# Trust Dynamics in Human-AV (Automated Vehicle) Interaction

**Ruikun Luo**
Robotics Institute
University of Michigan
Ann, Arbor, MI 48109, USA
ruikunl@umich.edu

**Jian Chu**
Department of Mechanical Engineering
The University of Texas at Austin
Austin, TX 78712, USA
jian_chu@utexas.edu

**X. Jessie Yang**
Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109, USA
xijyang@umich.edu


*CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.*

ACM ISBN 978-1-4503-6819-3/20/04.
https://dx.doi.org/10.1145/3334480.3382970

## Abstract
Despite many benefits of automated driving, such as reducing fuel consumption, traffic congestion and crashes, a lack of trust hinders the adoption of automated vehicles (AVs). Prior research focused on people's trust in AVs based on AVs' overall performance. The present study is focused on people's trust change in AVs over time in a sequential decision making context. We conducted a human-in-the-loop experiment with 16 participants in a virtual 3D environment wherein participants acted as passengers riding an AV. We manipulated two independent variables: level of stochasticity (high vs. low) and source of stochasticity (external vs. internal). Dependent variables included participants' moment-to-moment trust in AVs and post-experiment trust. Our results revealed that when the stochasticity was due to internal errors (e.g. AV's sensor errors) as compared to external errors (e.g. traffic jams or road blocks), participants' trust in AVs decreased more significantly. Also, the larger the cost due to an error, the larger the trust decrement.

## Author Keywords
Trust dynamics; Markov Decision Process; Human-automation interaction; Human-robot interaction; Internal and external risk

## CCS Concepts

**•Human-centered computing → HCI theory, concepts and models;**

## Introduction

Despite the many benefits of automated driving, such as reducing fuel consumption, traffic congestion and crashes, a lack of trust hinders the deployment of automated vehicles (AVs) on the road. Therefore, research efforts have now been undertaken to promote trust in and acceptance of AVs [3, 10, 11, 14].

For example, Waytz et al. [12] showed that anthropomorphism of the AV significantly affects the drivers' trust in AVs. Koo et al. [8] found that when AV explained what it was going to do before its action, it decreased the participants' anxiety and increased their preference for the AV. Forster et al. [5] showed that the explaining the AV's next action with voice was rated as superior on AV trust, anthropomorphism and usability compared to the interface without explanation. More recently, Haspiel et al. [6] and Du et al. [3] showed that the AV passengers had higher trust in the AV when it provided explanations before an unexpected event than after an unexpected event happened.

Existing studies shed light on the factors influencing drivers' and passengers' trust in and acceptance of AVs. However, they primarily examined trust once at the end of an experiment, with little emphasis on the form and evolution of trust over time [2, 13]. To address this limitation, we aim to examine people's trust dynamics when interacting with an AV. Specifically, how do the level of stochasticity and source of stochasticity affect people's trust dynamics in AVs?

An AV plans its actions based on all available information, such as time to destination and fuel economy. Suppose an AV plans its optimal policy for moving from point A to point B (i.e., the set of optimal actions the AV should take at each intersection to move from point A to point B). Most of the time, the AV will be able to reach its intended location at each intersection. However, occasionally due to some unexpected stochastic events, such as traffic jams, road blocks or sensor errors, the AV will not reach the intended location. In this study, we used Markov Decision Process to model this sequential decision making problem. Of interest in this study are the level of stochasticity - the extent to which an AV is able to reach its intended location by following its optimal policy, and the source of stochasticity - whether the unexpected stochastic event is due to external factors which are difficult to anticipate in advance (e.g. traffic jams or road blocks), or due to internal factors attributed to the AV (e.g. AV's sensor errors). In addition, we examined how the consequence of an stochastic event would influence people's trust in AVs.

We hypothesized that level of stochasticity, source of stochasticity and outcome of the AV's decision affect human drivers' trust in AVs and tested the following hypotheses:

- **H1**: Higher stochasticity would lower human users' trust toward AVs.
- **H2**: The stochasticity resulting from internal factors would lead to greater trust decrement than that from external factors.
- **H3**: The higher the cost resulted from the unexpected event, the greater the trust decreases.

## Method

*Participants*

16 participants (Age: Mean = $25.44$ years, $SD$ = 2.56 years) took part in the experiment. Among the 16 participants, 6


**Figure 1:** Map of the city. Red dots: Origin and destination pairs, $O_1, D_1, O_2, D_2$. Magenta arrow: AV's current location and orientation. Black solid line: path has already passed. Green solid line: planned optimal path to the destination. Dotted light green line: previous planned optimal path to the destination before error occurs.

were female and 10 were male. 15 out of the 16 participants had valid driving licenses in or outside of U.S. Participants were compensated with $\$20$ upon completion of the experiment.

*Apparatus and task*

We developed a 3D simulation environment as shown in Figure 2. In this environment, each participant acted as a passenger of an AV. The AV drove in the city from the origin to the destination (i.e. from $O_1$ to $D_1$ and from $O_2$ to $D_2$ as shown in Figure 1). The naviagation problem was modeled as a Markov Decision Process (MDP) and the optimal policy was computed via policy iteration. The MDP provides a mathematical framework for modeling sequential decision making where the outcomes are stochastic. We considered

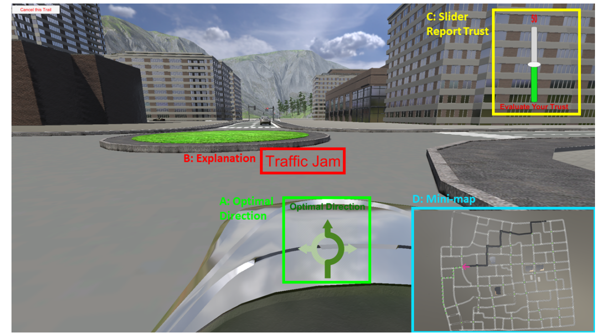


**Figure 2:** The 3D Environment with Graphics User Interface (GUI). A: Bottom of the screen shows the optimal direction at each intersection. B: Center of screen shows the explanations of different unexpected events. C: Top right of the screen shows the slider used by participants to report their trust. D: Bottom right of the screen shows the mini-map.

two sources of stochasticity in the simulation environment: external factors such as traffic jams and road blocks, and internal factors such as AV's sensor errors.

*Design of the Graphical User Interface (GUI)*

Figure 2 shows the GUI. At the bottom of the screen, the GUI indicates the optimal direction that the AV initially plans at an intersection. Due to the inherent stochasticity (i.e. traffic jams, road blocks, AV's sensor errors), the AV may not be able to reach the intended direction. When an unexpected event happens, an explanation in both auditory and visual modalities (in the center of the screen) will be provided to the participants as show in Figure 3.

The top right corner indicates the participants' moment-to-moment trust. After passing an intersection, participants were asked to report their trust in AVs via the slider.


**(a)**


**(b)**


**(c)**

**Figure 3:** External factors: (a)traffic jam, (b)road block. Internal factors: (c)sensor error.

The mini-map at the bottom right corner shows the city map. The mini-map tracks the AV's location in real time and shows the planned path from the AV as shown in Figure 1. The magenta arrow indicates the current location and orientation of the vehicle. The solid black line indicates the path that the vehicle has already passed. The solid green line indicates the planned optimal path to the destination. The dotted green line indicates the previous planned optimal path to the destination before the unexpected event occurs.

*Experimental design*

The experiment used a 2x2 within-subjects design with two independent variables - level of stochasticity (high vs. low) and source of stochasticity (internal vs. external). We selected two pairs of origins and destinations (i.e., $O_1$, $D_1$, $O_2$, $D_2$ as shown in Figure 1). Combining the two different levels of stochasticity and two pairs of origins and destinations, each participant experienced four rides. The presentation of the four rides followed a $4 \times 4$ Latin square to eliminate potential order effects.

**Independent variables:** There were two levels of stochasticity (high vs. low). The level of stochasticity was the probability that the AV cannot achieve the optimal direction at each intersection (high: 40% vs. low: 15%). Furthermore, there were two sources of stochasticity. One was external factors which were from the environment and hence were difficult to anticipate (e.g. road blocks and traffic jams) and the other was internal factors attributed to the AV (e.g., sensor errors). Within each ride, internal (sensor errors) and external stochastic (traffic jams and road blocks) events were generated randomly with equal probability. The probability of traffic jam and road block were also the same.

**Dependent variables:** Dependent measures in this experiment included participants' subjective trust and the length of the planned optimal path. We measured participants' subjective trust in AVs using two methods: participants indicated their moment-to-moment trust after each intersection, and rated their post-experiment trust in the AV after each ride using Moray's trust survey [9]. We measured the change of optimal path length when an stochastic event happened (e.g., At an intersection, the AV may fail to follow the optimal path it initially planned due to sensor errors. And this re-routing will likely affect the the AV's subsequent path and the length of the path.)

*Procedure*

Upon arrival, participants signed an informed consent and filled in a demographic survey. After that participants received a training on the simulator and the map of the city. During the map viewing session, participants were presented with the map of the city and the length of each road on the map. Participants were then presented with 28 pairs of paths. Each pair had the same origin and destination, but different routes (magenta and blue path in Figure 4). The length of each route was given on the top of the figure. The participants were required to review the map and the 28 pairs of paths for at least 10 minutes.

## Results

We conducted linear mixed effects analysis with each participant as a random intercept. Results are reported as significant for $\alpha < .05$.

*Level of Stochasticity*

The results indicated that stochasticity level significantly affects participants' post-trial trust and moment-to-moment trust in AVs. Higher stochasticity level leads to lower post-trial trust in AVs ($F(1, 15) = 15.568, p = .001$). Similarly, higher stochasticity leads to lower moment-to-moment trust report ($F(1, 1377.401) = 285.626, p < .001$) as shown in Figure 5.

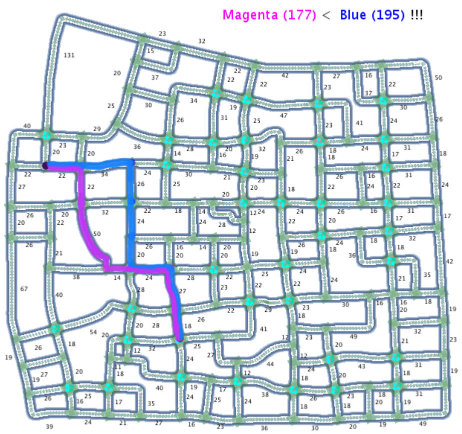


**Figure 4:** Training map. Magenta and blue path indicated two routes from the same origin and destination. Length of each route is listed on top of the image.

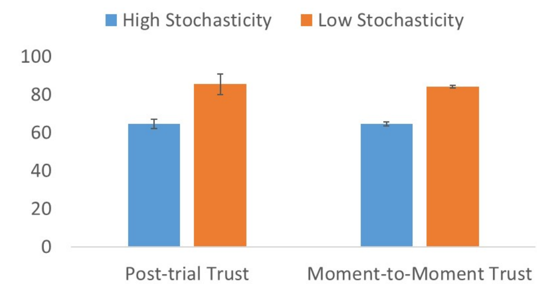


**Figure 5:** Mean and standard error (SE) of post-trial trust and moment-to-moment trust

| | Trust change |
|---|---|
| Unexpected | $-6.37 \pm 0.78$ |
| Expected | $3.31 \pm 0.34$ |
| External | $-0.91 \pm 0.68$ |
| Internal | $-11.35 \pm 1.28$ |

**Table 1:** Mean and standard error (SE) of participants' moment-to-moment trust change

*Source of Stochasticity*

The results indicated that the occurrence of stochastic events had a significant impact on participants' moment-to-moment trust changes ($F(1, 1391) = 177.813$, $p < .001$). When the AV successfully followed the intended optimal direction, participants' trust increased by $3.31 \pm 0.34$. However, when stochastic events happened (i.e. AV failed to reach the intended optimal direction due to traffic jams, road blocks or sensor errors), participants' moment-to-moment trust decreased by $6.37 \pm 0.78$ as shown in Table 1.

The results also indicated that the source of stochasticity (external factors vs. internal factors) significantly affects the participants' trust changes ($F(1, 423.383) = 58.229$, $p < .001$). If the source of stochasticity was due to external factors (i.e. traffic jams, road blocks), participants' moment-to-moment trust decreased by $-0.91 \pm 0.68$. However, if the source of stochasticity was due to internal factors (i.e. sensor errors), participants' moment-to-moment trust decreased by $-11.35 \pm 1.28$ as shown in Table 1.

*Change of the Optimal Path Length*

The results indicated that when unexpected stochastic events occurred, the change of the optimal path length significantly affected participants' moment-to-moment trust change ($F(1, 426.097) = 4.012$, $p = .046$) as shown in Figure 6. The larger the increment of path length (i.e., longer distance), the larger the trust decrement.

## Discussion & conclusion

Consistent with our previous finding [7], results from the present study indicates that higher stochasticity leads to lower post-trial trust in AVs. In addition, we analyzed the trust value and trust change value. Our results also indicated that level of stochasticity significantly influence the participants' moment-to-moment trust, which supported our hypothesis **H1**.

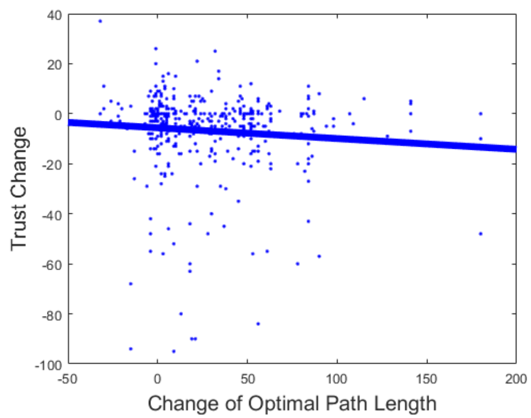


**Figure 6:** Trust change vs. change of the optimal path length

Our results also supported **H2**, that stochasticity resulted from internal factors leads to greater trust decrement than that from external factors. [4] built a belief-based trust computation model while distinguishing external and internal factors, however, they did not study how external and internal factors influenced human's trust and assigned the same weights on both factors. In the present study, we found a significant difference between external and internal factors.

**H3** was also supported by the results. When stochastic events occurred, the larger the negative consequence was, the larger the trust decrement. This finding suggests that people display outcome biases when assessing their trust in AVs. Outcome bias is one type of decision-making biases, that people tend to evaluate the quality of a decision based on the outcome of the decision. When outcome bias occurs, the same decision will be evaluated as worse when it happens to produce bad outcome rather than good outcome, even if the outcome is determined by chance. Previous research has been studied this phenomena in different contexts, such as medical decisions, monetary gambles [1].

Our findings have implications on the design of AVs. Our results showed that people's moment-to-moment trust decrement was significantly smaller when an unexpected stochastic event was due to external factors. This suggests that framing the source of errors as external may prevent drastic loss of trust. For example, sensor errors could be due to harsh weather, which may be perceived as external instead of internal errors. It could also be useful to educate AV riders on the various factors (e.g., traffic jam, road block) that are difficult to anticipate and predict, which may lower the riders' expectations of AVs and preserve riders' trust even when unexpected events occur.

There are several limitations of this study. First, the MDP based navigation algorithm assumes that the AV can observe the states accurately. This assumption is often violated in the real life. Therefore, we aim to explore Partially Observable MDP in future works. Second, in this study, the explanations of unexpected events (i.e. traffic jams, road blocks, sensor errors) as shown in Figure 3 were treated given by the 3D simulator instead of a display of the AV. In reality, those information should be provided by the AV and may not be correct, i.e. the AV reports traffic jam but there is no traffic jam. In the future work, we will address this issue by claiming that the explanations are given by the AV and add uncertainties to these explanations.

## REFERENCES


[1] Jonathan Baron and John C Hershey. 1988. Outcome bias in decision evaluation. *Journal of personality and social psychology* 54, 4 (1988), 569.

[2] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. 2013. Impact of

robot failures and feedback on real-time trust. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 251–258.

[3] Na Du, Jacob Haspiel, Qiaoning Zhang, Dawn Tilbury, Anuj K. Pradhan, X. Jessie Yang, and Lionel P. Robert. 2019. Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transportation Research Part C: Emerging Technologies* 104 (jul 2019), 428–442.

[4] Rino Falcone, Giovanni Pezzulo, and Cristiano Castelfranchi. 2002. A fuzzy approach to a belief-based trust computation. In *Workshop on Deception, Fraud and Trust in Agent Societies*. Springer, 73–86.

[5] Yannick Forster, Frederik Naujoks, and Alexandra Neukum. 2017. Increasing anthropomorphism and trust in automated driving functions by adding speech output. *IEEE Intelligent Vehicles Symposium, Proceedings* 2, Iv (2017), 365–372.

[6] Jacob Haspiel, Na Du, Jill Meyerson, Lionel P Robert Jr, Dawn Tilbury, X Jessie Yang, and Anuj K Pradhan. 2018. Explanations and expectations: Trust building in automated vehicles. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 119–120.

[7] Changhoon Kim, Mengyuan Zhang, Chongjie Zhang, and X. Jessie Yang. 2018. Trust dynamics in sequential decision making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62, 1 (2018), 165–166.

[8] Jeamin Koo, Dongjun Shin, Martin Steinert, and Larry Leifer. 2016. Understanding driver responses to voice alerts of autonomous car operations. *International journal of vehicle design* 70, 4 (2016), 377–392.

[9] Bonnie M. Muir and Neville Moray. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 3 (mar 1996), 429–460.

[10] Luke Petersen, Dawn Tilbury, Xi Jessie Yang, and Lionel Robert. 2017. Effects of augmented situational awareness on driver trust in semi-autonomous vehicle operation. (2017).

[11] Frank MF Verberne, Jaap Ham, and Cees JH Midden. 2012. Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human factors* 54, 5 (2012), 799–810.

[12] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology* 52 (2014), 113–117.

[13] X Jessie Yang, Vaibhav V Unhelkar, Kevin Li, and Julie A Shah. 2017. Evaluating effects of user experience and system transparency on trust in automation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, 408–416.

[14] Feng Zhou, X Jessie Yang, and Xin Zhang. 2019. Takeover Transition in Autonomous Vehicles: A YouTube Study. *International Journal of Human–Computer Interaction* 0, 0 (2019), 1–12.