seeks to answer these questions. We begin by reviewing existing literature on trust adjustment in human-automation interaction. Next, we discuss the two types of decision-making heuristics/biases, namely outcome bias and contrast effect. We hypothesize that outcome bias and contrast effect would affect human operators' trust adjustments when interacting with automation. Although there are other types of decision-making heuristics/biases, these two types are particularly relevant to trust adjustment and are our focus in the present study.

1.1 Trust as a Dynamic Variable

Trust in automation, or more recently, trust in autonomy, has attracted substantial research attention in the past three decades. The majority of prior literature adopted a *snapshot* view of trust and typically evaluated trust through questionnaires administered at the end of an experiment. More than two dozen factors have been identified to influence one's (snapshot) trust in automation, including individual factors such as culture and age (McBride, Rogers, & Fisk, 2011; Rau, Li, & Li, 2009), system factors such as reliability and level of automation (Du, Huang, & Yang, 2020; Parasuraman, Sheridan, & Wickens, 2000; Wickens & Dixon, 2007; Wickens et al., 2009), and environmental factors such as multi-tasking requirement (Zhang & Yang, 2017). This snapshot view, however, does not fully acknowledge that trust is a dynamic variable that can change over time (Figure 1). With few exceptions, we have little understanding of how trust strengthens or decays due to moment-to-moment interactions with automation (de Visser et al., 2020; Guo & Yang, 2020; Yang, Guo, & Schemanske, To appear; Yang, Unhelkar, Li, & Shah, 2017).

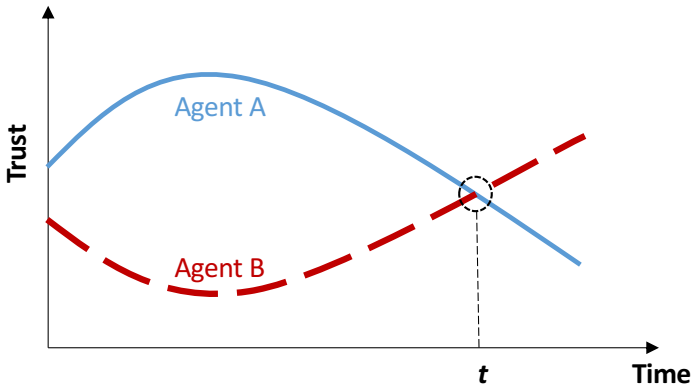


Figure 1. The static *snapshot* view of trust versus the *dynamic* view of trust. If taking a snapshot at time t, both agents have the same trust level, but their trust dynamics differ.

The limited amount of research on trust adjustment reveals two major findings. First, human operators' trust adjustments are significantly influenced by automation performance: trust increases after automation successes and decreases after automation failures (Lee & Moray, 1992; Moray, Inagaki, & Itoh, 2000; Yang et al., 2017; Yang, Wickens, & Hölttä-Otto, 2016). Second, automation failures engender a much stronger influence on trust adjustment than automation successes – Trust is difficult to build but can be lost quickly (Lee & Moray, 1992, 1994; Manzey, Reichenbach, & Onnasch, 2012).

In their seminal work, Lee and Moray (1992) employed a simulated pasteurization task, in which participants controlled two pumps and one heater, each of which could be set to automatic or manual control. Participants completed 10 training trials and 50 experimental trials over three days. Two pump faults were introduced on trials 26 and 40, at which point the pump failed to respond accurately. After each trial, participants rated their level of trust on a 10-point Likert scale. Trust plotted over the 60 trials showed a mild trust increment after interacting with a reliable pump but a large trust decrement after trials 26 and 40. Based on the results, Lee and Moray (1992) developed an autoregressive time-series model of trust. Trust at time t was modeled as a function of trust at time $t-1$, the pasteurization output, and the occurrence of pump faults.

Along the same line, Manzey et al. (2012) used the AutoCAMS multi-tasking platform (Hockey, Wastell, & Sauer, 1998) to track the human operators' trust over 20

positive interactions (automation successes) and 1 or 2 negative interactions (automation failures). Participants rated their subjective trust 5 times throughout the experiment. Results of the experiment reveal a sharp decline of trust after the automation failures. Manzey et al. (2012) concluded that human operators adjust their trust in automation based on positive and negative feedback loops. The positive loop is triggered by the experience of automation success, and the negative loop by the experience of automation failures.

Recently, Yang et al. (To appear) summarized the two major findings as two properties of trust dynamics, namely *continuity* and *negativity bias*, and identified the third property, *stabilization* (i.e., An average person's trust will stabilize over repeated interactions with the same automation.) Based on the three properties, a computational model for predicting the moment-to-moment trust was developed (Guo, Shi, & Yang, 2021; Guo & Yang, 2020; Yang et al., To appear). The computational model proposes that trust at any time, t, follows a Beta distribution. The model outperforms existing models in prediction accuracy and guarantees good model e generalizability and explainability.

1.2 Decision Making Heuristics/Biases and Trust Adjustment

When making decisions, people rarely use the normative approach. Instead, their decision-making is often subject to various types of heuristics/biases. Below we discuss three types of decision-making heuristics/biases, namely, hindsight bias, outcome bias, and contrast effect.

Hindsight bias, as described by Fischhoff (1975), is the tendency to adjust the estimates of various likelihoods of possible event outcomes in uncertain situations after the event has occurred and the outcome is known. Hindsight bias is also known as the creeping determinism (Fischhoff, 1975) and knew-it-all-along effect (Wood, 1978). It has been documented in auditory processing, labor disputes, medical diagnoses, consumer satisfaction, personnel management, sporting events, political strategy, legal proceedings, and nuclear accident analysis (Bernstein, Wilson, Pernat, & Meilleur, 2012;

Hawkins & Hastie, 1990; Roese & Vohs, 2012). The typical experimental procedure presents participants with a situation that may lead to several possible outcomes. Participants estimate the likelihood of each outcome, learn the actual outcome, and estimate the likelihood of each outcome again (Guilbault, Bryant, Brockway, & Posavac, 2004). Hindsight bias occurs when the participants rate the actual outcome as more likely in the second estimate and the other possible outcomes as less likely.

Rather than observing a change in the estimated likelihood of each outcome, outcome bias observes a change in the perceived quality of the decision made (Baron & Hershey, 1988; Henriksen & Kaplan, 2003). Outcome bias has been observed in laboratory tasks evaluating medical decisions (Baron & Hershey, 1988), gambling (Baron & Hershey, 1988; Brownback & Kuhn, 2019; Sezer, Zhang, Gino, & Bazerman, 2016), and business decisions (Gino, Moore, & Bazerman, 2009), and in real-world evaluations of soccer player performance (Kausel, Ventura, & Rodríguez, 2019). Following a similar experimental paradigm, outcome bias studies ask participants to rate the quality of a decision that influences but do not entirely determine outcomes. After that, participants learns the actual outcome and rate the decision quality again. The probabilities of each outcome are held constant and in some cases even made explicit to participants (Baron & Hershey, 1988; Brownback & Kuhn, 2019), suggesting that any differences between the pre- and post-outcome decision quality judgments can be attributed to the only new information provided – the outcome (Baron & Hershey, 1988). Outcome bias occurs when the same decision is evaluated to be of lower quality when it happens to produce bad rather than good outcome.

Another decision making heuristics/biases people may use in trust adjustment is the contrast effect. The contrast effect occurs when people's judgments are unintentionally affected by previous or simultaneous stimuli. It has been observed in judgments of shape perception (Suzuki & Cavanagh, 1998), facial recognition (Hsu & Lee, 2016), job candidate interviews (Wexley, Yukl, Kovacs, & Sanders, 1972), physical attractiveness (Kenrick & Gutierres, 1980; Thornton & Maurice, 1997), and consumer reports (Lynch, Chakravarti, & Mitra, 1991). In their study of sequential effects on

perceptions of job candidates, Wexley et al. (1972) found that up to 12% of variance in ratings of candidates could be attributed to a contrast effect from the two candidates preceding the target candidate.

1.3 The Present Study

The primary objectives of the present study are two-folded. First, we aim to provide further evidence on the effects of automation successes and failures on a person's trust adjustment. Prior studies employing a small number of automation failures showed that a person's moment-to-moment trust increases after automation successes and decreases after automation failures (Lee & Moray, 1992; Moray et al., 2000; Yang et al., 2017, 2016), and automation failures have a larger influence on trust adjustment (Lee & Moray, 1992, 1994; Manzey et al., 2012). We will replicate and further examine the two empirical findings with more frequent occurrences of automation failures.

Second, we aim to investigate the effects of outcome bias and contrast effect on trust adjustment. The two types of biases/heuristics have been observed in various judgment and decision-making tasks (Baron & Hershey, 1988; Henriksen & Kaplan, 2003; Hsu & Lee, 2016; Suzuki & Cavanagh, 1998). In the context of trust adjustment in human-automation interaction, we postulate that people's moment-to-moment trust adjustment will be affected by the final outcome of a task and will be influenced by whether a task could be completed successfully by a human operator him-/her-self manually. To our knowledge this is the first experiment to examine how outcome bias and contrast effect influence people's moment-to-moment trust adjustment. In particular, we test the following hypotheses:

H1a: Trust in automation will increase as a result of automation successes and decrease as a result of automation failures.

H1b: The magnitude of trust decrements due to automation failures will be greater than that of trust increments due to automation successes.

H2a: An automation success will lead to a larger increment of trust if the final outcome of a task is desirable.

H2b: An automation failure will lead to a larger decrement of trust, if the final outcome of a task is undesirable.

H3a: An automation success will lead to a greater increment of trust if a human operator fails the task on his or her own.

H3b: An automation failure will lead to a greater decrement of trust if a human operator succeeds the task on his or her own.

Along with the primary objectives, we were interested in exploring any potential impact of the overall automation reliability on trust adjustment. Previous research has revealed consistently a positive relationship between automation reliability and the (snapshot) trust in automation measured at the end of an experiment (de Visser & Parasuraman, 2011; Wickens & Dixon, 2007). However, little research, if any, has examined the effects of automation reliability on moment-to-moment trust adjustment. In the present study, automation reliability was set to be above 70% (i.e., 70%, 80% and 90%) based on a meta-analysis showing that 70% is the threshold above which using automation improves task performance and below which such use may harm performance (Wickens & Dixon, 2007).

2. METHOD

This research complied with the American Psychological Association code of ethics and was approved by the Institutional Review Board at the University of Michigan.

2.1 Participants

A total of 75 adults (Age: Mean = 23.4 years, SD = 4.1 years) participated. The number of participants was determined using a power analysis for the 2 (pattern) $\times$ 3 (reliability level) mixed design F test. The power analysis was completed by assuming a large effect, a α of .05, and a statistical power of .80. The required sample size is 51. All participants had normal or corrected-to-normal vision. Participants were paid a base rate of $10 dollars plus a bonus ranging from $1 to $5 depending on their performance.

Of the 75 participants, 60 performed the experiment in a face-to-face setting and the remaining 15 in a remote control setting where they controlled the experimental PC

remotely (due to the COVID-19 pandemic). We notice no systematic differences in participants' behavior or performance between the two groups except that the participants in the remote control setting took longer to complete the experiment, probably due to internet lags when controlling the experimental PC remotely.

2.2 Apparatus and Stimuli

The study employs a simulated memory recognition task adapted from Tulving (1981). Figure 2 shows the flowchart of the experimental task.

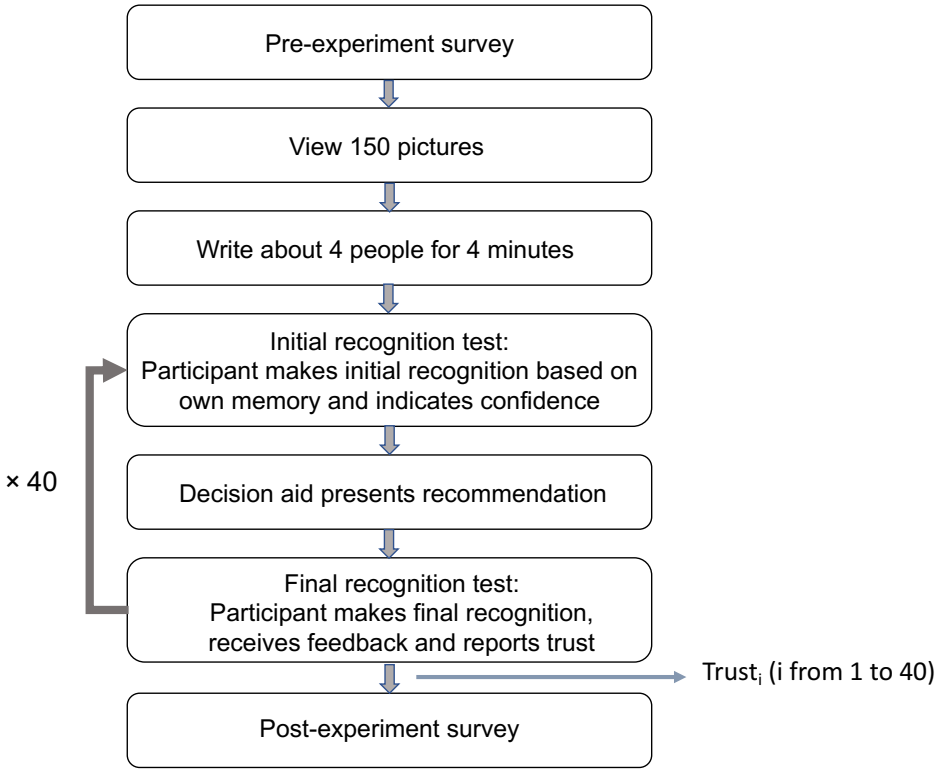


Figure 2. Flow chart of the aided memory recognition task

Before the experiment, participants fill in a demographic survey (i.e., age and gender) and a trust propensity survey gauging their propensity to trust automation. Trust propensity has been shown to influence a person's (snapshot) trust in automation after interacting with an automated system (Merritt & Ilgen, 2008). Please refer to Appendix A for the items used in the survey.

In the experiment, each participant first views a block of 150 pictures (i.e., A, B, C, ...), each for two seconds, as shown in Figure 3. Sixty of the 150 images are targets

for later recognition, and 90 are buffer images to increase cognitive load. Next, participants perform an interpolated memory task to write down as much information as they could about four famous people in four minutes. The interpolated task is used in the study of Tulving (1981) to bring the hit rate into the middle performance range. The number of famous people and the duration of the interpolated task are determined in a preliminary study (Yang et al., 2016).

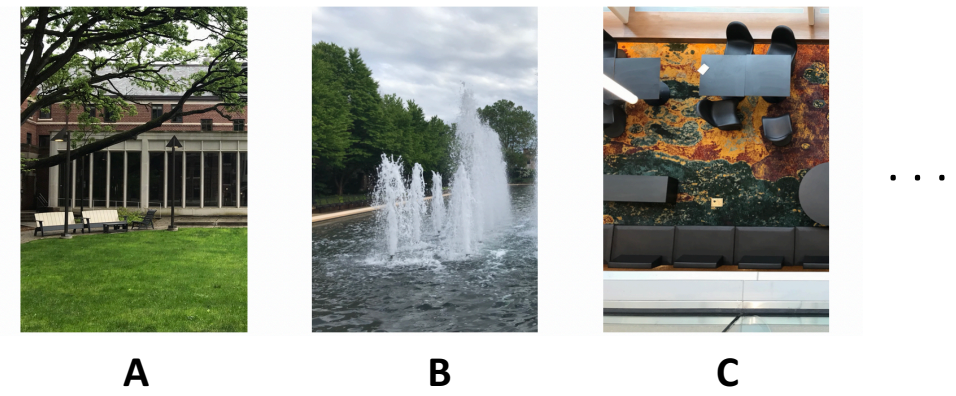


Figure 3. Illustration of target pictures A, B, C, ...

Participants are then given the recognition test that consists of 40 trials of a two alternative forced choice image recognition task (2AFC), in which participants identify a target image when it is presented with a distractor. The 40 trials of target-distractor pairs are created from the 60 target pictures (Figure 4). In each trial, a target picture could be presented with a distractor that resembles itself (e.g, Distractor A' resembles target A) or a target picture could be presented with a distractor that resembles another target picture (e.g., Distractor C' resembles target C). Therefore, 60 target pictures only generate 40 target-distractor pairs.

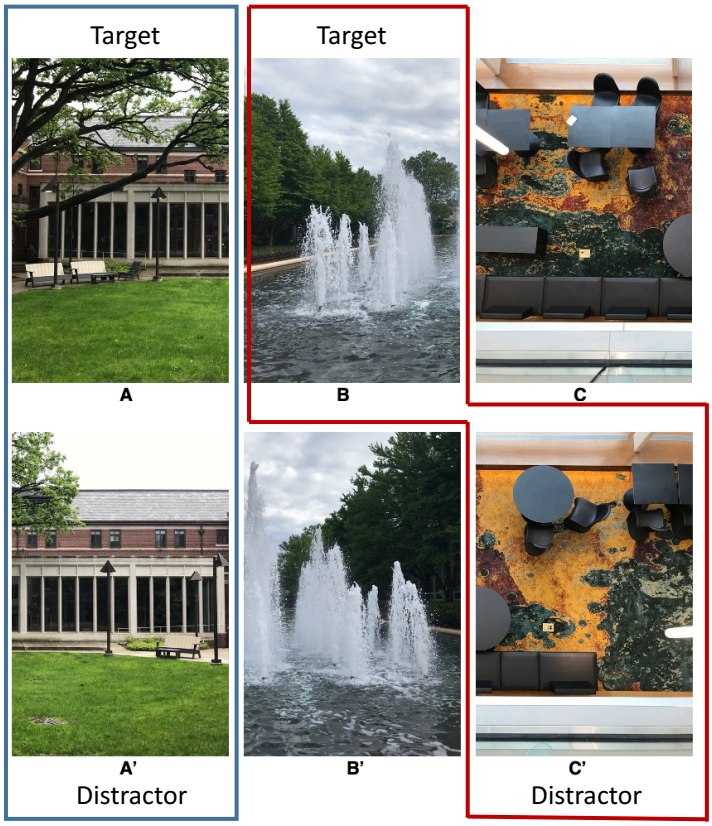


Figure 4. Illustration of target-distractor pairs

During each recognition trial, participants first selects the image they recall seeing previously entirely based on their memory by clicking on it with the cursor (Figure 5). The selected image will be highlighted in a frame. Participants then rate their confidence using a visual analog scale, with the leftmost point labeled "I'm completely guessing." and the rightmost point "I'm completely certain."

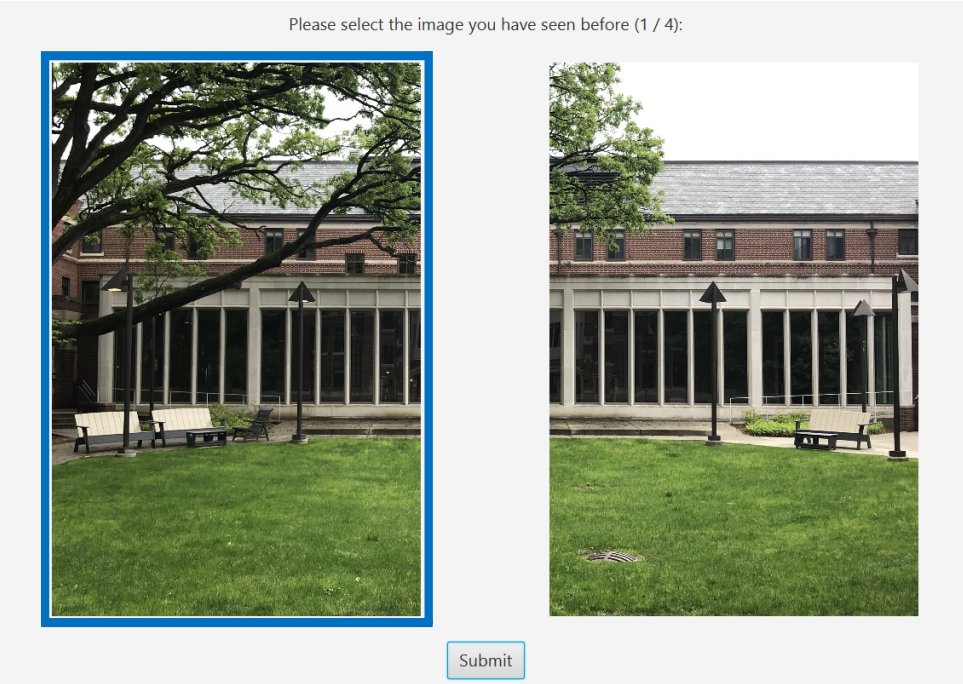


Figure 5. Illustration of the two-alternative forced-choice (2AFC) test

After that, the participants are reminded of their initial choice, and presented with the recommendation from an automated decision aid (i.e., displaying "The image recognition algorithm suggests LEFT/RIGHT" on screen). The participants are then asked to make their final recognition selection (Figure 6). Once the participants make their final recognition selection, they receive feedback on the correctness of their final choice (i.e., "Your final answer is CORRECT/WRONG" on screen).

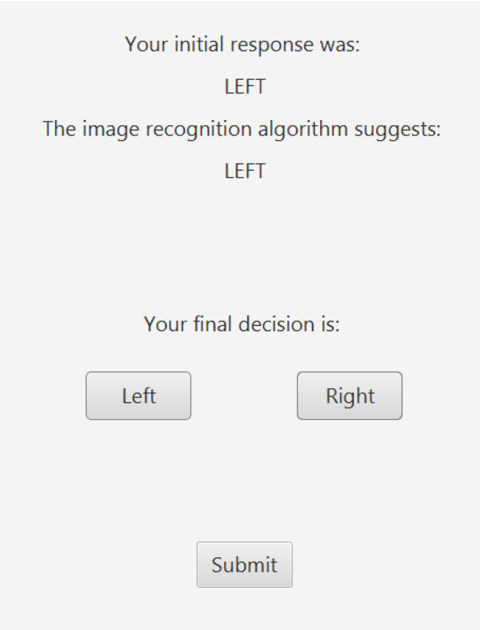


Figure 6. The automated aid's recommendation interface

After that, participants rate their trust towards the automated decision aid using a visual analog scale (Manzey et al., 2012) (Figure 7), with the leftmost point labelled "I don't trust it at all." and the rightmost point "I trust it completely." All visual analog scales are then converted to 0-100 scales for data analysis.

Figure 7. Illustration of the visual analog scale for measuring $trust_i$ after each trial

After participants complete the experiment, they fill in two post-experiment trust survey adapted from Jian, Bisantz, and Drury (n.d.) and Muir and Moray (1996). Please refer to Appendixes B and C for the items used in the two surveys.

2.3 Experimental Design

The experiment employs a mixed design with two independent variables. The between-subjects variable is automation reliability, varied in three levels: 70%, 80%, and 90%. The within-subjects variable is performance pattern. In the experiment, participants perform an initial recognition, received a recommendation from the

automated decision aid, and performed a final recognition. All the three steps could be either correct or wrong, resulting in 8 performance patterns (Table 1). For example, Pattern 3 indicates that the participants make a wrong recognition initially (i.e., 0 in binary format), but after receiving a correct recommendation from the automated aid (i.e., 1 in binary format), change the answer, and the final recognition is correct (i.e., 1 in binary format). Note that each participant does not necessarily exhibit each of the eight patterns because the performance patterns are a result of their recognition performance.

2.4 Measures

Trust propensity. At the beginning of the experiment, participants complete a survey gauging their propensity to trust automation, adapted from Merritt, Heimbaugh, LaChapell, and Deborah (2013).

Trust adjustment. After each 2AFC trial i, participants report their $trust(i)$ in the decision aid. We calculate a trust adjustment as:

$$Trust\ adjustment(i) = Trust(i) - Trust(i-1), \text{ where } i = 2, 3, ..., 40$$

Since the moment-to-moment trust is reported after each trial, only 39 trust adjustments are obtained from each participant.

Post-experiment Trust Survey. After the experiment, participants complete a 12-item trust survey (Jian et al., n.d.) and an 8-item trust survey (Muir & Moray, 1996).

2.5 Experimental Procedure

Prior to the experiment, participants provided informed consent and completed a demographic survey and the trust propensity survey. They were oriented to steps of the experiment and walked through each of the screens they would see during the experiment. After that, participants completed a practice session, wherein they viewed 12 images, performed the interpolated four well-known-people memory task, and performed four 2AFC trials. Participants then proceeded to the experimental trials, viewed 150 new images, performed the interpolated task, and completed 40 2AFC trials

with their assigned automation reliability level. The images and the well-known people used in the practice session were different from the ones in the actual experiment. At the end of the experiment, participants reported their post-experiment trust toward the automated aid using two scales (Jian et al., n.d.; Muir & Moray, 1996). Participants were told that the automated decision aid was imperfect, but they were not informed of the exact reliability level. The experiment spanned roughly half an hour, and the average time for the 40 2AFC test trials was 8 minutes and 54 seconds (SD = 1 minute and 58 seconds) for the in-person group, and 13 minutes and 38 seconds (SD = 5 minutes and 35 seconds) for the remote group.

3. RESULTS

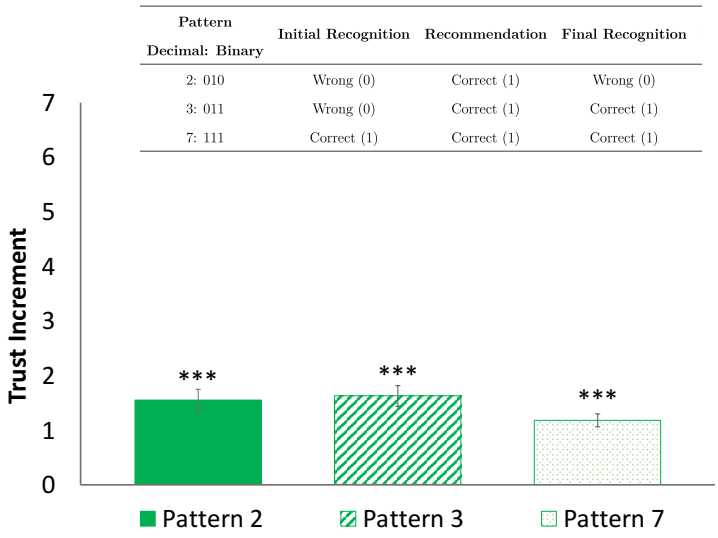
After the experiment was completed, the number of occurrences for each performance pattern was calculated. Table 1 summarizes the number of occurrences for each pattern and the corresponding trust adjustments. As the occurrence of each pattern can only be determined posteriorly, participants might not necessarily display each performance pattern. Due to the extremely low number of occurrences for patterns 1 and 6, these two patterns are discarded from data analysis. As a result, a full factorial analysis is inappropriate. Instead, we conduct a series of planned comparisons. For the planned comparisons, we first conduct Analysis of Covariance (ANCOVAs) with both automation reliability and performance pattern as independent variables, and trust propensity as the covariate. Results show that neither automation reliability (i.e., 70%, 80%, 90%) or trust propensity had significant effects nor interaction effects in all the planned comparisons. Therefore, trust propensity (as the covariate) is removed from the analysis and the data are collapsed across the levels of automation reliability. The following analysis is conducted on the collapsed dataset using one sample and paired samples t-tests. Data points outside of three standard deviations of the mean are considered outliers and excluded from the data analysis.

TABLE 1: *8 ($2 \times 2 \times 2$) possible performance patterns based on the combinations of human operator's initial recognition, the recommendation provided by the automated decision aid, and the operator's final recognition*

Performance Pattern Decimal: Binary	**Initial Recognition**	**Recommendation**	**Final Recognition**	**# of Participants**	**Trust Adjustment Mean (SD)**
0: 000	Wrong (0)	Wrong (0)	Wrong (0)	68	-4.2 (4.2)
* 1: 001	Wrong (0)	Wrong (0)	Correct (1)	1	NA
2: 010	Wrong (0)	Correct (1)	Wrong (0)	72	1.6 (1.7)
3: 011	Wrong (0)	Correct (1)	Correct (1)	71	1.9 (1.3)
4: 100	Correct (1)	Wrong (0)	Wrong (0)	63	-5.0 (4.6)
5: 101	Correct (1)	Wrong (0)	Correct (1)	63	-3.7 (4.4)
* 6 : 110	Correct (1)	Correct (1)	Wrong (0)	5	NA
7: 111	Correct (1)	Correct (1)	Correct (1)	73	1.2 (1.0)

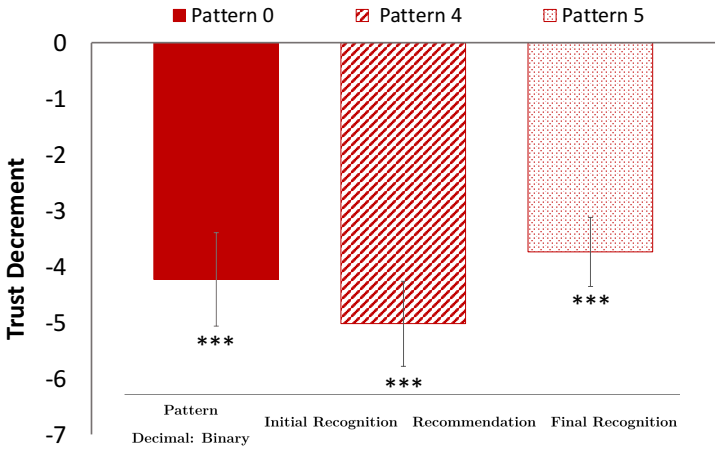
Note: Asterisks denote the exclusion of a pattern due to its low number of occurrences.

Recalling that ***H1a*** hypothesizes that an automation success would result in trust increment and an automation failure trust decrement, we compare the magnitude and direction of trust adjustment of patterns 2, 3, and 7 (correct recommendations) against zero, and of patterns 0, 4, and 5 (wrong recommendation) against zero. One sample t-tests with Bonferroni adjustments ($\alpha = 0.017$) show that correct recommendations increase trust (Pattern 2: $t(1, 71) = 7.81, p < 0.001$, Cohen's $d = 0.92$; Pattern 3: $t(1, 70) = 12.6, p < 0.001$, Cohen's $d = 1.48$; Pattern 7: $t(1, 72) = 9.94, p < 0.001$, Cohen's $d = 1.17$) (Figure 8) and wrong recommendations decrease trust (Pattern 0: $t(1, 67) = -8.29, p < 0.001$, Cohen's $d = 1.00$; Pattern 4: $t(1, 62) = -8.72, p < 0.001$, Cohen's $d = 1.10$; Pattern 5: $t(1, 62) = -6.74, p < 0.001$, Cohen's $d = 0.85$) (Figure 9).



Pattern Decimal: Binary	Initial Recognition	Recommendation	Final Recognition
2: 010	Wrong (0)	Correct (1)	Wrong (0)
3: 011	Wrong (0)	Correct (1)	Correct (1)
7: 111	Correct (1)	Correct (1)	Correct (1)

Figure 8. Mean and Standard Error(SE) values of ***trust increment*** in patterns 2, 3 and 7 ($^{***}p < .001$).

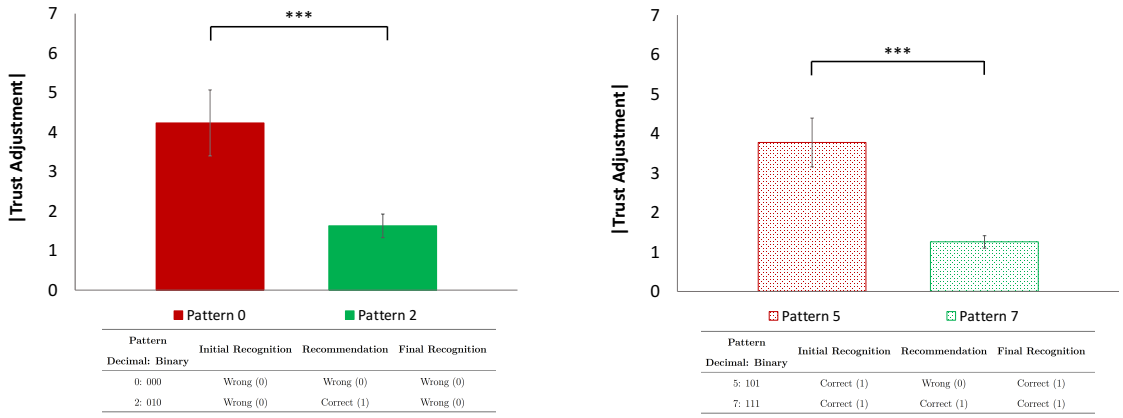


Pattern Decimal: Binary	Initial Recognition	Recommendation	Final Recognition
0: 000	Wrong (0)	Wrong (0)	Wrong (0)
4: 100	Correct (1)	Wrong (0)	Wrong (0)
5: 101	Correct (1)	Wrong (0)	Correct (1)

Figure 9. Mean and Standard Error (SE) values of ***trust decrement*** in patterns 0, 4 and 5 ($^{***}p < .001$).

H1b hypothesizes that the magnitude of trust decrements would be larger than that of trust increments. We conduct two paired samples t-tests with Bonferroni adjustments ($\alpha = 0.025$): the *magnitude* of pattern 0 (wrong initial recognition-wrong recommendation-wrong final recognition) versus the magnitude of pattern 2 (wrong initial recognition-correct recommendation-wrong final recognition), and the magnitude of pattern 5 (correct initial recognition-wrong recommendation-correct final recognition)

versus the magnitude of pattern 7 (correct initial recognition-correct recommendation-correct final recognition). The only difference between each pair is the correctness of the automation's recommendation. Results reveal higher magnitude of trust decrements for wrong recommendations in pattern 0 compared to correct recommendations in pattern 2 ($t(1, 65) = 5.34, p < 0.001$, Cohen's $d = 0.66$), and for wrong recommendations in pattern 5 compared to correct recommendations in pattern 7 ($t(1, 60) = 4.81, p < 0.001$, Cohen's $d = 0.62$) (Figure 10)



Pattern Decimal: Binary	Initial Recognition	Recommendation	Final Recognition
0: 000	Wrong (0)	Wrong (0)	Wrong (0)
2: 010	Wrong (0)	Correct (1)	Wrong (0)

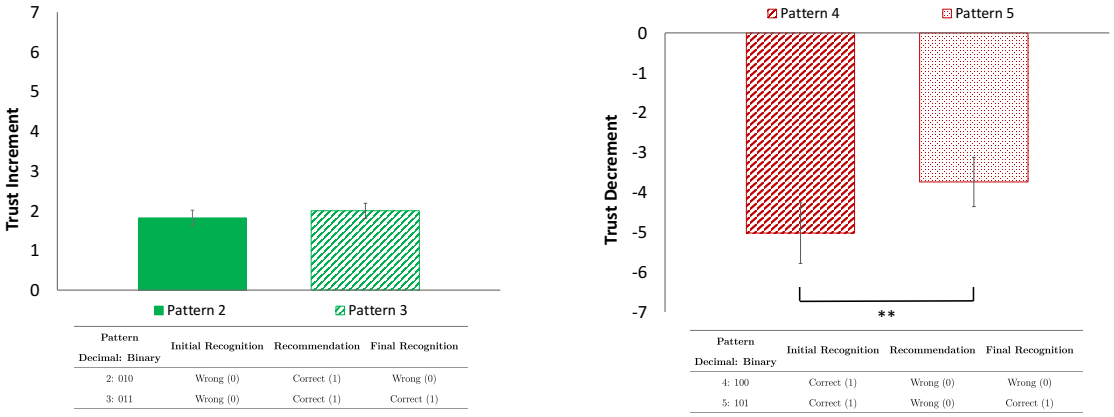
Pattern Decimal: Binary	Initial Recognition	Recommendation	Final Recognition
5: 101	Correct (1)	Wrong (0)	Correct (1)
7: 111	Correct (1)	Correct (1)	Correct (1)

(a) (b)

Figure 10. Comparing the ***magnitude*** of trust adjustment between (a) patterns 0 and 2 and between (b) patterns 5 and 7 ($***p < .001$).

H2a hypothesizes that an automation success would lead to a larger trust increment if the final outcome is good. We compare pattern 2 (wrong initial recognition - correct recommendation - wrong final recognition) versus pattern 3 (wrong initial recognition - correct recommendation - correct final recognition), where the only difference is the final recognition performance. The t-test shows that the difference is not significant ($t(1, 67) = -1.08, p = .29$) (Figure 11(a)).

H2b hypothesizes that an automation failure would lead to a larger trust decrement if the final outcome is undesirable. The comparison between pattern 4 (correct initial recognition - wrong recommendation - wrong final recognition) and pattern 5 (correct initial recognition - wrong recommendation - correct final recognition) reveals a significantly larger decrement for pattern 4 (wrong final recognition) ($t(1, 52) = -2.63, p = .01$, Cohen's $d = 0.36$) (Figure 11(b)).



Pattern Decimal: Binary	Initial Recognition	Recommendation	Final Recognition
2: 010	Wrong (0)	Correct (1)	Wrong (0)
3: 011	Wrong (0)	Correct (1)	Correct (1)

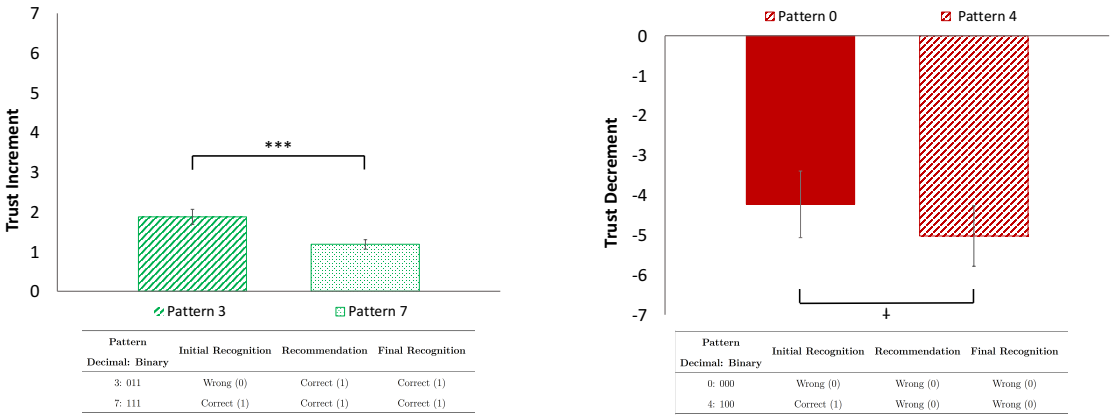
Pattern Decimal: Binary	Initial Recognition	Recommendation	Final Recognition
4: 100	Correct (1)	Wrong (0)	Wrong (0)
5: 101	Correct (1)	Wrong (0)	Correct (1)

(a) (b)

Figure 11. (a) Comparing ***trust increment*** between patterns 2 and 3. (b) Comparing ***trust decrement*** between patterns 4 and 5 ($^{**}p < .01$).

We hypothesize in ***H3a*** that an automation success would produce a greater trust increment if the human operator fails the task. We compare pattern 3 (wrong initial recognition - correct recommendation - correct final recognition) versus pattern 7 (correct initial recognition - correct recommendation - correct final recognition). In the two patterns, the automated aid provides correct recommendations, and the only difference is the human operators' initial recognition. A paired-samples t-test reveals a significantly larger trust increment in pattern 3 compared to pattern 7 ($t(1, 69) = 4.40, p < 0.001$, Cohen's $d = 0.53$) (Figure 12(a)).

In ***H3b***, we hypothesize that an automation failure would lead to a greater trust decrement if the human operator is capable of performing the task. We compare pattern 0 (wrong initial recognition - wrong recommendation - wrong final recognition) versus pattern 4 (correct initial recognition - wrong recommendation - wrong final recognition). The analysis show a marginally significant difference between the two patterns ($t(1, 57) = 1.83, p = .07$, Cohen's $d = 0.24$) (Figure 12(b)).



(a) (b)

Figure 12. (a) Comparing ***trust increment*** between patterns 3 and 7 ($^{***}p < .001$). (b) Comparing ***trust decrement*** between patterns 0 and 4 ($^{\dagger}p < .1$).

After examining the moment-to-moment trust adjustments, we further explore how the accumulation of moment-to-moment adjustments contribute to the participants' trust at the end of the experiment. Table 2 summarizes participants' trust propensity, trust after the 40th trial (i.e., after the entire interaction experience with automation), and post-experiment trust at each automation reliability level.

TABLE 2: *Mean and standard deviation values of participants' trust propensity, Trust*(40)*, and post-experiment trust*

Reliability (%)	**Trust Propensity**	**Trust after 40th Trial** *Trust*(40)	**Post-experiment Trust Scale of Jian et al. (n.d.)**	**Post-experiment Trust Scale of Muir and Moray (1996)**
70	72.6 (14.8)	56.6 (25.7)	50.6 (18.0)	53.3 (16.9)
80	69.4 (10.4)	73.8 (17.0)	54.1 (12.1)	55.9 (12.4)
90	69.4 (14.4)	81.6 (15.0)	64.9 (16.7)	72.1 (13.5)

We construct four linear regression models to explore how automation reliability affect trust propensity, and how reliability and trust propensity jointly influence participants' trust after they interact with the automation. The results show that automation reliability is not a significant predictor of trust propensity ($F(1, 73) = .73, p = .40$). Both automation reliability ($\beta = 1.32, t(72) = 4.91, p < .001$) and trust propensity ($\beta = 0.45, t(72) = 2.72, p < .01$) significantly predict participants' trust after the 40th trial. Automation reliability ($\beta = 0.77, t(72) = 3.60, p < .001$) and

trust propensity ($\beta = 0.37, t(72) = 2.74, p < .01$) significantly predict participants' post-experiment trust measured by the scale of Jian et al. (n.d.). Automation reliability ($\beta = 1.00, t(72) = 5.10, p < .001$) and trust propensity ($\beta = 0.39, t(72) = 3.16, p < .01$) significantly predict participants' post-experiment trust measured by the scale of Muir and Moray (1996).

4. DISCUSSION

Consistent with previous work of Lee and Moray (1992), Moray et al. (2000) and Yang et al. (2017, 2016), we also find that trust in automation increases as a result of automation successes and decreases as a result of automation failures (supporting ***H1a***). Our finding that the magnitude of trust decrements is larger than that of trust increments (supporting ***H1b***) is in line with previous studies in which the number of automation successes significantly exceeded that of automation failures (Lee & Moray, 1992; Manzey et al., 2012). In Lee and Moray (1992), participants experienced two automation failures among 50 experimental trials. In Manzey et al. (2012), participants had one or two automation failure(s) among 20 automation successes. Both studies consistently showed that the strength of automation failures on trust adjustment is considerably stronger than automation successes. Our study show that the stronger effect of automation failures still holds when there is a considerable number of automation failures (i.e., 12 failures among 40 trials).

With respect to the outcome bias, we find supporting evidence for ***H2b*** that an automation failure leads to a larger trust decrement if the final outcome (i.e., the final recognition performance) is undesirable. In line with previous research in medical, gambling and business decisions (Baron & Hershey, 1988; Brownback & Kuhn, 2019; Gino et al., 2009; Sezer et al., 2016), human operators demonstrate outcome biases in trust adjustments when the automation is wrong: an automation error is forgiven to some extent if the error does not lead to detrimental outcomes. This finding is disconcerting because the final outcome of a task can be due to a combination of factors. The automation influences but do not entirely determine the outcome of the

task. This finding should be explored further in follow-up experiments with more cognitively demanding tasks, such as medical decision-making, as the experimental task used in the present study is fairly simple.

However, results of the present study does not support ***H2a*** that a correct recommendation would lead to a larger trust increment if the final outcome (i.e., the final recognition performance) is good. This non-significant result could have been due to the self-serving bias (Duval & Silvia, 2002; Miller & Ross, 1975; Weiner, 1985). The self-serving bias is any cognitive or perceptual process that is distorted by the need to maintain and enhance self-esteem. It is particularly evident when individuals attribute the cause of outcomes. When explaining positive outcomes, their attributions emphasize the causal impact of internal, dispositional causes, but when identifying the causes of negative events, they stress external, situational factors. If we take a close look at the comparisons between pattern 2 (wrong initial recognition - correct recommendation - wrong final recognition) and pattern 3 (wrong initial recognition - correct recommendation - correct final recognition), the contrast between the two patterns suggest that the final correct outcome in Pattern 3 is largely due to the correct automation recommendation, and the final wrong outcome in Pattern 2 is largely due to the human operators' wrong initial recognition. According to the self-serving bias, the human operator would likely distort the trust adjustment process to maintain self-esteem, by appreciating the correct recommendation less than they should have done in Pattern 3.

Results of the present study also provide support for ***H3a*** that an automation success produces a greater trust increment if the human operator fails the task and marginal support for ***H3b*** that an automation failure produces a greater trust decrement if the human operator succeeds the task. The experimental paradigm used in the present study allows direct assessment of human operators' ability (i.e., manual performance without the help of automation aids). Instead of assessing the participants' ability directly, prior literature often evaluated participants' self-confidence in performing a task manually and considered operators' self-confidence and their trust in

automation two independent constructs (i.e., Two constructs have no association). For example, de Vries, Midden, and Bouwhuis (2003) and Lee and Moray (1994) found that trust and self-confidence predict participants' automation dependence behaviors – human operators use automation when trust exceeds self-confidence and use manual control when self-confidence exceeds trust. Our findings reveal that human operators' ability and their trust adjustment are not independent of each other. Because of the strong association between self-confidence and ability (Wixted & Wells, 2017), people's self-confidence and trust are probably not independent either, and therefore should not be viewed as independent constructs.

Viewing the results on ***H1***–***H3*** holistically, our findings suggest that human operators are rational only to a certain extent when adjusting trust in automation. They are rational in the sense that they increase trust in automation after automation successes and decrease trust upon automation failures. On the contrary, they are irrational in the sense that their trust adjustments are significantly influenced by decision-making heuristics/biases.

Along with the primary hypotheses, we are interested in exploring any potential impact of the overall automation reliability on trust adjustment. We find no significant effect on any of the one-way or pairwise t-tests (i.e., automation reliability does not significantly influence trust adjustment). On the contrary, we find that automation reliability significantly predicts the (snapshot) trust in automation measured at the end of an experiment, which is in line with previous research (de Visser & Parasuraman, 2011; Du et al., 2020; Wickens & Dixon, 2007). Our findings suggest that effect of automation reliability on the (snapshot) trust at the end of an experiment is due to the accumulation of the moment-to-moment trust adjustments over time. A less reliable automation produces more automation failures, leading to a lower (snapshot) trust at the end.

5. CONCLUSION

In contrast to the snapshot view of trust, this study considers trust a dynamic variable and examines how human operators adjust their trust in automation as a result of moment-to-moment interaction with automation. Understanding the trust adjustment process enables accurate prediction of an operator's moment-to-moment trust in automation, which can be used to design trust-aware adaptive automation. To the best of our knowledge, it is the first to show that operators' trust adjustments are subject to decision-making heuristics/biases. It also provides further empirical evidence that automation failures have a greater impact on trust adjustments than automation successes.

Moreover, we present a novel experimental paradigm that can be used to examine human operators' trust dynamics. The paradigm allows us to track participant's moment-to-moment trust over time. In addition, it distinguishes participants' ability to perform a task manually, the automated decision aid's performance, and the final performance. By eliciting human operators' answers pre- and post-automation recommendation, researchers could take a deep dive into how trust dynamics can be influenced by the interplay between participants' performance without the aid of automation, automation performance, and performance with the aid of automation.

We note the following limitations. Similar to a few previous studies (Manzey et al., 2012; Yang et al., 2017), we used a one-item scale. It is possible, however, that one item will fail to capture all of the sub dimensions of trust compared to the use of multi-dimension scales such as the 12-item trust scale in Jian et al. (n.d.). Further research should investigate the possibility of a succinct multi-scale trust scale that can be used in querying trust over repeated interactions with automation. Second, the occurrence of patterns 1 and 6 was excluded from data analysis as their occurrences were rare. Further research could be conducted to purposely induce the occurrences of these two patterns.

Key points

- Human operators adjust their trust in automation due to moment-to-moment interaction with automation. The trust adjustment process is moderated by decision-making heuristics/biases including outcome bias and contrast effect.
- An automation failure is forgiven to a certain extent if the failure does not harm the final task outcome.
- An automation success engenders a larger trust increment if the human operator fails the task by him-/her-self. An automation failure leads to a marginally larger trust decrement if the human operator succeed the task.
- The stronger effect of automation failures on trust adjustment still holds when the occurrence of automation failures is up to 30% (i.e., Automation reliability is 70%).

Appendix A

Trust Propensity Survey (adapted from Merritt, Heimbaugh, LaChapell & Deborah, 2013)

1	I usually trust machines/automated technologies until there is a reason not to.
2	For the most part, I distrust machines/automated technologies.
3	In general, I would rely on an automated machine/technology to assist me.
4	My tendency to trust machines/automated technologies is high.
5	It is easy for me to trust machines/automated technologies to do their job.
6	I am likely to trust a machine/automated technology even when I have little knowledge about it.

Appendix B

Post-experiment Trust Survey (adopted from Jian, Bisantz & Drury, 2000)

1	The automated decision aid is deceptive.
2	The automated decision aid behaves in an underhanded manner.
3	I am suspicious of the automated decision aid's intents, actions, or outputs.
4	I am wary of the automated decision aid.
5	The automated decision aid's actions will have a harmful or injurious outcome.
6	I am confident in the automated decision aid.
7	he automated decision aid provides security.
8	The automated decision aid has integrity.
9	The automated decision aid is dependable.
10	The automated decision aid is reliable.
11	I can trust the automated decision aid.
12	I am familiar with the automated decision aid.

Appendix C

Post-experiment Trust Survey (adopted from Muir and Moray, 1996)

1	To what extent does the automated decision aid perform its function properly?
2	To what extent can the automated decision aid's behavior be predicted from moment to moment?
3	To what extent can you count on the automated decision aid to do its job?
4	To what extent does the automated decision aid perform the task it was designed to do in the system?
5	To what extent does the automated decision aid respond similarly to similar circumstances at different points in time?
6	My degree of faith that the automated decision aid will be able to cope with other system states in the future:
7	My degree of trust in the automated decision aid to respond accurately:
8	My overall degree of trust in the automated decision aid:

Note: One item,"My degree of trust in the automated decision aid's display" from the original survey was not included.

References

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569. doi: https://doi.org/10.1037/0022-3514.54.4.569

Bernstein, D. M., Wilson, A. M., Pernat, N. L., & Meilleur, L. R. (2012). Auditory hindsight bias. *Psychonomic Bulletin & Review*, *19*(4), 588–593. doi: https://doi.org/10.3758/s13423-012-0268-0

Brownback, A., & Kuhn, M. A. (2019). Understanding outcome bias. *Games and Economic Behavior*, *117*, 342–360. doi: https://doi.org/10.1016/j.geb.2019.07.003

de Visser, E. J., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming. *Journal of Cognitive Engineering and Decision Making*, *5*(2), 209–231. doi: https://doi.org/10.1177/1555343411410160

de Visser, E. J., Peeters, M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerincx, M. A. (2020). Towards a theory of longitudinal trust calibration in human–robot teams. *International Journal of Social Robotics*, *12*(2), 459–478. doi: https://doi.org/10.1007/s12369-019-00596-x

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies*, *58*(6), 719–735. doi: https://doi.org/10.1016/S1071-5819(03)00039-9

Du, N., Huang, K. Y., & Yang, X. J. (2020). Not all information is equal: Effects of disclosing different types of likelihood information on trust, compliance and reliance, and task performance in human-automation teaming. *Human Factors*, *62*(6), 987-1001. doi: https://doi.org/10.1177/0018720819862916

Duval, T. S., & Silvia, P. J. (2002). Self-awareness, probability of improvement, and the self-serving bias. *Journal of Personality and Social Psychology*, *82*(1), 49–61. doi: https://10.1037/0022-3514.82.1.49

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome

knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288. doi: https://doi.org/10.1037/0096-1523.1.3.288

Gino, F., Moore, D. A., & Bazerman, M. H. (2009). See no evil: When we overlook other people's unethical behavior. *Social decision making: Social Dilemmas, Social Values, and Ethical Judgments*, 241–263. doi: https://doi.org/10.1037/0096-1523.1.3.288

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, *26*(2-3), 103–117. doi: https://doi.org/10.1207/s15324834basp2602&3_1

Guo, Y., Shi, C., & Yang, X. J. (2021). Reverse Psychology in Trust-Aware Human-Robot Interaction. *IEEE Robotics and Automation Letters*, *6*(3), 4851–4858. doi: https://doi.org/10.1109/LRA.2021.3067626

Guo, Y., & Yang, X. J. (2020). Modeling and predicting trust dynamics in human–robot teaming: A bayesian inference approach. *International Journal of Social Robotics*. doi: https://doi.org/10.1007/s12369-020-00703-3

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311. doi: https://doi.org/10.1037/0033-2909.107.3.311

Henriksen, K., & Kaplan, H. (2003). Hindsight bias, outcome knowledge and adaptive learning. *BMJ Quality & Safety*, *12*(suppl 2), ii46—-ii50. doi: https://doi.org/10.1136/qhc.12.suppl_2.ii46

Hockey, G. R. J., Wastell, D. G., & Sauer, J. (1998). Effects of sleep deprivation and user interface on complex performance: A multilevel analysis of compensatory control. *Human Factors*, *40*(2), 233–253. doi: https://doi.org/10.1518/001872098779480479

Hsu, S.-M., & Lee, J.-S. (2016). Relative judgment in facial identity perception as revealed by sequential effects. *Attention, Perception, & Psychophysics*, *78*(1), 264–277. doi: https://doi.org/10.3758/s13414-015-0979-1

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (n.d.). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. doi: https://doi.org10.1207/S15327566IJCE0401_04

Kausel, E. E., Ventura, S., & Rodríguez, A. (2019). Outcome bias in subjective ratings of performance: Evidence from the (football) field. *Journal of Economic Psychology*, *75*, 102132. doi: https://doi.org/10.1016/j.joep.2018.12.006

Kenrick, D. T., & Gutierres, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology*, *38*(1), 131. doi: https://doi.org/10.1037/0022-3514.38.1.131

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270. doi: https://doi.org/10.1080/00140139208967392

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*(1), 153–184. doi: https://doi.org/10.1006/ijhc.1994.1007

Lynch, J. G., Jr., Chakravarti, D., & Mitra, A. (1991). Contrast effects in consumer judgments: Changes in mental representations or in the anchoring of rating scales? *Journal of Consumer Research*, *18*(3), 284–297. doi: https://doi.org/10.1086/209260

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.

McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011). Understanding the effect of workload on automation use for younger and older adults. *Human Factors*, *53*(6), 672–686. doi: https://doi.org/10.1177/0018720811421909

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Deborah, L. (2013). I trust it, but i don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors*, *55*(3), 520-534. doi:

https://doi.org/10.1177/0018720812465081

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. doi: https://doi.org/10.1518/001872008X288574

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, *82*(2), 213–225. doi: https://doi.org/10.1037/h0076486

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology Applied*, *6*(1), 44–58.

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. doi: https://doi.org/10.1080/00140139608964474

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. doi: https://doi.org/10.1109/3468.844354

Rau, P. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior*, *25*(2), 587–595. doi: https://doi.org/10.1016/j.chb.2008.12.025

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426. doi: https://doi.org/10.1177/1745691612454303

Sezer, O., Zhang, T., Gino, F., & Bazerman, M. H. (2016). Overcoming the outcome bias: Making intentions matter. *Organizational Behavior and Human Decision Processes*, *137*, 13–26. doi: https://doi.org/10.1016/j.obhdp.2016.07.001

Suzuki, S., & Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(5), 1315. doi: https://doi.org/10.1037/0096-1523.24.5.1315

Thornton, B., & Maurice, J. (1997). Physique contrast effect: Adverse impact of

idealized body images for women. *Sex Roles*, *37*(5-6), 433–439. doi: https://doi.org/10.1023/A:1025609624848

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior*, *20*, 479-496.

Weiner, B. (1985). An Attributional Theory of Achievement Motivation and Emotion. *Psychological Review*, *92*(4), 548–573. doi: https://doi.org/10.1037/0033-295X.92.4.548

Wexley, K. N., Yukl, G. A., Kovacs, S. Z., & Sanders, R. E. (1972). Importance of contrast effects in employment interviews. *Journal of Applied Psychology*, *56*(1), 45-48. doi: https://doi.org/10.1037/h0032132

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. doi: https://doi.org/10.1080/14639220500370105

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a "cry wolf" effect? *Human Factors*, *51*(4), 446–462. doi: https://doi.org/10.1177/0018720809344720

Wixted, J. T., & Wells, G. L. (2017). The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis. *Psychological Science in the Public Interest*, *18*(1), 10–65. doi: https://doi.org/10.1177/1529100616686966

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(2), 345. doi: https://doi.org/10.1037/0096-1523.4.2.345

Yang, X. J., Guo, Y., & Schemanske, C. (To appear). From Trust to Trust Dynamics: Combining Empirical and Computational Approaches to Model and Predict Trust Dynamics in Human-Autonomy Interaction. In V. G. Duffy, S. J. Landry, J. D. Lee, & N. A. Stanton (Eds.), *Human-automation interaction: Transportation.*

Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017). Evaluating effects of user

experience and system transparency on trust in automation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 408–416). New York, NY, USA: ACM. doi: https://doi.org/10.1145/2909824.3020230

Yang, X. J., Wickens, C. D., & Hölttä-Otto, K. (2016). How users adjust trust in automation: Contrast effect and hindsight bias. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, pp. 196–200). doi: https://doi.org/10.1177/1541931213601044

Zhang, M. Y., & Yang, X. J. (2017). Evaluating effects of workload on trust in automation, attention allocation and dual-task performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, pp. 1799–1803). doi: https://doi.org/10.1177/1541931213601932

Biographies

X. Jessie Yang is an Assistant Professor in the Department of Industrial and Operations Engineering at the University of Michigan Ann Arbor. She obtained a PhD in Mechanical and Aerospace Engineering (Human Factors) from Nanyang Technological University, Singapore in 2014.

Christopher Schemanske is an MSE student in the Department of Industrial and Operations Engineering at the University of Michigan Ann Arbor. When the present work was conducted, he was an undergraduate student in the same department.

Christine Searle is a MS student at the Robotics Institute, University of Michigan Ann Arbor. She obtained a BA in Computer Science and Psychology in 2014 from Indiana University Bloomington.