Associations between Trust Dynamics and Personal Characteristics

Hyesun Chung
Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, USA
hyesunc@umich.edu

X. Jessie Yang
Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, USA
xijyang@umich.edu

***Abstract*—While personal characteristics influence people's snapshot trust towards autonomous systems, their relationships with trust dynamics remain poorly understood. To address this gap, we conducted a human-subject experiment involving 130 participants performing a simulated surveillance task aided by an automated threat detector. A comprehensive pre-experimental survey gathered participants' personal characteristics measured in 12 constructs and 28 dimensions. We clustered participants' trust dynamics into three groups, namely Bayesian decision-makers, disbelievers, and oscillators. Subsequently, we identified their distinctive attributes regarding personal characteristics. Results showed that the clusters differ in seven personal characteristics (masculinity, positive affect, extraversion, neuroticism, intellect, performance expectancy, high expectations). Our findings provide implications for designing specialized trust prediction algorithms tailored to each type of trust dynamics and trust-aware agents.**

***Index Terms*—trust dynamics, personal characteristics, clustering, human-autonomy interaction, human-robot interaction**

I. Introduction

The increasing adoption of automated and autonomous systems in various domains, including transportation, healthcare, education, and defense, has led to a growing interest in understanding how trust develops between humans and these systems [1], [2]. Early work on trust in automation/autonomy[1] is focused on identifying antecedents of trust. Numerous factors have been identified as influencing trust (see [4]–[9] for reviews). These factors can be categorized into those related to the human, the automated/autonomous system, and the context or environment [5]. In these early works, a person's trust in automation is measured at specific time points (*snapshot* trust), predominately at the end of an experiment after a series of interactions and/or sometimes at the beginning of an experiment.

More recently, acknowledging that a person's trust can change dynamically while interacting with autonomous technologies, there is a shift of research focus from snapshot trust to trust dynamics - how humans' trust in autonomy forms and evolves due to moment-to-moment interaction with autonomous technologies [10]–[12].

[1]To be consistent with early literature, we use the two terms automation and autonomy interchangeably in this paper while acknowledging the difference between the two [3].

One line of research on trust dynamics explores how trust is formed, violated, and recovered when humans interact with autonomous and robotic agents [10], [11], [13]–[15]. Numerous studies have demonstrated the effectiveness of different trust repair strategies aimed at restoring trust following violations [16], including apologies [17]–[19], denials [18], explanations [17], [20], [21], and promises [22]. The second line of research is focused on developing real-time trust prediction models and clustering trust dynamics [23]–[30]. Several studies revealed the existence of different types of trust dynamics [24], [25], [27], [28].

Despite existing research efforts on trust dynamics, one significant research gap remains. Most of the studies about the association between personal characteristics and trust center on post-experimental trust (measured at the end of a series of interactions) [31]–[34] and/or a person's trust propensity (i.e., an individual's inherent tendency to trust automation) [35]–[37]. While the importance of investigating trust dynamics over snapshot trust has been recognized [10], [12], [38], there is limited knowledge concerning the relationships between trust dynamics and personal characteristics. This gap makes it challenging to comprehend which personal characteristic factors impact trust dynamics the most, how they do so, and how diverse individuals exhibit distinct trust dynamics.

To address this gap, in this study, we conducted a human-subject experiment involving a total of 130 participants performing a simulated surveillance task aided by an imperfect automated threat detector. Based on key literature review studies on trust in automation, we constructed a comprehensive list of 12 personal characteristics constructs and 28 dimensions that could potentially influence a person's trust dynamics when interacting with autonomous agents. The survey was administered before the experiment. During the experiment, participants' dynamic trust when interacting with the automated threat detector was recorded. With the experimental data, we first performed clustering analysis and obtained three clusters of trust dynamics: Bayesian decision-maker, disbeliever, and oscillator. Subsequently, Analyses of Variance (ANOVAs) were conducted to explore differences across the three trust dynamics clusters.

II. Method

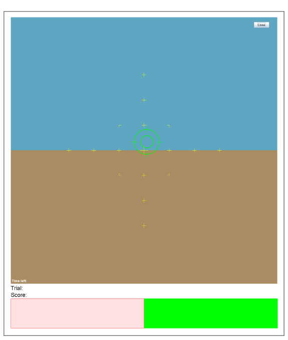
We conducted a human-subject experiment where participants engaged in a simulated surveillance task with the help of an imperfect automated threat detector. The research studies complied with the American Psychological Association code of ethics and were approved by the institutional review board.

A. Participants

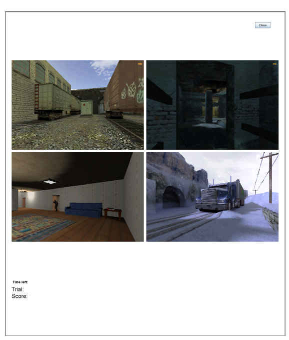
We gathered data from 130 individuals (average age=22.6 years, SD=3.5). All participants had normal or corrected-to-normal vision. None of them had previously taken part in a similar study or had any prior experience with the specific testbed used in the current research. Participants were compensated with \$20 upon completion of the experiment. Additionally, there was an opportunity to receive a performance-based bonus ranging from \$2.5 to \$10.

B. Apparatus and stimuli

During the experiment, the participants worked with a swarm of drones to perform a surveillance task at 100 sites (i.e. 100 trials). Each trial lasted for 10 seconds. During a trial, the participants had to perform two tasks simultaneously. They had to maintain a level flight of the drones, which was essentially a compensatory tracking task, while detecting potential threats from the photo feeds captured by the drones (Figure 1). At the beginning of each trial, participants started on the tracking display. They had access to either the tracking task or the detection task display at any given time and needed to switch between the two displays.



(a) Tracking task



(b) Detection task

Fig. 1: Simulated surveillance task consisting of the (a) tracking task and the (b) detection task

Tracking task. For each trial, participants engaged in a tracking task that persisted for 10 seconds. Using a joystick, participants had to control the movement of a green circle, which randomly drifted on the screen. The goal was to maneuver the circle towards a crosshair positioned at the center of the display (Figure 1(a)) and thereby minimize the distance between the green circle and the crosshair.

The performance of the tracking task was assessed based on the average distance between the circle's position and the center. This measurement captured how effectively participants could control the circle and maintain its proximity to the crosshair during the 10-second duration of each trial.

Detection task. In addition to the tracking task, participants were responsible for detecting threats. Each trial involved participants receiving a new set of four static images from the drones for threat detection. The images were presented as shown in Figure 1(b). The threat being detected was represented by a person, as illustrated in Figure 2(a), and only one threat appeared among the four images. There were no distractors included in the images so that the participants did not have to determine whether the person depicted was a friend or foe. The distribution of threats across the four images followed a uniform distribution, ensuring randomness.

(a) Images with threats

(b) Images without threats

Fig. 2: Examples of a threat

An imperfect automated threat detector assisted the participants in performing the threat detection task. If the detector identified a threat, an immediate visual and auditory alert in the form of a red "Danger" signal and a synthetic "Danger" sound were triggered at the beginning of each trial. Participants were expected to accurately and promptly identify the presence of the threat by pressing the "Report" button on the joystick. On the other hand, in cases where the detector did not identify any threat, the alert signal was "Clear", presented in green, with an auditory notification as well. Participants were not required to report the absence of a threat; their expected action was to take no action when there was no threat.

In both cases, participants had the option to either follow the decisions of the threat detector blindly or personally examine the images and make their own judgments. The performance of the detection task was assessed based on three metrics: detection time, detection accuracy, and detection score.

C. Experimental Design

The experimental data was collected under five different automation reliability levels: 62%, 64%, 66%, 68%, and 70%. Participants were randomly assigned to one of the five levels.

The number of each case (Hit, False alarm, Miss, Correct rejection) was configured using Signal Detection Theory (SDT) [39]. Benchmarking prior literature [40], the criterion c was set at -0.29, and the sensitivity d' was set to 1.09. Subsequently, the occurrences of each case were calculated and rounded, as presented in Table I.

TABLE I: Occurrences of hits, misses, false alarms, and correct rejections in each reliability level

Reliability	Alert	Threat	No Threat
62%	Danger	8	36
	Clear	2	54
64%	Danger	16	32
	Clear	4	48
66%	Danger	24	28
	Clear	6	42
68%	Danger	32	24
	Clear	8	36
70%	Danger	40	20
	Clear	10	30

D. Measures

In our study, we collected a range of subjective, behavioral and performance data.

Dynamic Trust. As mentioned before, there were 100 trials in the experiment. After each trial, participants' subjective trust rating was measured using a visual analog scale [11], [14], [38], [41]. The scale had "I don't trust the detector at all" as the leftmost anchor to "I absolutely trust the detector" as the rightmost anchor. The ratings provided by participants on the visual analog scale were then automatically converted to a value between 0 and 1.

Blindly Following and Crosschecking Behaviors. In each trial, participants could blindly follow the recommendation provided by the automated threat detector without crosschecking the detection display, or they could choose to check the detection display themselves. The two types of behaviors were referred to as blindly following and crosschecking.

Personal Characteristics. We collected a diverse set of personal characteristics of the participants. To identify relevant factors that affect users' trust in autonomy, we conducted a thorough review of key literature review and meta-analysis papers [4]–[9]. From these papers, we compiled all personal characteristics previously shown to be associated with trust in autonomy. As a result, we had a list of 12 distinct constructs, encompassing a total of 28 dimensions that related to personal characteristics. For each construct, we utilized well-established surveys to gather data (Table II).

Post-experiment Survey. In addition to the trust ratings obtained after each experimental trial, we also measured participants' subjective evaluations of the automated threat detector after completing the entire experimental trial. We asked for their overall trust, satisfaction, and understanding of the detector. Furthermore, we inquired about their self-confidence regarding their performance.

TABLE II: List of personal characteristics and measurement scales

Construct	Reference
Measures of culture	CVScale [42]
Measures of attentional control	ACS [43]
Measures of mood	PANAS [44]
Personality	Mini-IPIP [45]
Risk propensity	RPS [46]
Decision-making style	GDMS [47]
Reasoning test	CRT [48]
Propensity to trust	PTS [49]
Measures of negative attitude towards autonomous systems	NARS [50]
Measures of expectancy towards autonomous systems	UTAUT [51] Questions from [52]
Self-efficacy towards autonomous systems	Computer efficacy beliefs [53]
Perfect automation schema	PAS [54]

E. Experimental procedure

Upon arrival, participants were asked to provide informed consent and fill out a pre-experiment survey, including the demographics and the personal characteristics surveys. After that, participants received a training and practice session on the tasks. The practice session consisted of 30 trials with only the tracking task and eight trials combining both the tracking and detection tasks. The eight trials involved equal numbers of hits, misses, false alarms, and correct rejections, enabling participants to experience all possible cases. Participants were informed that the automated threat detector used during the practice was solely for illustration purposes and that the detector's reliability would not be the same during the experimental trials. The actual experiment consisted of a total of 100 trials, with a five-minute break after the $50th$ trial. After completing all trials, participants were asked to respond to the post-experiment survey.

III. Data Analysis and Results

In this study, we performed trust prediction for the 130 participants, followed by a clustering analysis. After assigning each participant to specific cluster groups based on the clustering analysis, we conducted ANOVA to investigate if there were any significant differences between the clusters.

A. Predicting Temporal Trust and Clustering Trust Dynamics

We employed the Beta random variable model [12], [24], [25] to predict a human's temporal trust. This model has been shown to outperform other models [23], [55] in prediction accuracy. Additionally, it provides good model explainability and generalizability because it complies with three properties of trust dynamics identified from empirical studies.

- Continuity: Trust at the current moment i is significantly associated with trust at the previous moment $i-1$.
- Negativity bias: Negative experiences exert a stronger influence on trust compared to positive experiences.
- Stabilization: Over time, a person's trust tends to stabilize during repeated interactions with the same system.

After an autonomous system completes the i^{th} task, the human's temporal $trust_i$ follows a Beta distribution. The predicted trust $\hat{trust}_i$ is calculated by the mean of $trust_i$

$$trust_i \sim Beta(\alpha_i, \beta_i) \tag{1}$$

$$\hat{trust}_i = E(trust_i) = \frac{\alpha_i}{\alpha_i + \beta_i} \tag{2}$$

$$\alpha_i = \begin{cases} \alpha_{i-1} + \omega^s, & if\ performance_i = 1\ (agent's\ success) \\ \alpha_{i-1}, & if\ performance_i = 0\ (agent's\ failure) \end{cases} \tag{3}$$

$$\beta_i = \begin{cases} \beta_{i-1} + \omega^f, & if\ performance_i = 0\ (agent's\ failure) \\ \beta_{i-1}, & if\ performance_i = 1\ (agent's\ success) \end{cases} \tag{4}$$

where

- $performance_i$: Performance of the autonomous system on the ith interaction
- α_i, β_i: Parameters of the Beta distribution after the ith interaction
- ω^s: Gains due to the human's positive experience toward the autonomous system after its success
- ω^f: Gains due to the human's negative experience toward the autonomous system after its failure

After n tasks/interaction, the system succeeds in n^s tasks and fails n^f tasks. Then

$$trust_i \sim Beta(\alpha_0 + n^s w^s, \beta_0 + n^f w^f) \tag{5}$$

$$\hat{trust}_i = E(trust_i) = \frac{\alpha_0 + n^s w^s}{\alpha_0 + n^s w^s + beta_0 + n^f w^f} \tag{6}$$

where α_0 and β_0 represent a person's *a priori* positive and negative experience with autonomy in general.

In this study, we personalized the trust model for each participant and learned the parameters $\{\alpha_0, \beta_0, w^s, w^f\}$ using their self-reported trust ratings as the ground truth labels. Additionally, since each participant reported their trust rating after every trial, we utilized each rating to iteratively update the model's parameters. The learned parameters were then used to predict $\hat{trust}_i$. For a detailed description of the model, please refer to [12], [24], [25].

Clustering based on trust dynamics. We applied k-means clustering to the participants within each reliability level. Following prior studies [12], [24], [25], the clustering was based on two factors: *Average logarithm trust* and *RMSE (Root Mean Square Error)*. The average logarithm trust represents participants' overall levels of trust, and the RMSE measures the deviation between the self-reported ground truth and the predicted trust value, indicating how closely the participant's trust dynamics follow the abovementioned properties of trust dynamics.

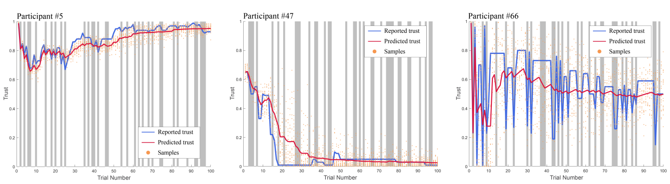
The two input features were calculated as follows:

$$Average\ logarithm\ trust = \frac{1}{100}\Sigma_{i=1}^{i=100} \log Trust_i \tag{7}$$

$$RMSE = \sqrt{\frac{1}{100}\Sigma_{i=1}^{i=100}(Trust_i - \hat{trust}_i)^2} \tag{8}$$

where $Trust_i$ is the participant's self-reported trust toward the autonomous system after the ith task and $\hat{trust}_i$ is the predicted trust score.

The number of clusters was determined using the elbow rule, which decides on the optimal number of clusters that best capture the patterns in the data. Consistent with prior literature [12], [24], [25], the k-means clustering resulted in three distinct clusters: Bayesian Decision Makers (BDMs, $n = 91$), disbelievers ($n = 25$), and oscillators ($n = 14$). The Bayesian decision makers update their trust in a Bayesian manner, the disbelievers display constantly low trust in autonomy, and the oscillators' trust fluctuates dramatically (Figure 3).



(a) BDM (b) Disbeliever (c) Oscillator

Fig. 3: Three distinct clusters of trust dynamics: Bayesian decision maker (BDM), disbeliever, and oscillator.

B. Association between Trust Dynamics, Personal Characteristics and Behaviors

We conducted Analyses of Variance (ANOVAs) to examine differences among the three clusters. For this study, our focus was on the personal characteristics measures collected through the pre-experiment survey. Table III presents the mean and standard deviation (SD) values for the dimensions of personal characteristics that displayed differences across the three cluster groups.

TABLE III: Mean and Standard Deviation (SD) of the three different clusters for key personal characteristics

Dimension	BDM	Disbeliever	Oscillator
Masculinity (/5) *	1.81 (0.67)	1.68 (0.77)	2.38 (1.12)
Positive affect (/5) *	2.80 (0.71)	2.75 (0.76)	3.37 (0.65)
Extraversion (/5) *	2.99 (0.95)	2.74 (0.96)	3.54 (1.08)
Neuroticism (/5) *	2.67 (0.73)	3.09 (0.83)	2.61 (0.90)
Intellect (/5) *	3.77 (0.72)	3.57 (0.93)	4.23 (0.46)
Performance expectancy (/7) **	5.73 (0.70)	5.24 (1.39)	6.07 (0.68)
PAS-High expectations (/5) **	1.88 (0.57)	1.55 (0.54)	2.14 (0.79)

Personal Characteristics. Results revealed significant differences in seven personal characteristics dimensions: *masculinity* ($F(2, 127) = 4.21, p = 0.02$), *positive affect* ($F(2, 127) = 4.11, p = 0.019$), *extraversion* ($F(2, 127) = 3.06, p = 0.050$), *neuroticism* ($F(2, 127) = 3.16, p = 0.046$), *intellect* ($F(2, 127) = 3.63, p = 0.029$), *performance expectancy* ($F(2, 127) = 4.76, p = 0.010$), and *PAS-high* *expectations* ($F(2,127) = 5.08, p = 0.008$). With respect to *masculinity*, the oscillator group scored significantly higher than the BDM ($p = 0.029$) and disbeliever groups ($p = 0.019$). Similarly, for *positive affect*, the oscillators had significantly higher scores than BDMs ($p = 0.021$) and disbelievers ($p = 0.032$). Concerning *extraversion*, oscillators had higher ratings than disbelievers ($p = 0.045$). For *neuroticism*, the disbeliever group had the highest scores, and its difference between that of BDM showed marginal significance ($p = 0.053$). In terms of *intellect*, the oscillator showed the highest mean score, which is larger than the BDM ($p = 0.095$) and disbeliever groups ($p = 0.025$). For *performance expectancy*, the disbeliever group had significantly lower scores compared to both BDM ($p = 0.041$) and oscillator ($p = 0.015$). Likewise, regarding *PAS-High expectations*, the mean value of disbelievers was the lowest compared to BDMs ($p = 0.041$) and oscillators ($p = 0.010$).

IV. Discussion and Conclusion

The objective of the current study was to explore the relationships between personal characteristics and trust dynamics in the context of human-autonomy interaction. To achieve this, we conducted a laboratory experiment to collect trust dynamics data from 130 participants as they engaged in a simulated surveillance task aided by an automated threat detector. Additionally, we gathered comprehensive personal characteristic data from participants through a survey, along with post-experiment subjective ratings. The analysis of the trust dynamics data revealed three distinct trust dynamics clusters: Bayesian Decision Maker (BDM), disbeliever, and oscillator. Subsequently, we conducted ANOVA and found seven personal characteristics that were significantly different across the three groups: *masculinity* from the measures of culture, *positive affect* from the measures of mood, *extraversion*, *neuroticism* and *intellect* from the personality measures, *performance expectancy* from the measures of expectancy towards autonomous systems, *high expectations* from the measures of perfect automation schema.

In contrast to prior research focusing on snapshot trust, our study delved deep into trust dynamics and employed a relatively large participant sample ($n = 130$). We first predicted a person's trust in real-time employing the Beta random variable model and clustered the trust dynamics. In contrast to other studies that utilized features associated with participants' behavior for clustering trust dynamics [27], [28], our approach focused solely on metrics derived from trust ratings. This emphasis aligns with the definition of trust –"the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability [6]". Consequently, our goal was to center the analysis on this attitude itself. This led to the identification of three distinct cluster groups, each exhibiting unique trust dynamics characteristics. It's worth noting that the clustering outcomes and their distribution exhibited similarities with those reported in [24], [25]. Our study benefits from a larger sample size ($n = 130$) and further proves the generalizability of the three clusters of trust dynamics across multiple datasets.

In addition, the study shed light on the existence of a group of individuals whose trust exhibits significant fluctuations over time (oscillator). Their trust dynamics made it very difficult to accurately predict their trust using the Bayesian inference model. This finding presents both a limitation and a future research avenue, suggesting the need for a more robust trust prediction model specifically designed to address the unique behavior of the oscillator group.

Acknowledgment

This work is supported by the National Science Foundation under Grant No. 2045009.

References

[1] T. B. Sheridan, "Human–robot interaction: status and challenges," *Human factors*, vol. 58, no. 4, pp. 525–532, 2016.

[2] National Academies of Sciences, Engineering, and Medicine, *Human-AI Teaming: State-of-the-Art and Research Needs*. Washington, DC: The National Academies Press, 2022.

[3] M. R. Endsley, "From Here to Autonomy: Lessons Learned from Human-Automation Research," *Human Factors*, vol. 59, no. 1, pp. 5–27, 2017. [Online]. Available: https://doi.org/10.1177/0018720816681350

[4] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Human factors*, vol. 57, no. 3, pp. 407–434, 2015.

[5] P. A. Hancock, T. T. Kessler, A. D. Kaplan, J. C. Brill, and J. L. Szalma, "Evolving trust in robots: specification through sequential and comparative meta-analyses," *Human factors*, vol. 63, no. 7, pp. 1196–1229, 2021.

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

[7] A. D. Kaplan, T. T. Kessler, J. C. Brill, and P. Hancock, "Trust in artificial intelligence: Meta-analytic findings," *Human factors*, vol. 65, no. 2, pp. 337–359, 2023.

[8] K. E. Schaefer, J. Y. Chen, J. L. Szalma, and P. A. Hancock, "A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems," *Human factors*, vol. 58, no. 3, pp. 377–400, 2016.

[9] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human factors*, vol. 53, no. 5, pp. 517–527, 2011.

[10] E. J. de Visser, M. M. Peeters, M. F. Jung, S. Kohn, T. H. Shaw, R. Pak, and M. A. Neerincx, "Towards a theory of longitudinal trust calibration in human–robot teams," *International Journal of Social Robotics*, vol. 12, no. 2, pp. 459–478, 2020.

[11] X. J. Yang, C. Schemanske, and C. Searle, "Toward quantifying trust dynamics: How people adjust their trust after moment-to-moment interaction with automation," *Human Factors*, vol. 65, no. 5, pp. 862–878, 2023.

[12] X. J. Yang, Y. Guo, and C. Schemanske, "From trust to trust dynamics: Combining empirical and computational approaches to model and predict trust dynamics in human-autonomy interaction," in *Human-Automation Interaction: Transportation*. Springer, 2022, pp. 253–265.

[13] J. D. Lee and N. Moray, "Trust, self-confidence, and operators' adaptation to automation," *International Journal of Human-Computer Studies*, vol. 40, no. 1, pp. 153–184, 1994.

[14] D. Manzey, J. Reichenbach, and L. Onnasch, "Human performance consequences of automated decision aids: The impact of degree of automation and system experience," *Journal of Cognitive Engineering and Decision Making*, vol. 6, no. 1, pp. 57–87, 2012.

[15] C. Esterwood and L. P. Robert Jr, "Three strikes and you are out!: The impacts of multiple human–robot trust violations and repairs on robot trustworthiness," *Computers in Human Behavior*, vol. 142, p. 107658, 2023.

[16] R. M. Kramer and R. J. Lewicki, "Repairing and enhancing trust: Approaches to reducing organizational trust deficits," *The Academy of Management Annals*, vol. 4, no. 1, pp. 245–277, 2010.

[17] M. Natarajan and M. Gombolay, "Effects of anthropomorphism and accountability on trust in human robot interaction," in *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, 2020, pp. 33–42.

[18] S. C. Kohn, A. Momen, E. Wiese, Y.-C. Lee, and T. H. Shaw, "The consequences of purposefulness and human-likeness on trust repair attempts made by self-driving vehicles," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2019, pp. 222–226.

[19] A. Mahmood, J. W. Fung, I. Won, and C.-M. Huang, "Owning mistakes sincerely: Strategies for mitigating ai errors," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–11.

[20] S. M. Faas, J. Kraus, A. Schoenhals, and M. Baumann, "Calibrating pedestrians' trust in automated vehicles: does an intent display in an external hmi support trust calibration and safe crossing behavior?" in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–17.

[21] Z. Ashktorab, M. Jain, Q. V. Liao, and J. D. Weisz, "Resilient chatbots: Repair strategy preferences for conversational breakdowns," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–12.

[22] Y. Albayram, T. Jensen, M. M. H. Khan, M. A. A. Fahim, R. Buck, and E. Coman, "Investigating the effects of (empty) promises on human-automation interaction and trust repair," in *Proceedings of the 8th International Conference on Human-Agent Interaction*, 2020, pp. 6–14.

[23] A. Xu and G. Dudek, "OPTIMo: Online probabilistic trust inference model for asymmetric human-robot collaborations," in *Proceedings of the 2015 ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, 2015, pp. 221–228.

[24] Y. Guo and X. J. Yang, "Modeling and predicting trust dynamics in human–robot teaming: A bayesian inference approach," *International Journal of Social Robotics*, vol. 13, no. 8, pp. 1899–1909, 2021.

[25] S. Bhat, J. B. Lyons, C. Shi, and X. J. Yang, "Clustering trust dynamics in a human-robot sequential decision-making task," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8815–8822, 2022.

[26] M. Chen, S. Nikolaidis, H. Soh, D. Hsu, and S. Srinivasa, "Planning with Trust for Human-Robot Collaboration," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*, 2018, pp. 307–315.

[27] G. McMahon, K. Akash, T. Reid, and N. Jain, "On modeling human trust in automation: Identifying distinct dynamics through clustering of markovian models," *IFAC-PapersOnLine*, vol. 53, pp. 356–363, 2020.

[28] J. Liu, K. Akash, T. Misu, and X. Wu, "Clustering human trust dynamics for customized real-time prediction," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 1705–1712.

[29] Y. Guo, X. J. Yang, and C. Shi, "Enabling Team of Teams: A Trust Inference and Propagation (TIP) Model in Multi-Human Multi-Robot Teams," in *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, July 2023.

[30] J. Ayoub, L. Avetisian, X. J. Yang, and F. Zhou, "Real-Time Trust Prediction in Conditionally Automated Driving Using Physiological Measures," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 12, pp. 14 642–14 650, 2023.

[31] W. Cai, Y. Jin, and L. Chen, "Impacts of personal characteristics on user trust in conversational recommender systems," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–14.

[32] J. Zhou, S. Luo, and F. Chen, "Effects of personality traits on user trust in human–machine collaborations," *Journal on Multimodal User Interfaces*, vol. 14, pp. 387–400, 2020.

[33] N. N. Sharan and D. M. Romano, "The effects of personality and locus of control on trust in humans versus artificial intelligence," *Heliyon*, vol. 6, no. 8, 2020.

[34] H.-Y. Huang, M. Twidale, and M. Bashir, "'if you agree with me, do i trust you?': An examination of human-agent trust from a psychological perspective," in *Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2*. Springer, 2020, pp. 994–1013.

[35] L. Tang, P. Ferronato, and M. Bashir, "Do users' values influence trust in automation?" in *International Conference on Intelligent Human Computer Interaction*. Springer, 2022, pp. 297–311.

[36] H.-Y. Huang and M. Bashir, "Users' trust in automation: a cultural perspective," in *Advances in Human Factors in Robots and Unmanned Systems: Proceedings of the AHFE 2017 International Conference on Human Factors in Robots and Unmanned Systems, July 17- 21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8*. Springer, 2018, pp. 282–289.

[37] S.-Y. Chien, K. Sycara, J.-S. Liu, and A. Kumru, "Relation between trust attitudes toward automation, hofstede's cultural dimensions, and big five personality traits," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 60, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2016, pp. 841–845.

[38] X. J. Yang, V. V. Unhelkar, K. Li, and J. A. Shah, "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*, 2017, pp. 408–416.

[39] H. Stanislaw and N. Todorov, "Calculation of signal detection theory measures," *Behavior research methods, instruments, & computers*, vol. 31, no. 1, pp. 137–149, 1999.

[40] D. Manzey, N. Gérard, and R. Wiczorek, "Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload," *Ergonomics*, vol. 57, no. 12, pp. 1833–1855, 2014.

[41] N. Du, K. Y. Huang, and X. J. Yang, "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*, vol. 62, no. 6, pp. 987–1001, 2020.

[42] B. Yoo, N. Donthu, and T. Lenartowicz, "Measuring hofstede's five dimensions of cultural values at the individual level: Development and validation of cvscale," *Journal of international consumer marketing*, vol. 23, no. 3-4, pp. 193–210, 2011.

[43] D. Derryberry and M. A. Reed, "Anxiety-related attentional biases and their regulation by attentional control." *Journal of abnormal psychology*, vol. 111, no. 2, p. 225, 2002.

[44] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the panas scales." *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.

[45] M. B. Donnellan, F. L. Oswald, B. M. Baird, and R. E. Lucas, "The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality." *Psychological assessment*, vol. 18, no. 2, p. 192, 2006.

[46] R. M. Meertens and R. Lion, "Measuring an individual's tendency to take risks: the risk propensity scale 1," *Journal of applied social psychology*, vol. 38, no. 6, pp. 1506–1520, 2008.

[47] D. P. Spicer and E. Sadler-Smith, "An examination of the general decision making style questionnaire in two uk samples," *Journal of Managerial Psychology*, vol. 20, no. 2, pp. 137–149, 2005.

[48] M. E. Toplak, R. F. West, and K. E. Stanovich, "Assessing miserly information processing: An expansion of the cognitive reflection test," *Thinking & reasoning*, vol. 20, no. 2, pp. 147–168, 2014.

[49] S. M. Merritt, H. Heimbaugh, J. LaChapell, and D. Lee, "I trust it, but i don't know why: Effects of implicit attitudes toward automation on trust in an automated system," *Human factors*, vol. 55, no. 3, pp. 520–534, 2013.

[50] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Measurement of negative attitudes toward robots," *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, vol. 7, no. 3, pp. 437–454, 2006.

[51] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS quarterly*, pp. 425–478, 2003.

[52] A. Alaiad and L. Zhou, "The determinants of home healthcare robots adoption: An empirical investigation," *International journal of medical informatics*, vol. 83, no. 11, pp. 825–840, 2014.

[53] T. Hill, N. D. Smith, and M. F. Mann, "Role of efficacy expectations in predicting the decision to use advanced technologies: The case of computers." *Journal of applied psychology*, vol. 72, no. 2, p. 307, 1987.

[54] S. M. Merritt, J. L. Unnerstall, D. Lee, and K. Huber, "Measuring individual differences in the perfect automation schema," *Human factors*, vol. 57, no. 5, pp. 740–753, 2015.

[55] J. Lee and N. Moray, "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, 1992.