Trust-Based Route Planning for Automated Vehicles

Shili Sheng
School of Engineering
University of Virginia
ss7dr@virginia.edu

Erfan Pakdamanian
School of Engineering
University of Virginia
ep2ca@virginia.edu

Kyungtae Han
Toyota InfoTech Labs
kyungtae.han@toyota.com

Ziran Wang
Toyota InfoTech Labs
ziran.wang@toyota.com

John Lenneman
Toyota Collaborative Safety Research Center
john.lenneman@toyota.com

Lu Feng
School of Engineering
University of Virginia
lu.feng@virginia.edu

arXiv:2101.03267v1 [cs.HC] 9 Jan 2021

ABSTRACT

Several recent works consider the personalized route planning based on user profiles, none of which accounts for human trust. We argue that human trust is an important factor to consider when planning routes for automated vehicles. This paper presents the first trust-based route planning approach for automated vehicles. We formalize the human-vehicle interaction as a partially observable Markov decision process (POMDP) and model trust as a partially observable state variable of the POMDP, representing human's hidden mental state. We designed and conducted an online user study with 100 participants on the Amazon Mechanical Turk platform to collect data of users' trust in automated vehicles. We build data-driven models of trust dynamics and takeover decisions, which are incorporated in the POMDP framework. We compute optimal routes for automated vehicles by solving optimal policies in the POMDP planning. We evaluated the resulting routes via human subject experiments with 22 participants on a driving simulator. The experimental results show that participants taking the trust-based route generally resulted in higher cumulative POMDP rewards and reported more positive responses in the after-driving survey than those taking the baseline trust-free route.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computing methodologies** → **Planning and scheduling**.

KEYWORDS

Trust, Automated Vehicle, Route Planning

ACM Reference Format:
Shili Sheng, Erfan Pakdamanian, Kyungtae Han, Ziran Wang, John Lenneman, and Lu Feng. 2021. Trust-Based Route Planning for Automated Vehicles. In *12th ACM/IEEE International Conference on Cyber-Physical Systems (with CPS-IoT Week 2021) (ICCPS '21), May 19–21, 2021, Nashville, TN, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/0

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICCPS '21, May 19–21, 2021, Nashville, TN, USA
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 0…$15.00
https://doi.org/0

1 INTRODUCTION

Recent years have witnessed significant advances in the development of automated vehicle, which have already been tested over millions of miles on public roads [4]. However, fully autonomous vehicles that do not require human intervention are still decades away due to technology, infrastructure, and regulation limitations [18]. The majority of automated vehicles available to the general public nowadays are Level 2 and Level 3 of automation [12], which allow the driver to turn attention away from the primary task of driving; but the driver must still be prepared to take over control of the vehicle when necessary. Human's decision on whether or not to rely on the automation is guided by trust. Prior studies have found that distrust is a main barrier to adoption of automated vehicles [27]; in addition, users with lower trust levels take over control of the vehicle more frequently [28]. On the other hand, overtrust in automation can lead to catastrophic outcomes (e.g., fatal Tesla autopilot crashes [3]). Therefore, in order to improve safety and user experience, there is a need for taking into account human trust in the system design of automated vehicles.

In this paper, we consider the design of route planning system for the navigation of automated vehicles. Existing route planning methods (e.g., [7, 19, 26]) mostly focus on computing routes that optimize metrics such as distance, travel time, and fuel consumption. Several recent works (e.g., [9, 13, 41]) consider the personalized route recommendation based on user profiles (e.g., mobility options, frequently visited places). However, none of the existing route planning methods explicitly account for human trust. We argue that human trust is an important factor to consider when planning routes for automated vehicles. For example, if the driver has lower trust in the automated vehicle's capability for safely navigating urban streets with pedestrians constantly crossing as opposed to freeways, the driver may prefer a freeway despite longer distance.

To the best of our knowledge, this paper presents the first work of *trust-based route planning* for automated vehicles. There are several challenges in developing this work. First, how to measure and model human trust in automation, which is a hidden mental state influenced by many factors and changes over time [38]. Second, how to incorporate the trust model into the route planning while accounting for the human-vehicle interaction (e.g., takeover decisions). Finally, how to evaluate the proposed trust-based route

planning approach. In the following, we provide an overview of how we address these challenges in this work.

We follow the notion of trust in automation defined in [34], which views human trust as delegation of responsibility for actions to the automation and willingness to accept risk (possible harm), while the decision to delegate is based on a subjective evaluation of the automation's capability for a particular task. To concretize the problem, we consider a motivating example where the automated vehicle may encounter three types of typical road incidents (i.e., pedestrian, obstacle, and oncoming truck). Trust is therefore affected by human's takeover decision and the vehicle's capability of handling an incident. We adopt the commonly used method of measuring the subjective belief of trust via user questionnaires. Specifically, we designed and conducted an online user study with 100 participants on the Amazon Mechanical Turk platform. We asked users to watch various driving videos recorded in the driver's view and answer questions about their trust in the automated vehicle's capability of safely handling the incident shown in the video in a 7-point Likert scale, as well as whether they would like to take over control of the vehicle imaging that they were the driver sitting inside the automated vehicle. We model the evolution of trust dynamics (i.e., how trust changes over time) as a linear Gaussian system using the data collected from the online user study. We also build data-driven models to predict human's takeover decisions.

We formalize the human-vehicle interaction as a partially observable Markov decision process (POMDP), which is a general modeling framework for planning under uncertainty [24]. We model trust as a partially observable state variable of the POMDP, representing human's hidden mental state. In addition, there are three observable state variables representing the vehicle position, incident type, and the success/failure of the vehicle handling an incident. The estimated trust dynamics model informs the probabilistic transition function of the trust variable in the POMDP. There are two actions: human's takeover decision and the vehicle's route choice. Since the vehicle does not know about human's actual takeover decision in advance, it assumes that human follows the data-driven takeover decision models estimated using the online user study data. The goal of POMDP planning is to compute an optimal policy that makes route choices to maximize the expectation of the cumulative reward, with a reward function designed to promote better safety and user experience of automated vehicles.

We applied the proposed trust-based route planning approach to the motivating example and obtained two routes: a trust-based route where human makes takeover decisions based on trust dynamics and incidents, and a trust-free route (as a baseline for comparison) where human's takeover decisions only depend on incidents. We evaluated and compared the performance of these two routes via human subject experiments on a driving simulator. We conducted experiments with 22 participants, who were randomly assigned to two equal-sized groups for the between-subject study (each group has 11 participants, who took one of the two routes). The experimental results show that participants taking the trust-based route generally resulted in higher cumulative POMDP rewards and reported more positive responses in the after-driving survey than those taking the trust-free route.

Contributions. We summarize the major contributions of this work as follows.

- We developed the first trust-based route planning approach for automated vehicles, which is based on a POMDP framework and uses data-driven models of trust dynamics and takeover decisions.
- We designed and conducted an online user study with 100 participants on the Amazon Mechanical Turk platform to collect data about users' trust in automated driving.
- We designed and conducted human subject experiments with 22 participants on a driving simulator to evaluate the proposed approach, which showed encouraging results.

Paper organization. The rest of the paper is organized as follows. We discuss the related work in Section 2, describe the motivating example in Section 3, present the trust-based route planning approach in Section 4, describe the driving simulator experiments in Section 5, and draw conclusions in Section 6.

2 RELATED WORK

In this section, we survey the related work in two topics: (1) route planning for vehicles, and (2) trust in automation. For each topic, we identify gaps in the state-of-the-art and discuss the connection with this paper.

2.1 Route Planning for Vehicles

The goal of route planning is to compute the optimal routes for vehicles. The most commonly used metrics include distance, travel time, and fuel consumption. Graph search algorithms such as Dijkstra's algorithm [14] and A^* algorithm [21] can be applied to find the shortest distance path between any two locations. Computing the fastest route (i.e., with the least travel time) is more challenging than finding the shortest distance route. Kanoulas et al. [26] extended A^* algorithm by considering the speed change at different time of the day to compute the fastest route. Gonzalez et al. [19] developed an adaptive fastest route planning method based on information learned from the historical traffic data, accounting for various factors (e.g., road quality, weather condition, area crime rate) that may influence vehicle speed patterns. Andersen et al. [7] proposed to find the most eco-friendly route by assigning eco-weights based on GPS and fuel consumption data.

There are several recent studies considering personalized route recommendation for various users. Campigotto et al. [9] developed a method for the personalized route planning by using Bayesian learning to update users' profile such as home location, work place, and mobility options. Dai et al. [13] recommended a personalized optimal route considering user preferences encoded as a ratio between different metrics such as distance, travel time, and fuel consumption. Zhu et al. [41] proposed a personalized and time-sensitive route planning method, in which they inferred users' preferences with locations and visiting time through historical data.

None of the aforementioned route planning methods considers human trust. In this paper, we aim to fill this gap by developing a trust-based route planning approach.

2.2 Trust in Automation

Trust in the context of human-technology relationships can be roughly classified into three categories: (1) *credentials-based*, which is used mainly in security and determines if a user can be trusted based on a set of credentials [25]; (2) *experience-based*, which includes reputation-based trust in peer-to-peer and e-commerce applications, determines an agent's trust value based on its own experience in predicting the probability of the execution of a certain action by another agent [30]; and (3) *cognitive trust*, which explicitly account for not only the human experience, but also subjective judgment about preferences and mental states [17]. In this paper, we are interested in human's trust in automated vehicles, and therefore consider cognitive trust that captures the human notion of trust. Specifically, we follow the notion of *trust in automation* proposed in [34], which indicates human's willingness to rely on automation.

Studies have found that human trust changes over time during the interaction with automation, affected by various factors such as the automation's reliability, predictability, and transparency [20, 38]. Studies have also shown that trust can influence human's reliance on automation and the system is likely to be under-utilized if human mistrust the automation [16]. For example, a recent study found that users with lower trust tended to take over control from automated vehicles more frequently [28]. Inspired by insights drawing from these prior studies, we develop a data-driven trust dynamics model to represent the evolution of human trust in automated vehicles, and a takeover decision model to associate the likelihood of human's takeover decision with trust.

Different method to measure trust have been proposed. User questionnaires are commonly used to evaluate the subjective belief of trust [36, 40]. For example, the study in [11] asked questions about users' trust in automated vehicles in a 7-point Likert scale. In addition, various sensing technologies have been used for the continuous measurement of human trust in real-time, including gaze tracking [22], gestures (e.g., face touching and arms crossed) [35], and biometrics (e.g., electroencephalogram and galvanic skin response) [23]. We measure human trust in a 7-point Likert scale via questionnaires in the online user study, and via continuous user control input (i.e., pressing buttons mounted on the steering wheel) in the driving simulator study.

Existing works about trust in automated vehicles include investigating factors that influence users' adoption of automated vehicles [32, 33, 39], studying the effect of alarm timing on driver's trust [5], designing forward collision warning system [29] and cruise control system [8] to improve users' trust. By contrast, this paper develops a route planning approach that accounts for trust to improve user experience of automated vehicles.

Several recent works have explored the idea of modeling trust with POMDPs. For example, a POMDP model for trust-workload dynamics in Level 2 driving automation was developed in [6], and a POMDP-based method for human-robot collaboration in table cleaning tasks was proposed in [10]. Our work is inspired by these methods. We focus on trust-based route planning for automated vehicles, which requires different POMDP modeling.

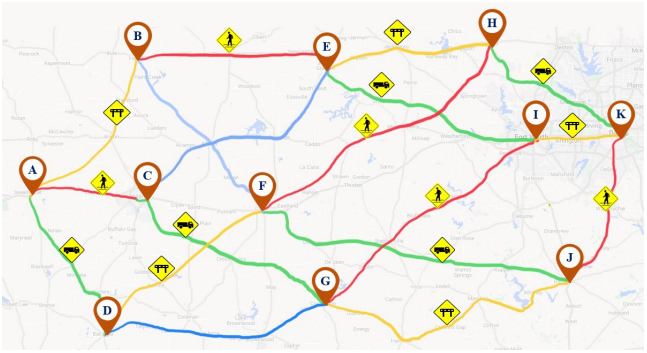


Figure 1: An example map with three types of road incidents (pedestrian, obstacle, and oncoming truck).

3 MOTIVATING EXAMPLE

We describe a motivating example of route planning for automated vehicles. Figure 1 shows an example map, where three types of typical incidents that may occur on the road are considered: (1) a pedestrian crossing the road, (2) an obstacle ahead of the lane, and (3) an oncoming truck in the neighboring lane. We can easily generalize to more complex examples with a richer set of incidents. For simplicity, we assume that each road segment may have up to one incident at a time. We also assume that the vehicle has information about the potential incident that it may encounter in the next road segment. Such information can be easily obtained, for example, via sensing and crowd sourcing traffic monitoring apps.

Figure 2 shows a schematic view of the automated vehicle traveling from one location to another. Suppose that the vehicle is approaching an incident in the autopilot mode. Due to safety concerns, the driver may decide to take over control of the vehicle and switch to manual driving. Such takeover decisions can be influenced by the driver's trust in the automated vehicle's capability of handling different types of incidents: the driver with lower trust is more likely to take over. In addition, the driver's trust evolves over time depending on the takeover decision and the vehicle's capability of handling an incident.

The goal of this work is to develop a trust-based route planning approach that computes an optimal route for the automated vehicle (e.g., navigating from A to K in the example map) while taking into account human trust dynamics and takeover decisions.

4 TRUST-BASED ROUTE PLANNING

We present a trust-based route planning approach for automated vehicles. The key idea is to model the human-vehicle interaction as a POMDP and compute the optimal vehicle route by solving the optimal policy of POMDP planning.

4.1 The Proposed POMDP Framework

Formally, a POMDP is denoted as a tuple $(S, A, \mathcal{T}, R, O, \delta, \gamma)$, where S is a finite of state, A is a set of actions, $\mathcal{T}$ is the transition function representing conditional transition probabilities between states, $R : S \times A \rightarrow \mathbb{R}$ is the real-valued reward function, O is a set of observations, δ is the observation function representing the conditional probabilities of observations given states and actions,

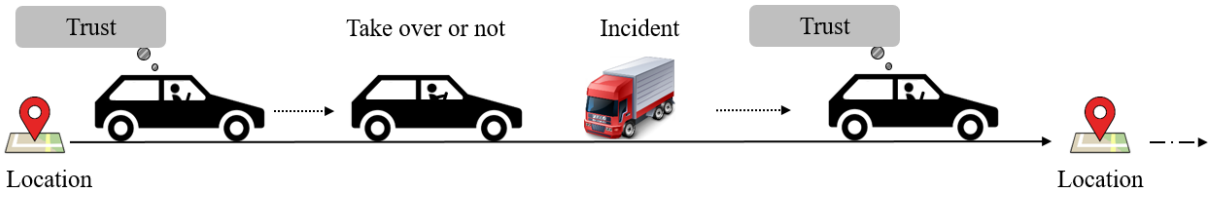


Figure 2: A schematic view of an automated vehicle navigating from one location to another. When approaching an incident, the driver may decide to take over and switch to manual driving. The takeover decision can be influenced by the driver's trust in the automated vehicle, which evolves over time.

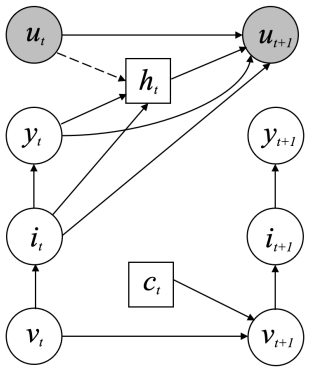


Figure 3: The POMDP graphical model for trust-based route planning. (Each node represents a state variable. Shadowed nodes are partially observable variables. Squares represent actions. Arrows represent transition functions.)

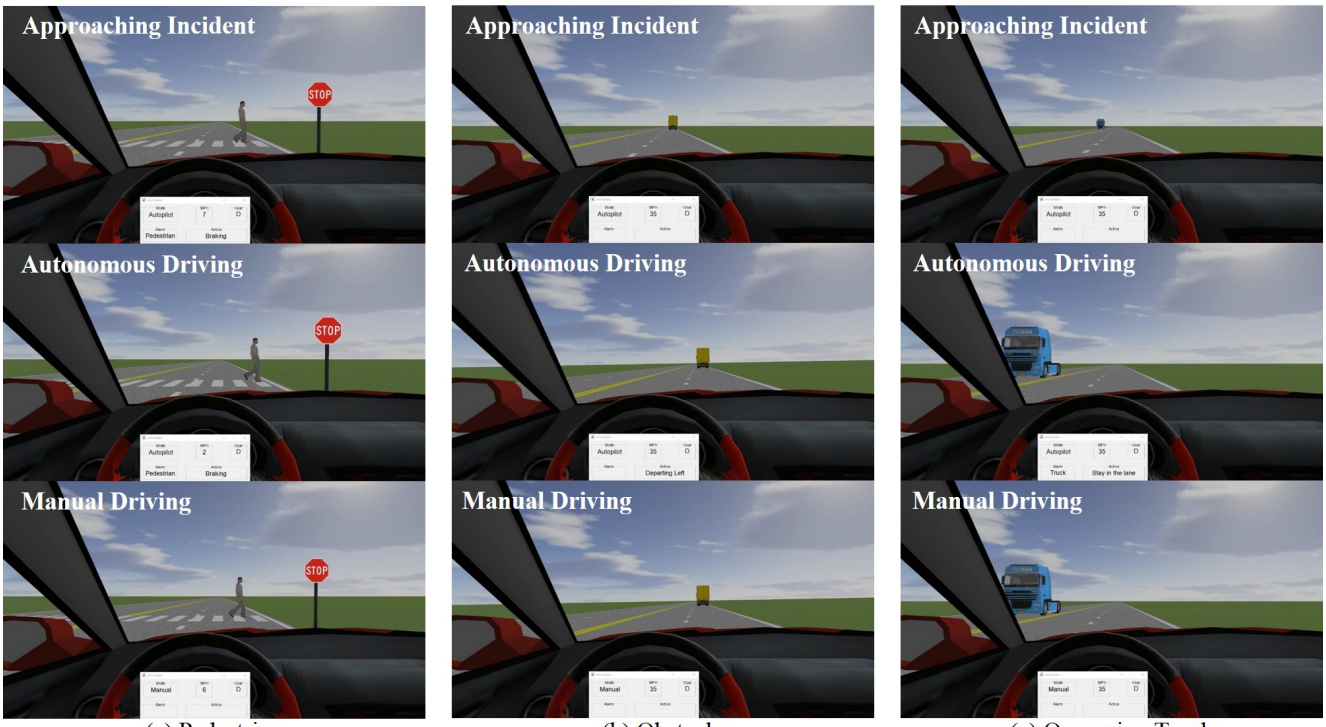
and $\gamma \in [0, 1]$ is the discount factor. At each time step t, given an action $a_t \in A$, a state $s_t \in S$ evolves to $s_{t+1} \in S$ with probability $\mathcal{T}(s_{t+1}|s_t, a_t)$. The agent receives a reward $R(s_t, a_t)$, and makes an observation $o_{t+1} \in O$ about the next state s_{t+1} with probability $\delta(o_{t+1}|s_{t+1}, a_t)$. The goal of POMDP planning is to compute the optimal policy that chooses actions to maximize the expectation of the cumulative reward $\mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)]$.

Figure 3 illustrates a graphical model of the proposed POMDP framework for trust-based route planning. We factor the state s_t at time t into four variables: v_t represents the vehicle position, i_t represents the road incident, y_t represents the automated vehicle's capability of safely handling the incident, and u_t is a partially observable variable representing human's trust in the automated vehicle (because trust is a hidden human mental state that cannot be directly observed by the vehicle agent). We factor the action a_t at time t into two variables: the vehicle route choice c_t and the human's takeover decision h_t. Given the vehicle's current position v_t and the route choice action c_t, we can determine the next vehicle position v_{t+1} by the transition function $\mathcal{T}(v_{t+1}|v_t, c_t)$. The potential incident i_t that the vehicle may encounter is determined by the vehicle position with probability $\mathcal{T}(i_t|v_t)$, and the automated vehicle's capability of safely handling the incident i_t is given by $\mathcal{T}(y_t|i_t)$. As discussed in Section 2, trust in automation can be influenced by many factors. Here, we model the evolution of trust dynamics with a probabilistic transition function $\mathcal{T}(u_{t+1}|u_t, y_t, i_t, h_t)$, based on a simplified assumption that trust evolves depending on the takeover decision and the vehicle's capability of handling an incident. The intuition is that trust may increase when the human chooses to not take over and witnesses the automated vehicle successfully handling an incident, and the trust may decrease if the automated vehicle fails to handle an incident.

The vehicle agent does not know about human's actual takeover action in advance, and it computes the optimal POMDP policy π^* of route choices c_t based on a model that predicts human's takeover decision h_t. We consider two different takeover decision models for comparison: (1) *trust-free model*, denoted by $\pi^h(h_t|i_t, y_t)$, where human decides whether to takeover depending on the incident and a fixed belief on the automated vehicle's capability to handle certain types of incidents; and (2) *trust-based model*, denoted by $\pi^h(h_t|i_t, y_t, u_t)$, where human make takeover decisions based on the incident and trust, indicating that human's belief on the automated vehicle's capability changes over time depending on the trust dynamics.

Considering the motivating example described in Section 3. The vehicle position v_t is one of the locations $\{A, B, \ldots, K\}$ shown in the map (Figure 1). The incident i_t can take one of the four values: null, pedestrian, obstacle, and truck. The vehicle's capability y_t of handling incidents has binary outcomes: success, and failure. Since human's trust is a partially observable variable u_t representing the hidden mental state, we use an observation variable $\hat{u}_t$ to represent the subjective trust in a 7-point Likert scale (1 and 7 indicate the lowest and highest levels of trust, respectively) measured via user questionnaires. The available route choices c_t are given by the map. For example, in location A, the vehicle may choose one of the three routes colored in yellow, red and green to navigate to B, C and D, respectively. The human takeover decision h_t is a binary choice of whether or not to take over control of the vehicle and resume manual driving. We can define the transition functions $\mathcal{T}(v_{t+1}|v_t, c_t)$ and $\mathcal{T}(i_t|v_t)$ based on the map. We can estimate $\mathcal{T}(y_t|i_t)$ based on the historical testing logs of the automated vehicle safely handling incidents. For the motivating example, we assume that the automated vehicle can always safely handle incidents (but the human driver has no prior knowledge about this assumption).

We design a reward function shown in Table 1 for the motivating example. Intuitively, we want to reward for better safety and user experience of automated vehicles. If the automated vehicle handles an incident successfully, we assign positive rewards based on the difficulty of driving tasks. When approaching a pedestrian incident, the automated vehicle needs to stop before the crosswalk and wait



(a) Pedestrian (b) Obstacle (c) Oncoming Truck

Figure 4: Screenshots of driving videos used in the online user study, covering three types of incidents: (a) a pedestrian crossing the road, (b) an obstacle (a stopped truck) ahead of the lane, (c) an oncoming truck in the neighboring lane. Each sub-figure shows: (top) the driver's view when the automated vehicle is approaching the incident, (middle) the view of autonomous driving if the driver chooses to not take over, (bottom) the view of manual driving if the driver chooses to take over.

Table 1: Reward function for the motivating example

	Pedestrian	Obstacle	Truck
Autopilot (Success)	3	2	1
Autopilot (Failure)	-9	-6	0
Manual driving	0	0	0

until the pedestrian crossing the road. When approaching an obstacle incident, the automated vehicle needs to perform lane changing in order to avoid collision with the obstacle. When there is an oncoming truck in the neighboring lane, the automated vehicle needs to keep driving in the same lane. Thus, we rank the pedestrian incident as the most difficult task and assign the highest reward value of 3, followed by the obstacle incident with a reward value of 2 and the truck incident with a reward value of 1. On the other hand, if the automated vehicle fails to handle an incident safely, we assign rewards based on the severity of incidents (e.g., striking a pedestrian can cause more serious damages than colliding with an obstacle). We assign zero reward to manual driving, because we want to promote better user experience and let the driver to enjoy non-driving tasks (e.g., reading or using mobile devices) in the automated vehicle. In addition, we assign a reward value of 5 to empty road (i.e., no incident thus no failure or takeover) to indicate this as the most favorable choice.

For the rest of this section, we describe the design of an online user study for data collection in Section 4.2; we present a data-driven method to estimate trust dynamics $\mathcal{T}(u_{t+1}|u_t, y_t, i_t, h_t)$ and the observation function $\delta(\hat{u}_t|u_t)$ in Section 4.3; we describe the data-driven modeling of trust-free takeover decision $\pi^h(h_t|i_t, y_t)$ and the trust-based takeover decision $\pi^h(h_t|i_t, y_t, u_t)$ in Section 4.4; and finally, we apply the proposed approach to the motivating example and present the computed optimal routes in Section 4.5.

4.2 Online User Study for Data Collection

We designed and conducted an online user study[1] with 100 anonymous participants on the Amazon Mechanical Turk platform. The objective of this study is to collect data about human's trust on automated vehicles. In particular, we investigated how trust evolves with respect to different incidents on the road and how human's takeover decisions are affected by incidents and trust. We created a set of driving videos using the PreScan driving simulation software [1]. Figure 4 shows the screenshots of example videos covering three types of incidents (i.e., pedestrian, obstacle, and oncoming truck) used in the motivating example.

[1]This study was approved by the Institutional Review Board at the University of Virginia.

During the online user study, we first established the baseline by asking participants about their trust in automated vehicles in a 7-point Likert scale (i.e., trust ranges from 1 to 7). Then, we showed a video of the automated vehicle approaching an incident on the road from the driver's view, and asked participants if they would like to takeover control of the vehicle and switch to manual driving, imaging that they were the driver siting inside the automated vehicle. Depending on the participant's response of takeover or not, we showed the next video of either the vehicle is driven autonomously or manually to handle the incident. After that, we asked participants to fill in a questionnaire which estimates their updated trust in automated vehicle. We adapted the Muir's questionnaire [37] and asked participants to answer the following questions in 7-point Likert scale:

(1) To what extent can you predict the automated vehicle's behavior from moment to moment?
(2) To what extent can you count on the automated vehicle to do its job?
(3) What degree of faith do you have that the automated vehicle will be able to cope with similar incidents in the future?
(4) Overall how much do you trust the automated vehicle?

We averaged a participant's responses to these four questions into a single rating between 1 and 7 to represent the participant's updated trust. We repeated the above process nine times (three times per incident type) with a randomized order of incidents.

We did not include any vehicle crash video in this study, because we assume that the automated vehicle is capable of handling all incidents safely. For example, the vehicle would automatically stop and wait for the pedestrian to cross the lane, or change the lane to avoid the obstacle. However, participants are not aware of such information in advance. They make takeover decisions based on their trust beliefs about the automated vehicle's capability to safely handle certain incident, and the trust levels may change based on their experience of watching prior incident videos.

The data we collected from each participant has the following format: $\mathcal{D} = \{\hat{u}_0, i_0, h_0, \hat{u}_1, \ldots, i_8, h_8, \hat{u}_9\}$, where $\hat{u}_t$ is the measured user trust, i_t is the incident type, h_t is the user decision of takeover or not, at each time step t. In order to guarantee the data quality, our study recruitment criteria required that participants must be able to read English fluently and have an above 95% approval rate on the Amazon Mechanical Turk platform. We also inserted questions for attention checks during the user study.

4.3 Data-Driven Trust Dynamics Model

As described in Section 4.1, the proposed POMDP framework for trust-based route planning represents human trust as a partially observable variable u_t at time step t, which evolves to u_{t+1} over time depending on human's takeover decision h_t and the automated vehicle's capability y_t to handle incident i_t. Using the data collected from the online user study described in Section 4.2, we model the trust dynamics and the POMDP observation function as a linear Gaussian system:

$$\mathcal{T}(u_{t+1}|u_t, y_t, i_t, h_t) = \mathcal{N}(\alpha_t u_t + \beta_t, \sigma_t^2)$$

$$\hat{u}_t \sim \mathcal{N}(u_t, \sigma_u^2)$$

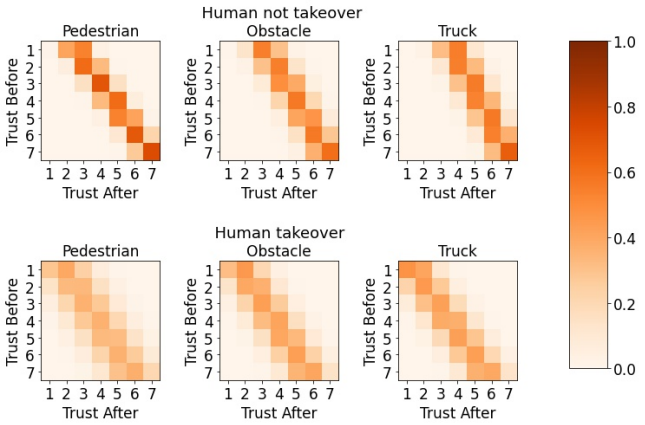


Figure 5: Visualization of probabilistic transition matrices of the learned trust dynamics model, where u_t and u_{t+1} are shown as trust before and trust after values ranging from 1 to 7, and each matrix corresponds to a pair of incident and takeover decision.

where $\mathcal{N}(\mu, \sigma^2)$ represents the Gaussian distribution with the mean μ and the variance σ; α_t and β_t are linear coefficients of trust dynamics given y_t, i_t and h_t; and $\hat{u}_t$ represents the observations of trust measured via subjective questionnaires in the online user study. We estimate these parameter values using full Bayesian inference with Hamiltonian Monte Carlo sampling algorithm [15].

Figure 5 illustrates a visualization of the learned trust dynamics model. There are six probabilistic transition matrices, corresponding to all combinations of three road incidents and binary human takeover decisions. Each transition matrix indicates the probability of changing from u_t (trust before value) to u_{t+1} (trust after value). We observe that trust values are more likely to increase when human decides to not take over (top row of Figure 5), while trust values tend to be constant or decrease when there is a takeover decision (bottom row of Figure 5). These observations are consistent with the insight from the prior studies (see Section 2) that takeover decisions are often correlated to trust.

4.4 Data-Driven Takeover Decision Models

In the POMDP framework, we use the variable h_t to denote human's takeover decisions (i.e., whether or not to take over control of the vehicle) when approaching an incident i_t at time step t. Such takeover decisions may also be influenced by human trust u_t. In the following, we present two takeover decision models based on whether or not to consider trust as an influencing factor.

Trust-free takeover decision model. Let b^i denote human's belief on the automated vehicle's capability of safely handling an incident i, which remains constant in the trust-free model. Let p_t denote the probability of human deciding to not take over at time step t. We define $p_t = \mathcal{S}(b^i r^{\mathrm{s},i} + (1 - b^i) r^{\mathrm{f},i})$, where $\mathcal{S}(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function, $r^{\mathrm{s},i}$ and $r^{\mathrm{f},i}$ are rewards of the automated vehicle handling the incident i with success and failure (see Table 1), respectively. We model the takeover decision with a Bernoulli distribution, denoted by $h_t \sim \mathcal{B}(p_t)$.

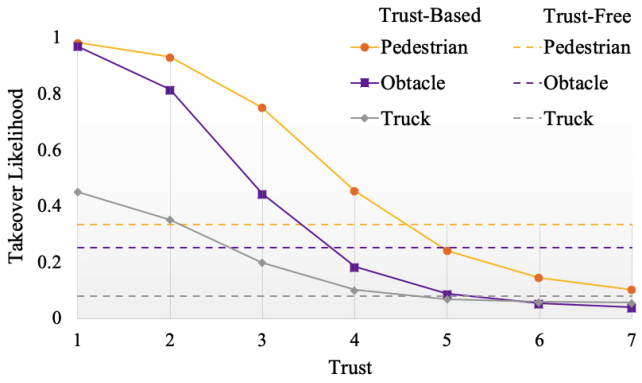


Figure 6: Predictions of takeover likelihood with respect to trust and incidents, using trust-based and trust-free takeover decision models.

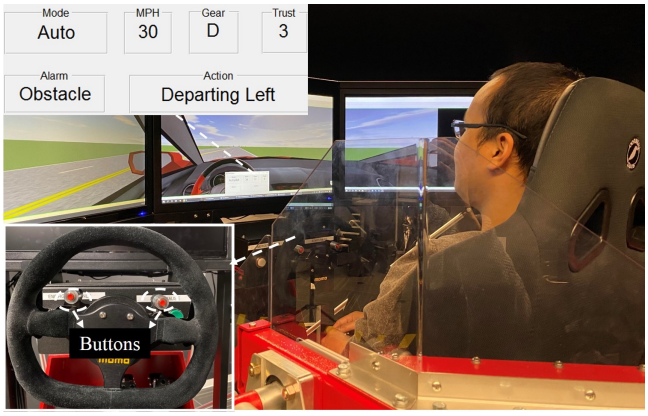


Figure 7: Driving simulator setup. The top zoomed-in view shows the GUI displaying the driver's current trust value, along with other information such as driving mode, velocity, gear, incident alarm, vehicle action. The bottom zoomed-in view shows the steering wheel with buttons for takeover commands and user trust input.

Trust-based takeover decision model. Let b_t^i denote human's belief on the automated vehicle's capability of safely handling an incident i at time step t, which evolves over time depending on the human trust u_t. Thus, we model the belief as a sigmoid function $b_t^i = \mathcal{S}(\kappa^i u_t + \lambda^i)$, where κ^i and λ^i are linear coefficients associated with the incident i. We assume that the human trust u_t follows a Gaussian distribution, denoted by $\hat{u}_t \sim \mathcal{N}(u_t, \sigma_u^2)$ where $\hat{u}_t$ are the measured trust values from the online user study. We define the probability of human deciding to not takeover as $p_t = \mathcal{S}(b_t^i r^{s,i} + (1 - b_t^i) r^{f,i})$, which is defined similarly to the trust-free model, but using the dynamic belief b_t^i instead of the constant b^i. Finally, the takeover decision is given by the Bernoulli distribution $h_t \sim \mathcal{B}(p_t)$.

Data-driven modeling results. We applied the full Bayesian inference with Hamiltonian Monte Carlo sampling algorithm [15] to estimate parameters in both the trust-free and trust-based models, using the data collected from the online user study. The results of log-likelihood show that the trust-based model (-359.37) fits better to the collected data as opposed to trust-free model (-446.83). The difference in log-likelihood results shows that accounting for trust in the takeover decision model can achieve better prediction performance, which supports our assumption that human takeover decisions is influenced by trust. Figure 6 shows model predictions of takeover probability with respect to trust and incidents. With the trust-free model, since the takeover decision does not depend on human trust, we observe three straight lines for three incidents. With the trust-based model, we observe the general trends of decreasing takeover likelihood with increasing trust, which are consistent with findings in the prior studies (see Section 2). Furthermore, we observe from the results of both models that it is more likely for human to decide to take over with riskier incidents: pedestrian with the highest takeover probability, followed by obstacle and truck.

4.5 Planning for the Motivating Example

We applied the Approximate POMDP Planning (APPL) Toolkit [2], which is an implementation of the point-based SARSOP algorithm for efficient POMDP planning [31], to compute the optimal policies of the proposed POMDP framework. For the motivating example, depending on the use of trust-based and trust-free takeover decision models, we obtained two optimal routes:

- trust-based route: A-D-G-J-K
- trust-free route: A-C-E-H-K

Note that the main difference between these two routes is the order of road incidents. In the trust-based route, the ordered incidents occurring in each road segment are oncoming truck (A-D), null (D-G), obstacle (G-J), and pedestrian (J-K). In the trust-free route, the incidents follows the order of pedestrian (A-C), null (C-E), obstacle (E-H), and oncoming truck (H-K). We evaluate and compare the performance of these two routes via human subject experiments[2] on a driving simulator, as described in the next section.

5 DRIVING SIMULATOR EXPERIMENTS

We describe the design, procedure, and results of our driving simulator experiments as follows.

5.1 Experiment Design

Apparatus. Figure 7 shows the driving simulator setup used for the experiments. The hardware platform is based on the Force Dynamics 401CR driving simulator, which is a four-axis motion platform that tilts and rotates to simulate the experience of being in a vehicle. The platform includes the seat, interlocked seat belt, interlocked doors, display screen, steering wheel, brake, paddle shifters, and throttle. There are two buttons mounted on the steering wheel (bottom zoomed-in view in Figure 7). We programmed the simulator's control input such that the driver can switch between automated and manual driving by pressing the two buttons simultaneously. In addition, we used the same set of buttons to measure participants' trust in automated vehicles during the experiments. The driver can

[2]This human subject study was approved by the Institutional Review Board at the University of Virginia.

press the left (*resp.* right) button to decrease (*resp.* increase) the trust value ranging from 1 to 7.

Driving scenario. We created a driving scenario based on the motivating example described in Section 3, using the PreScan driving simulation software [1]. We also programmed an autopilot controller for the simulated automated vehicle, which has the capability of leveraging the integrated sensors (e.g., radar, Lidar, and GPS) in PreScan for various driving tasks such as lane keeping, detecting and handling incidents.

Manipulated factor. We manipulate a single factor: the route that the autopilot controller follows. As stated in Section 4.5, the two conditions are: trust-based route and trust-free route.

Dependent measures. We are interested in studying the route which brings more cumulative reward. We recorded the participants' takeover decisions and calculated the cumulative POMDP reward using the reward function defined in Table 1.

Hypothesis. We hypothesize that participants taking the trust-based route can obtain higher cumulative POMDP rewards than those taking the trust-free route.

Subject allocation. We recruited 22 participants (average age: 23.7 years, SD=4.3 years, 31.8% female) from the university community. Each participant was compensated with a $20 gift card for completing the experiment. The recruitment criteria required all participants to have a valid driver license, at least one year of driving experience, and normal or corrected-to-normal vision. To avoid participants' bias, we adopted a between-subject study design: we randomly allocated 11 participants to take the trust-based route and the other 11 participants to experience the trust-free route.

5.2 Experiment Procedure

Upon arrival, a participant was instructed to read and sign a consent form approved by the Institutional Review Board. We conducted a five-minute training to help the participant get familiar with the driving simulator setup. Then, the participant was instructed to drive through the trust-based or trust-free route with the simulated automated vehicle, depending on the assigned study group. The journey started in the autopilot mode. When the vehicle approached an incident (i.e., pedestrian, obstacle, or truck), it alerted the participant by issuing an auditory alarm and displaying textual information about the incident type in the GUI. If the participant decided to not takeover, the vehicle would continue in the autopilot mode to handle the incident. The participant can take over control of the vehicle and switch to manual driving at any point during the experiment. If the participant did takeover, he was required to switch back to the autopilot mode after the vehicle passing that incident. We asked the participant to periodically record their trust in the automated vehicle using the buttons on the steering wheel (see bottom left in Figure 7). After the driving session, we asked the participant to answer the following survey questions in 7-point Likert scale (1 means strongly disagree, 4 is neural, 7 means strongly agree).

- Q1 I believe that the automated vehicle can get me to the destination safely.
- Q2 I find the route easy to drive.
- Q3 I find it easy to take over control of the automated vehicle.

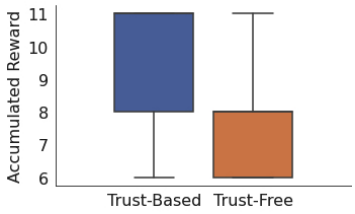


Figure 8: The cumulative rewards of participants taking trust-based and trust-free routes.

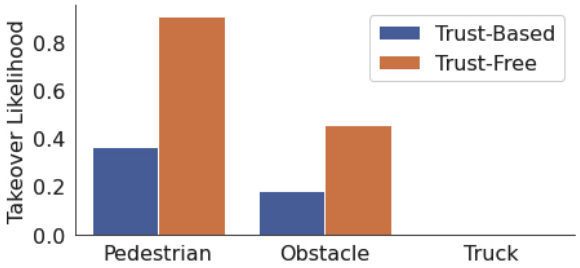


Figure 9: Participants' average takeover likelihood when the vehicle approaching different incidents in the trust-based and trust-free routes.

- Q4 I have concern about using the automated vehicle to drive through this route.
- Q5 I believe that the selected route is not dangerous.
- Q6 I think the selected route fits well with the way I would like to drive.
- Q7 I can depend on the reliability of the automated vehicle.

It took about 40 minutes for each participant to complete the entire experiment.

5.3 Results

We calculated the cumulative POMDP rewards (using the reward function defined in Table 1) for each participant, based on their takeover decisions when approaching incidents along the route. Figure 8 shows the box plot of cumulative rewards of all participants. We observe that participants taking the trust-based route tend to achieve higher cumulative rewards than participants taking the trust-free route, which is consistent with our study hypothesis. We also performed one-way analysis of variance (ANOVA) to evaluate this hypothesis, i.e, comparing the observed F-test statistics with $F(d_1, d_2)$ (F-distribution with between-group degree of freedom d_1 and within-group degree of freedom d_2). The observed statistics $F(1, 20) = 9.14$ is greater than the critical value at significance level 0.01. Thus, our study hypothesis is supported by ANOVA results statistically.

Figure 9 shows the average takeover likelihood of all participants, for different incidents along the two routes. It is not surprising to find that participants are more likely to take over in the trust-free route than the trust-based route. With both routes, participants have higher probabilities to take over when approaching a pedestrian than an obstacle, while none of them choose to take over the control

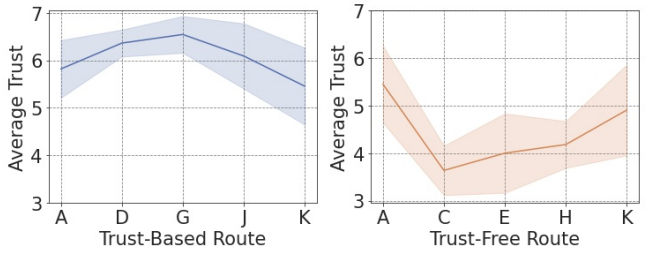


Figure 10: The evolution of participants' average trust along the trust-based and trust-free route. (The shadow represents the 95% confidence interval.)

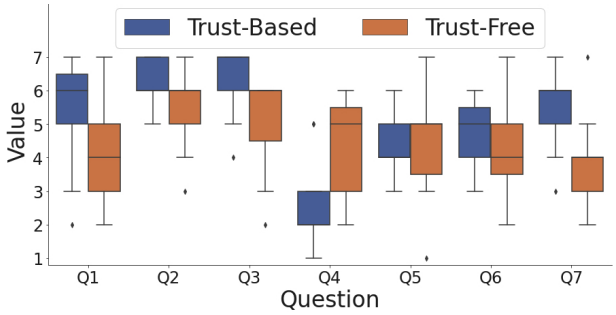


Figure 11: After-driving survey results. (Each box plot shows the maximum, the first quartile, the median, the third quartile, and the minimum. Each dot represents an outlier.)

when there was an oncoming truck in the neighboring lane. A possible explanation is that participants are more likely to take over when approaching incidents that are more challenging to handle or can cause more severe damages. These trends are consistent with the takeover predictions computed using the online user study data (see Figure 6).

Figure 10 shows how participants' average trust in the automated vehicle evolves as they were driving through different locations along the two routes. For the trust-based route, we observe that the average trust increases in the route segment A-D, this may result from the automated vehicle successfully handling the incident of oncoming truck in this segment. The trust continues to increase in the segment D-G, which is an empty road without any incident. However, the trust decreases in the next segment G-J where the vehicle needs to change lane to avoid an obstacle, and the trust further decreases in the last segment J-K where the vehicle needs to stop and wait for a pedestrian to cross the road. The decreasing of average trust may be explained by the occurring of more challenging and riskier incidents. For the trust-free route, we observe that the average trust drops sharply in the first route segment A-C with an pedestrian incident. However, the trust continues to increase slowly for the rest of the route. The average trust of participants taking the trust-based route is generally higher than taking the trust-free route.

Figure 11 summarizes the participants' responses to the after-driving survey questions. The results of Q1 indicate that participants experienced the trust-based route had higher belief in the automated vehicle's capability of driving safely than participants experienced the trust-free route. The results of Q2 show that participants found the trust-based route easier to drive than the trust-free route. The results of Q3 illustrate that participants driving through the trust-based route found it easier to take over control of the vehicle than those driving through the trust-free route. The results of Q4 show that participants experienced the trust-based route had less concern about the automated vehicle than those experienced the trust-free route. The results of Q5 indicate that participants tended to have a neutral opinion about how dangerous the routes are. The results of Q6 show that participants thought the trust-based route fits to the way they would like to drive better than the trust-free route in general. The results of Q7 find that participants driving through the trust-based route perceived higher reliability of the automated vehicle than those experienced the trust-free route. In summary, our human subject experimental results show that

- Participants taking the trust-based route generally resulted in higher cumulative POMDP rewards (where the reward function was designed to promote better safety and user experience of automated vehicles) than those taking the trust-free route.
- Participants were more likely to take over in the trust-free route than in the trust-based route; and riskier incidents led to higher takeover likelihood.
- Participants' trust in the automated vehicle evolved over time during the driving experience and was influenced by different types of incidents.
- Participants experienced the trust-based route had more positive responses in the after-driving survey than those driving through the trust-free route.

6 CONCLUSION

In this paper, we present a trust-based route planning approach for automated vehicles. We model the human-vehicle interaction as a POMDP and compute optimal routes for the vehicle by solving the POMDP planning. In order to incorporate trust into the route planning, we build data-driven models of trust dynamics and takeover decisions using data collected from an online user study with 100 participants on the Amazon Mechanical Turk platform. We applied the proposed trust-based route planning approach to a motivating example and obtained a trust-based route and a trust-free route (as a baseline for comparison). We evaluated these two routes via human subject experiments with 22 participants on a driving simulator. The results show that participants taking the trust-based route generally resulted in higher cumulative POMDP rewards (where the reward function was designed to promote better safety and user experience of automated vehicles), were less likely to take over control of the vehicle, and reported more positive responses in the after-driving survey than those taking the trust-free route. In addition, we observed that participants' trust changed over time during the study and was influenced by different road incidents. These observations are consistent with the findings of prior studies.

This work makes the first step towards incorporating human trust into route planning for automated vehicles. There are a few directions for future work. First, we would like to evaluate the scalability of the proposed approach. We believe that the proposed

POMDP-based approach can be applied to larger route planning problems (e.g., larger maps, more locations, and more route choices). However, the bottleneck lies in the evaluation. We will need to design and conduct new human subject experiments to evaluate the resulting routes of each problem, which can be costly and time consuming. Second, we would like to consider a richer set of incident types to reflect the complex road conditions that automated vehicles may encounter in the real-world. We will need to design and conduct new online user studies to collect data about trust in the automated vehicle's capability of safely handling these new incident types and build new data-driven trust dynamics model. Furthermore, we would like to explore the POMDP modeling of other factors that may influence human's trust in automated vehicles, such as the system transparency and predictability.

REFERENCES

[1] [n. d.]. PreScan Software, https://tass.plm.automation.siemens.com/prescan, Last Accessed: 2020-04-01.

[2] [n. d.]. Approximate POMDP Planning (APPL) Toolkit, https://github.com/AdaCompNUS/sarsop, Last Accessed: 2020-04-01.

[3] 2019. Tesla Model 3: Autopilot engaged during fatal crash. https://www.bbc.com/news/technology-48308852.

[4] 2020. Waymo's autonomous cars have driven 20 million miles on public roads. https://venturebeat.com/2020/01/06/waymos-autonomous-cars-have-driven-20-million-miles-on-public-roads/.

[5] Genya Abe and John Richardson. 2004. The effect of alarm timing on driver behaviour: an investigation of differences in driver trust and response to alarms according to alarm timing. *Transportation Research Part F: Traffic Psychology and Behaviour* 7, 4-5 (2004), 307–322.

[6] Kumar Akash, Neera Jain, and Teruhisa Misu. 2020. Toward Adaptive Trust Calibration for Level 2 Driving Automation. *arXiv preprint arXiv:2009.11890* (2020).

[7] Ove Andersen, Christian S Jensen, Kristian Torp, and Bin Yang. 2013. Ecotour: Reducing the environmental footprint of vehicles using eco-routes. In *2013 IEEE 14th International Conference on Mobile Data Management*, Vol. 1. IEEE, 338–340.

[8] Béatrice Cahour and Jean-François Forzy. 2009. Does projection into use improve trust and exploration? An example with a cruise control system. *Safety science* 47, 9 (2009), 1260–1270.

[9] Paolo Campigotto, Christian Rudloff, Maximilian Leodolter, and Dietmar Bauer. 2016. Personalized and situation-aware multimodal route recommendations: the FAVOUR algorithm. *IEEE Transactions on Intelligent Transportation Systems* 18, 1 (2016), 92–102.

[10] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. 2018. Planning with trust for human-robot collaboration. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 307–315.

[11] Jong Kyu Choi and Yong Gu Ji. 2015. Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction* 31, 10 (2015), 692–702.

[12] SAE On-Road Automated Vehicle Standards Committee et al. 2018. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. *SAE International: Warrendale, PA, USA* (2018).

[13] Jian Dai, Bin Yang, Chenjuan Guo, and Zhiming Ding. 2015. Personalized route recommendation using big trajectory data. In *2015 IEEE 31st international conference on data engineering*. IEEE, 543–554.

[14] Edsger W Dijkstra et al. 1959. A note on two problems in connexion with graphs. *Numerische mathematik* 1, 1 (1959), 269–271.

[15] Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. 1987. Hybrid monte carlo. *Physics letters B* 195, 2 (1987), 216–222.

[16] Mary T Dzindolet, Scott A Peterson, Regina A Pomranky, Linda G Pierce, and Hall P Beck. 2003. The role of trust in automation reliance. *International journal of human-computer studies* 58, 6 (2003), 697–718.

[17] Rino Falcone and Cristiano Castelfranchi. 2001. Social trust: A cognitive approach. In *Trust and deception in virtual societies*. Springer, 55–90.

[18] Daniel Gessner. 2020. Experts say we're decades from fully autonomous cars. Here's why. Jul. https://www.businessinsider.com/self-driving-cars-fully-autonomous-vehicles-future-prediction-timeline-2019-8.

[19] Hector Gonzalez, Jiawei Han, Xiaolei Li, Margaret Myslinska, and John Paul Sondag. 2007. Adaptive fastest path computation on a road network: a traffic mining approach. In *33rd International Conference on Very Large Data Bases, VLDB 2007*. Association for Computing Machinery, Inc, 794–805.

[20] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 5 (2011), 517–527.

[21] Peter E Hart, Nils J Nilsson, and Bertram Raphael. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics* 4, 2 (1968), 100–107.

[22] Sebastian Hergeth, Lutz Lorenz, Roman Vilimek, and Josef F Krems. 2016. Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human factors* 58, 3 (2016), 509–519.

[23] Wan-Lin Hu, Kumar Akash, Neera Jain, and Tahira Reid. 2016. Real-Time Sensing of Trust in Human-Machine Interactions. *IFAC-PapersOnLine* 49, 32 (2016), 48–53.

[24] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial intelligence* 101, 1-2 (1998), 99–134.

[25] Lalana Kagal, Tim Finin, and Anupam Joshi. 2001. Trust-based security in pervasive computing environments. *Computer* 34, 12 (2001), 154–157.

[26] Evangelos Kanoulas, Yang Du, Tian Xia, and Donghui Zhang. 2006. Finding fastest paths on a road network with speed patterns. In *22nd International Conference on Data Engineering (ICDE'06)*. IEEE, 10–10.

[27] Kanwaldeep Kaur and Giselle Rampersad. 2018. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. *Journal of Engineering and Technology Management* 48 (2018), 87–96.

[28] Moritz Körber, Eva Baseler, and Klaus Bengler. 2018. Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied ergonomics* 66 (2018), 18–31.

[29] Arnaud Koustanaï, Viola Cavallo, Patricia Delhomme, and Arnaud Mas. 2012. Simulator training with a forward collision warning system: Effects on driver-system interactions and driver trust. *Human factors* 54, 5 (2012), 709–721.

[30] Karl Krukow, Mogens Nielsen, and Vladimiro Sassone. 2008. Trust models in ubiquitous computing. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 366, 1881 (2008), 3781–3793.

[31] Hanna Kurniawati, David Hsu, and Wee Sun Lee. 2008. Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces.. In *Robotics: Science and systems*, Vol. 2008. Zurich, Switzerland.

[32] John D Lee and Kristin Kolodge. 2019. Exploring trust in self-driving vehicles through text analysis. *Human factors* (2019), 0018720819872672.

[33] John D Lee, Shu-Yuan Liu, Joshua Domeyer, and Azadeh DinparastDjadid. 2019. Assessing Drivers' Trust of Automated Vehicle Driving Styles With a Two-Part Mixed Model of Intervention Tendency and Magnitude. *Human factors* (2019), 0018720819880363.

[34] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[35] Jin Joo Lee, Brad Knox, and Cynthia Breazeal. 2011. *Modeling the Dynamics of Nonverbal Behavior on Interpersonal Trust for Human-Robot Interactions.* Ph.D. Dissertation. Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences.

[36] Nikolas Martelaro, Victoria C Nneji, Wendy Ju, and Pamela Hinds. 2016. Tell me more: Designing hri to encourage more trust, disclosure, and companionship. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 181–188.

[37] Bonnie Marlene Muir. 2002. Operators' trust in and use of automatic controllers in a supervisory process control task. (2002).

[38] Kristin E Schaefer, Jessie YC Chen, James L Szalma, and Peter A Hancock. 2016. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human factors* 58, 3 (2016), 377–400.

[39] Shili Sheng, Erfan Pakdamanian, Kyungtae Han, BaekGyu Kim, Prashant Tiwari, Inki Kim, and Lu Feng. 2019. A case study of trust on autonomous driving. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 4368–4373.

[40] Anqi Xu and Gregory Dudek. 2015. Optimo: Online probabilistic trust inference model for asymmetric human-robot collaborations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 221–228.

[41] Xiaoyan Zhu, Ripei Hao, Haotian Chi, and Xiaojiang Du. 2017. Fineroute: Personalized and time-aware route recommendation based on check-ins. *IEEE Transactions on Vehicular Technology* 66, 11 (2017), 10461–10469.