Exploring Factors Related to Drivers' Mental Model of and Trust in Advanced Driver Assistance Systems Using an ABN-Based Mixed Approach

Chunxi Huang, Jiyao Wang, Song Yan, and Dengbo He

Abstract—**Drivers' appropriate mental models of and trust in advanced driver assistance systems (ADAS) are essential to driving safety in vehicles with ADAS. Although several previous studies evaluated drivers' ADAS mental models of and trust in adaptive cruise control and lane-keeping assist systems, research gaps still exist. Specifically, recent developments in ADAS have made more advanced functions available but they have been under-investigated. Further, the widely adopted proportional correctness-based scores may not differentiate drivers' objective ADAS mental model and subjective bias towards the ADAS. Lastly, most previous studies adopted only regression models to explore the influential factors and thus may have ignored the underlying association among the factors. Therefore, our study aimed to explore drivers' mental models of and trust in emerging ADAS by using the sensitivity (i.e., *d'*) and response bias (i.e., *c*) measures from the signal detection theory. We modeled the data from 287 drivers using Additive Bayesian Network (ABN) and further interpreted the graph model using regression analysis. We found that different factors might be associated with drivers' objective knowledge of ADAS and subjective bias towards the existence of functions/limitations. Further, drivers' subjective bias was more associated with their trust in ADAS compared to objective knowledge. The findings from our study provide new insights into the influential factors on drivers' mental models of ADAS and better reveal how mental models can affect trust in ADAS. It also provides a case study on how the mixed approach with ABN and regression analysis can model observational data.**

Index Terms— **ADAS, Mental models, Trust, Factors, Additive Bayesian Network, Regression analysis**

This work was supported by the National Natural Science Foundation of China (Grant No. 52202425), the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone under Grant HZQB-KCZYB-2020083 and the Guangzhou Municipal Science and Technology Project (No. 2023A03J0011).

Chunxi Huang is with the Robotics and Autonomous Systems, Division of Emerging Interdisciplinary Areas (EMIA) under the Interdisciplinary Programs Office (IPO), The Hong Kong University of Science and Technology, Hong Kong SAR, China (e-mail: tracy.huang@connect.ust.hk).

Jiyao Wang and Song Yan are with the Intelligent Transportation Thrust, Systems Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China. (e-mail: jwang297@connect.hkust-gz.edu.cn and syan931@connect.hkust-gz.edu.cn).

Dengbo He is with the Intelligent Transportation Thrust, Systems Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, and HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen, China (e-mail: dengbohe@hkust-gz.edu.cn)

I. INTRODUCTION

In recent years, advanced driver assistance systems (ADAS), especially SAE Level 2 [1] ADAS, are becoming increasingly prevalent in modern vehicles [2], [3]. Traditionally, ADAS at the SAE Level 2 can provide longitudinal vehicle control via adaptive cruise control (ACC) or cruise control (CC), and lateral vehicle control via lane-keeping assistance (LKA) or lane centering control (LCC) [1].

However, before drivers are fully exempted from driving responsibilities (e.g., at SAE Level 5), drivers' understanding of driving automation is still essential to driving safety [4], which has been framed as the mental model in driving studies. The mental model was defined as "*a rich and elaborate structure, reflecting the user's understanding of what the system contains, how it works, and why it works that way*" [5]. In this study, drivers' ADAS mental models specifically refer to drivers' perceptions and understanding of the functions, limitations, and capabilities of the ADAS [6]. A growing body of research has shown that drivers' mental models of ADAS can influence their behavior, trust, and safety while driving. For example, drivers who have a better understanding of ADAS functions and limitations are more likely to use the automation appropriately and be prepared to intervene when needed [7]. In contrast, drivers who have inaccurate or incomplete mental models may over-rely on the ADAS [8], fail to notice or respond to critical events [9], and be less able to regain control of the vehicle in case of a system failure [10]. Further, the mental model of ADAS may influence drivers' trust in ADAS, which can further affect users' reliance on ADAS, leading to over-reliance if the trust is too high [11].

Given the importance of the mental model, previous studies have investigated influential factors of drivers' ADAS mental models using the survey-based method. For example, [12] evaluated users' understanding of the ADAS through "yes or no" responses to statements regarding ACC functions and limitations. Similarly, [13] used an online survey to evaluate drivers' understanding of ACC and LKA using 51 "true or false" questions regarding the functions and limitations of ACC/LKA. They found that drivers were not aware of the system's capabilities when they purchased the vehicle [12], and owning and using ACC did not result in a better understanding of the system [13]. Further, to account for users' uncertainties about their answers in the survey, rating scales were also used in several studies to allow participants to indicate their confidence in their responses [14]–[16]. They found that owners had a strong bias toward believing in the existence of ADAS capabilities [16]. Other studies also evaluated drivers' mental models in driving simulators [7], [17], [18]. For example, [18] found that drivers who had been exposed to edge-case events of ACC in driving simulators had better mental models of ACC. In [7], it was found that drivers with stronger mental models responded faster in edge-case situations compared to drivers with weaker mental models. Further, [17] found that although exposure to ADAS can improve drivers' mental models, training was still more effective in calibrating drivers' mental models. It should be noted that, although subjective bias may exist in survey studies, it can collect data from a larger sample size compared to the driving simulator studies and thus may have better generalizability.

However, previous studies mostly focused on traditional ADAS functions (i.e., ACC and LCC). Recent advancements in hardware and software have made new emerging ADAS functions beyond ACC and LCC available, including automated lane changing, automated overtaking, flexible speed control, automatic stop-and-go, road assistance, and driver monitoring & warning. Vehicle manufacturers have also started to bundle all ADAS functions into a single system without differentiating each single component to customers, such as Tesla's Navigate on Autopilot (NOA) [19] and XPeng's Navigation Guided Pilot (NGP) systems [20]. This brings new challenges to traffic safety, given that the emerging Level 2 ADAS still requires drivers to be responsible for the driving task and be prepared to take over control of the vehicle in case of emergencies, but take more tactical driving responsibilities (e.g., automatic lane changing) compared to traditional ADAS that only support operational tasks in vehicles (e.g., keeping constant speed).

Therefore, this study aims to explore the factors associated with drivers' mental models of and trust in emerging ADAS beyond ACC and LCC to facilitate appropriate usage of emerging ADAS. However, the percent-correctness score adopted in previous studies mixed users' objective knowledge of and subjective tendency towards believing whether a function/limitation exists and thus provides limited guidance on moderating users' mental models. Thus, in our study, we adopted d-prime (d') and criterion location (c) in signal detection theory (SDT) [21] to measure participants' sensitivity and response bias of their ADAS mental models. Further, to explore the influential factors of ADAS mental models, most previous research used linear regressions, which were not able to capture the structured relationships among the potential influential factors. To overcome this problem, a mixed approach combining Additive Bayesian Network (ABN) [22] and linear regressions was adopted in our study.

In light of the fast-growing market penetration of emerging ADAS functions in China [23] and the increasing number of ADAS-related accidents in China [24], the survey in our study targeted Chinese drivers. To the best of our knowledge, all existing studies focusing on ADAS mental models were conducted in North America (e.g., [13], [14], [16], [25], [26]) and Europe (e.g., [10], [12], [27], [28]). Given that cultural differences have been identified as a potentially influential factor in users' trust in automation [29], our study targeting towards Asian market may provide valuable insights into the development of tailored in-vehicle interfaces and driver education programs.

In summary, the contribution of this study is four-fold: (1) This is one of the first studies that investigated Chinese drivers' mental models of and trust in ADAS, which may guide the design of customized training programs from a cultural difference perspective of view; (2) This is also one of the first studies that targeted towards the emerging ADAS functions capable of handling tactical driving tasks; (3) This is the first study that adopted the SDT to quantify the subjective and objective components of ADAS mental models; (4) We introduced a novel ABN method to model the structures in questionnaire data, which may inspire future research.

II. Related Work

A. The Influence of ADAS Mental Models on Driving Safety

Drivers' appropriate understanding of when, how, and under what circumstances an ADAS can be used is essential to driving safety. In other words, drivers' mental models of ADAS (i.e., ADAS with only ACC and LCC or similar systems) were positively associated with driving safety in vehicles with Level-2 ADAS [7], [8], [14] and this is especially the case in situations where drivers were required to regain control of their vehicles [30]. For example, the research found that drivers who were unaware or uncertain of ACC limitations were more likely to use the automation in situations beyond the system's capabilities [8]. Another study found that drivers with better-calibrated mental models of ACC responded more quickly to edge-case situations (e.g., ACC failed to detect an approaching object in front of the vehicle) [14]. The worse performance in urgent scenarios can be attributed to deteriorated attention allocation strategies as a result of an inappropriate ADAS mental model. Specifically, drivers who were unaware of the limitations of ACC and LCC were found to pay less attention to the roadway and engaged more in non-driving related tasks, leading to decreased preparedness to take back control of the vehicle when necessary [9], [10], [31]. In a more recent driving simulator study [7], researchers manipulated participants' mental models by controlling the information participants received during training. They found that drivers' mental models of ADAS could predict their takeover performance, as measured by the mean absolute lateral position and standard deviation of the lateral position of the vehicle after a takeover.

However, it should be re-emphasized that the ADASs investigated in these previous studies had only ACC and LCC functions. Given the complexity introduced by the emerging ADAS functions these years, it is thus imperative to understand how well drivers understand these systems and what we can do to calibrate users' ADAS mental models.

B. The Association between Drivers' ADAS Mental Models and Drivers' Trust in ADAS

The link between the mental model and trust can potentially explain the impact of the mental model on driving safety. Trust in automation refers to "*an individual's attitude that an agent will assist them in achieving their objectives in a situation marked by uncertainty and vulnerability*" ([11], p. 54). Previous research revealed that users' trust in ADAS was associated with their ability to perceive the environment [32], respond to takeover events [33], and deal with hazards [34]. Drivers' trust in ADAS, on the other hand, can be affected by drivers' mental model of ADAS [11], [29]. For instance, a survey study conducted in North America found that for those who did not own a vehicle with ADAS, a better ADAS mental model was associated with a lower trust in the ADAS (i.e., ACC and LKA) [16].

However, it should be noted that the formation of trust in ADAS is a complex process. According to the framework in [29], human-automation trust variability originates from three layers, i.e., dispositional trust, situational trust, and learned trust. The mental model can be categorized into the layer of learned trust, but other factors from the dispositional (e.g., age and education) [35] and situational layers (e.g., traffic density) [36] may also influence drivers' trust in ADAS. Therefore, it is necessary to consider the moderating effects of these factors when exploring the relationship between the mental model and drivers' trust in ADAS.

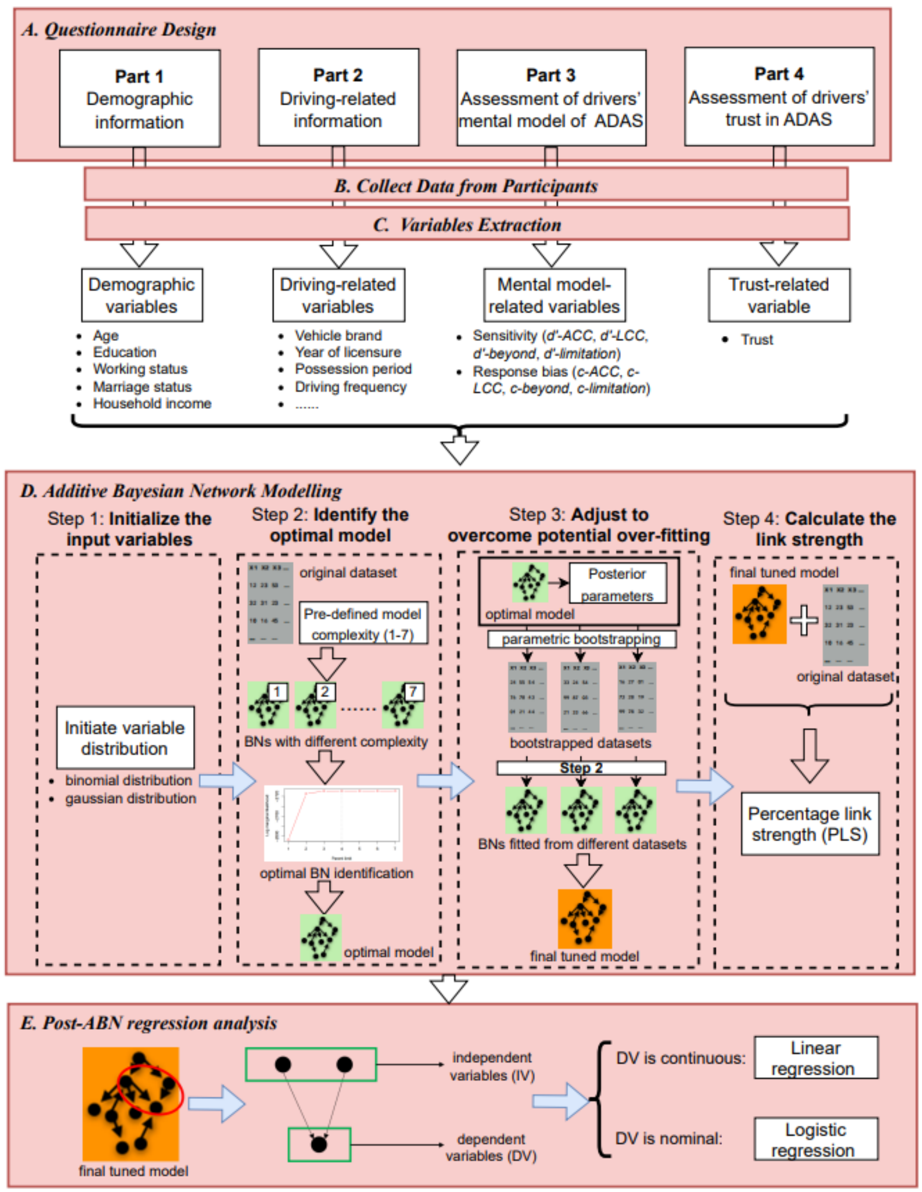

FIG. 1. THE OVERALL METHODOLOGICAL FRAMEWORK OF THIS STUDY

C. Drivers' Mental Models of Traditional ADAS

Despite the importance of drivers' ADAS mental models, research has shown that drivers often had less than ideal mental models of Level-2 ADAS. For example, two early survey studies in the 2010s found that only 42% [8] and 28% [37] of ACC users were aware of the limitations of ACC. Another survey found that 30% of Volvo XC60 owners were unaware of the limited system capabilities in curves and roundabouts [12]. Recent studies suggest that despite being in the market for decades, drivers' mental model of ACC did not improve. A survey study in the U.S. conducted in 2016 found that only 17% of respondents were able to correctly answer questions related to the ACC system [4], and another study in 2021 found that 47% of ACC owners held misperceptions of ACC systems [13]. Drivers' understanding of other emerging driving automation functions was even worse. For example, a survey study found that 81% of respondents were unaware of any limitations of the sensor-based backing aid system [38]. Another driving simulator study found that 20 out of 24 drivers mistakenly believed that the lane departure warning (LDW) system would work at any speed [39]. These findings suggest that more work is still needed to help calibrate users' mental models of ADAS. More importantly, the introduction of more recent and complex emerging ADAS functions may pose even greater challenges for drivers and thus it is urgent to assess drivers' mental models of emerging Level-2 ADAS.

III. Materials and methods

The overall methodological framework of this study is presented in **Fig. 1**. Overall, using the data collected from a survey study, an ABN model was built to explore the relationships among the factors of interest. The ABN was further tuned through parametric bootstrapping to avoid potential overfitting. Then, a post-ABN regression analysis was conducted to interpret the relationships among the targeted factors.

A. Questionnaire Design

A questionnaire was designed specifically to collect the data for this study. The questionnaire included four parts: (1) demographic questions; (2) driving-related questions; (3) assessment of drivers' mental model of ADAS; and (4) assessment of drivers' trust in ADAS:

1) Demographic information

Participants' demographic information was collected, including age, gender, education level, working status, marriage status, household income, and self-reported technology familiarity. Three questions adopted from [16], [40] were used to evaluate participants' self-reported technology familiarity, i.e., "*your level of experience with technology*," with possible responses ranging from 1 ("very inexperienced") to 10 ("very experienced"); "*the degree to which you consider yourself as an early adopter of technology*," with possible responses ranging from 1 ("absolutely no") to 10 ("absolutely yes"); and "*how easy you find it to learn new technology*," with possible responses ranging from 1 ("very difficult") to 10 ("very easy").

2) Driving-related questions

The driving-related information included participants' vehicle brand, possession period of their vehicle, years of licensure, driving frequency, weekly driving distance, ADAS experience, ADAS frequency, and ADAS familiarity. For ADAS familiarity, similar questions used for the assessment of technology familiarity were used, with the word 'technology' replaced by "ADAS".

3) Assessment of the ADAS mental model

A total of 49 statements were developed to evaluate the mental model of ADAS among drivers. These statements were developed based on a review of previous studies [4], [13], [16], [27], [28], [41], [42] and user manuals or official training materials from various automobile manufacturers. The statements were designed to assess the functions and limitations of Level-2 ADAS systems currently available in the market, including both traditional (i.e., ACC and LCC) and emerging ADAS functions. The term "ADAS", instead of the names of the sub-systems (e.g., ACC and LCC), was used throughout the questionnaire, as this is how vehicle manufacturers introduce ADAS to consumers. Following [43], to provide a more structured assessment of drivers' ADAS mental models, the statements used to assess drivers' ADAS mental models were categorized into four parts, each targeting one part of ADAS-related knowledge, i.e., ACC-related functions, LCC-related functions, functions beyond ACC & LCC, and ADAS limitations. Generally, the ACC-related functions and LCC-related functions parts included statements of what ACC can do (e.g., "*when you drive with ADAS on, the system can help you maintain a pre-set speed*") and LCC can do (e.g., "*when you drive with ADAS on, the system can help maintain the vehicle in the center of the lane*"). The functions beyond ACC & LCC part assessed participants' understanding of what the sub-systems beyond ACC and LCC can do (e.g., "*automatic lane changing will be triggered when the lead vehicle is too slow*"). The ADAS-limitations part included statements regarding the limitations and boundaries of ADAS (e.g., "*when you drive with ADAS on, the system may have difficulty when driving through construction zones*"). The mental-model-related questions and their grouping are provided in Appendix A. For 23 out of the 49 statements, real-world photos or videos were included to enhance understanding of the situations. Participants were instructed to rate their level of agreement with each statement on a scale of 1 to 6, with 1 indicating "*strongly disagree*" and 6 indicating "*strongly agree*".

Among all 49 statements, 34 of them can be treated as signals in SDT as they stated ADAS functions or limitations that indeed exist in the real world; while 15 of them can be treated as noises as they stated functions or limitations that do not exist. We further checked the assumptions of SDT, i.e., Gaussian distribution of the evidence when only noise is present, and when both signal and noise are present, and equal variance of the distributions [44]. Both assumptions were met. In this study, the questions designed for assessing drivers' mental model of ADAS were ordinal (i.e., level of agreement on the statement regarding the existence of a limitation or function), thus a fuzzy SDT [45] was adopted.

4) Assessment of drivers' trust in ADAS

To evaluate drivers' trust in ADAS, following [16], we utilized a five-item scale developed by [46]. Participants were asked to indicate their level of agreement on a five-point Likert scale, ranging from 1 ("*Strongly Disagree*") to 5 ("*Strongly Agree*"), for the following five statements, i.e., "*I can trust the ADAS system*," "*The ADAS system is reliable*," "*I am confident in the ADAS system*," "*I am familiar with the ADAS system*," and "*The ADAS system is dependable*."

B. Participants

To explore drivers' mental models of emerging ADAS, we specifically recruited participants among drivers of five vehicle brands (i.e., Tesla, Xpeng, NIO, LI, and WEY) that sold vehicles equipped with emerging ADAS functions in the Chinese market, via online posters in interest groups of car owners on social media (e.g., WeChat group of car owners). A total of 781 participants completed the questionnaire, of which 160 were excluded because they did not own a vehicle with Level-2 ADAS. Then, 268 samples were further excluded based on attention check questions and survey completion time, following previous studies [47], [48]. Lastly, 66 samples were excluded through manual review as they provided inconsistent or invalid information (e.g., the vehicle model does not match the vehicle brand), and 287 valid responses (262 males and 25 females) were kept. Each participant was compensated with 10 RMB for the valid response.

C. Variable Extraction

1) *Drivers'* ADAS mental model

For questions regarding the mental model, we scored drivers' responses by following the user manuals of the corresponding vehicle models (i.e., the vehicle model that a driver claimed to own) to avoid the influence of potential differences in ADAS functions or limitations between different vehicle models. As mentioned previously, we adopted the *d'* and *c* as measures of drivers' ADAS mental models. Given the non-binary responses in this study, we adopted the fuzzy SDT [45] as it can handle rating-scale-based responses without arbitrarily categorizing ratings into binary classes. In this study, we denoted the response from the participant for the i^{th} question as PR_i. Then, we calculated the four states of the world (i.e., hit, miss, false alarm, and correct rejection) using the equations proposed by [45]:

$$Hit = min\,(r, s) \quad (1)$$

$$Miss = max\,(s - r, 0) \quad (2)$$

$$False\ Alarm = max\,(r - s, 0) \quad (3)$$

$$Correct\ Rejection = min\,(1 - s, 1 - r) \quad (4)$$

where, $r = \frac{5 - |PR_i - 6|}{5}$ and stands for the probability of a positive response (i.e., the confidence of saying "yes, signal present"); while s represents the probability of a signal (which was 0 when the signal was absent and 1 when the signal was present). The minimum functions used in Equation (1) and (4) can be viewed as quantifying the degree of overlap between the fuzzy state of the real world (i.e., signal present or absent) and the respondent's fuzzy belief that the signal is present or absent. While the maximum function in Equation (2) and (3) can be interpreted as the degree of over-confidence (i.e., r > s, false alarm) or under-confidence (r < s, miss) in the presence of a signal. They are constrained to be positive by the maximum function; otherwise, the over-confidence becomes under-confidence, and vice versa.

Then the hit rate ($p(H)$) and false alarm rate ($p(FA)$) can be calculated as:

$$p\,(H) = Hits/(Hits + Misses) \quad (5)$$

$$p\,(FA) = False\ Alarms/(False\ Alarms + Correct\ Rejections) \quad (6)$$

Extreme hit rates and false alarm rates (i.e., 0 or 1) were adjusted following the methods suggested in [49]. The mean and 95% confidence intervals (CI) for $p(H)$ and $p(FA)$ of all groups calculated based on [50] are presented in Appendix B. Finally, the *d'* and *c* were obtained using following equations:

$$d' = z\,(p(FA)) - z(p(H)) \quad (7)$$

$$c = 0.5 * (z\,(p(H)) + z(p(FA))) \quad (8)$$

in which $z\,()$ is the function that transforms the probability into z-scores. The calculation was conducted using the *psycho* package in R [51]. For each part of the ADAS functions, we obtained the *d'* and *c* of the corresponding mental model, i.e., *d'-ACC* and *c-ACC* for ACC-related functions; *d'-LCC*, *c-LCC* for LCC-related functions; *d'-beyond* and *c-beyond* for functions beyond ACC & LCC; and *d'-limitation* and *c-limitation* for ADAS limitations.

2) Drivers' trust in ADAS

Drivers' trust in ADAS was calculated as the average ratings of five trust-related questions in the questionnaires following [46]: the higher the rating, the higher the trust.

3) Other variables

Following [52], we calculated participants' technology familiarity and ADAS familiarity by averaging their responses to the corresponding questions. It should be noted that gender was not included as a factor as the collected data was highly imbalanced in terms of gender. All variables are presented in **Table I** along with their distributions.

TABLE I
QUESTIONS IN THE SURVEY, EXTRACTED VARIABLES, AND THE DISTRIBUTION OF THE VARIABLES

Part	Description	Variable [Type]	Distributions
	The age (in years of old) of the participant.	Age[C]	M = 29.9 (SD: 6.1) Min: 20, Max: 58
	The highest education level of the participant.	Education[O]	• Professional college or less (n=86) • Bachelor or above (n=201)

Demographic questions	The working status of the participant.	Working status[N]	• Full-time work (n=153) • Others (n=134)
	The marriage status of the participant.	Marriage status[N]	• Married (n=173) • Not married (n=114)
	The household income of the participant (in RMB).	Household income[O]	• < 250,000 (n=143) • ≥ 250,000 (n=144)
	Participant's self-reported familiarity with technology.	Technology familiarity[C]	M = 7.2 (SD: 1.3) Min: 4.3, Max: 10
Driving-related questions	The brand of the vehicle that the participant owned.	Vehicle brand[N]	• Tesla (n=141) • Others (n=146)
	The duration of possession of the current vehicle.	Possession period[O]	• < 1 year (n=95) • ≥ 1 year (n=192)
	For many years since participants obtained their first driver's license.	Years of licensure[C]	M = 7.4 (SD: 4.4) Min: 1, Max: 28
	Participant's self-reported driving frequency in the past year.	Driving frequency[O]	• Almost every day (n=148) • Other (less frequent) (n=139)
	Participant's self-reported weekly average driving distance in the past year.	Weekly driving distance[O]	• < 99 km (n=122) • ≥ 99 km (n=165)
	Participant's self-reported experience of emerging ADAS.	ADAS experience[O]	• < 6 months (n=150) • ≥ 6 months (n=137)
	Participant's self-reported frequency of using emerging ADAS.	ADAS frequency[O]	• < several times per week (n=99) • ≥ several times per week (n=188)
	Participant's self-reported familiarity with emerging ADAS.	ADAS familiarity[C]	M = 8.2 (SD: 1.3) Min: 4.3, Max: 10
Drivers' mental model of ADAS	The sensitivity of participant's mental model of ACC functions.	*d'-ACC*[C]	M = 0.4 (SD: 0.5), Min: -1.2, Max: 2.1 95%CI: [0.34, 0.46]
	The response bias of participant's mental model of ACC functions.	*c-ACC*[C]	M = -1.1 (SD: 0.4), Min: -1.4, Max: 0.6 95%CI: [-1.15, -1.05]
	The sensitivity of participant's mental model of LCC functions.	*d'-LCC*[C]	M= 0.3 (SD: 0.6), Min: -1.3, Max: 2.5 95%CI: [0.23, 0.37]
	The response bias of participant's mental model of LCC functions.	*c-LCC*[C]	M = -0.9 (SD: 0.5), Min = -1.2, Max: 0.9 95%CI: [-0.96, -0.84]
	The sensitivity of participant's mental model of functions beyond ACC and LCC.	*d'-beyond*[C]	M = 0.2 (SD: 0.4), Min: -1.2, Max: 1.7 95%CI: [0.15, 0.25]
	The response bias of participant's mental model of functions beyond ACC and LCC.	*c-beyond*[C]	M = -1.1 (SD: 0.6), Min: -1.6, Max: 0.8 95%CI: [-1.17, -1.03]
	The sensitivity of participant's mental model of ADAS limitations.	*d'-limitation*[C]	M = 0.4 (SD: 0.7), Min: -2.5, Max: 2.2 95%CI: [0.32, 0.48]
	The response bias of participant's mental model of ADAS limitations.	*c-limitation*[C]	M = -0.8 (SD: 0.7), Min: -1.4, Max: 1.4 95%CI: [-0.88, -0.72]
Drivers' trust in ADAS	Participant's self-reported trust in emerging ADAS.	Trust[C]	M = 4.1 (SD: 0.7) Min: 2, Max: 5

Note: C represents the continuous variable; N represents the nominal variable; O represents the ordinal variable. M stands for the mean; SD stands for standard deviation; 95%CI stands for 95% confidence intervals.

D. Additive Bayesian Network (ABN) Modelling

A mixed approach with ABN and regression analysis was used to explore the influential factors of drivers' mental models of emerging ADAS. Being different from most classical model selection techniques that focus on a single dependent variable, ABN provides a tool for modeling multivariate relationships among variables of interest. Thus, with ABN, we were able to model the relationships among all variables without causing multicollinearity issues.

All Bayesian Networks (BN) models consist of two mutually dependent components: a qualitative part (a pre-defined or data-driven structure) and a quantitative part (the learned parameters). The model structure was determined by a directed acyclic graph (DAG), which serves as a graphical representation of the joint probability distribution factorization among all random variables. In the DAG, each node represents a random variable, and the directed arcs denote the probabilistic dependencies between variables. Traditionally, the BN can only handle categorical variables and requires the discretization of the continuous variables if there are. Considering that we have continuous variables in the survey, in our study, an ABN was adopted [22]. As a special type of BN, the ABN model can be regarded as a graphical model that bundles multiple generalized linear models (GLM) [53], in which, each sub-structure represents a GLM, with the variable (or node) itself being used as the dependent variable and the variables (or nodes) directed to it as independent variables.

The ABN offers greater flexibility compared to traditional BN, as ABN can incorporate variables with various types of distributions, such as countable, categorical, and continuous variables. Further, unlike methods such as structural equation modeling [54], which relies on expert knowledge to determine an optimal model structure, ABN modeling is a data-driven approach [53], although expert knowledge can still be incorporated if needed. In our study, we employed a Bayesian approach for both structure discovery and parameter learning. Specifically, a uniform structural prior was applied (i.e., all eligible DAG structures were weighted equally in the absence of data), and uninformative priors were used for all parameters at each arrow (i.e., all values were equally probable in the absence of data). The ABN modelling in this study was conducted using R [51] and JAGS [55], which included the following four consecutive steps:

Step 1: Initialize the input variables

Each variable was assigned a specific probability distribution based on its type: binomial distribution for binary variables and Gaussian distribution for quantitative ones.

Step 2: Identify the optimal model

To identify the optimal model, we varied the maximum number of connected nodes allowed for each node, starting from one to seven. A heuristic search-based method [56] was adopted, i.e., gradually increasing the number of allowable connected nodes from one to seven. The highest goodness-of-fit was achieved when we reached the largest network score with the specific number of allowable connected nodes [22]. The R package '*abn*' (version 2.7-3) [22] was used in this step to identify the optimal model.

Step 3: Adjust to overcome potential overfitting

As BN is prone to overfitting the data [57], in our study, a parametric bootstrapping method using Markov Chain Monte Carlo (MCMC) simulations was adopted to overcome the potential overfitting of our model [58]. As shown in **Fig. 1**, first, the posterior parameters were estimated from the optimal model structure using Laplace approximations. Then, based on the optimal model structure obtained in Step 2 as well as the estimated posterior parameters, we used MCMC samplers (i.e., JAGS) to generate 10,000 bootstrap datasets. Each generated bootstrap dataset had the same size as the original dataset. Then, we performed a model search (i.e., repeating Step 2) on all generated bootstrap datasets, leading to 10,000 bootstrap ABN structures in total. Thus, we were able to calculate the frequency of occurrence for each arc in the bootstrap ABN structure. Only the arcs that occurred in over 50% of the bootstrap ABN structure were kept in the final pruned BN model.

Step 4: Calculate the link strength

Link strength is a valuable tool for both visualization and approximate inference, which could be regarded as a proxy for arc uncertainty. The link strength is particularly advantageous in BN modelling, as the conventional metric for assessing significance in frequentist statistics (i.e., p-value) is not appropriate for BN. Following [22], we used the percentage link strength (PLS) as the link strength metric. The PLS between variable X and Y is defined as the percentage reduction in uncertainty of variable Y given the state of X if the states of all other variables directed to Y are known, as expressed below:

$$PLS\ (X \rightarrow Y) = \frac{H(Y|Z) - H(Y|X,Z)}{H\ (Y|Z)} \qquad (9)$$

where Z are the remaining variables directed to Y (excluding X) and H is the entropy computed using the empirical distribution of the random variable. The link strength is set to zero if there is no arc connecting X and Y.

In this study, we constructed one graph model that includes all demographic variables (e.g., age, education), driving-related variables (e.g., vehicle brand, possession period), trust, and sensitivity and response bias of all four parts of the mental model (i.e., *d'-ACC, c-ACC, d'-LCC, c-LCC, d'-beyond, c-beyond, d'-limitation, c-limitation*). At the same time, as the main focus of this study is to explore factors associated with drivers' mental model of and trust in different ADAS functions, the associations across sensitivity and response bias related to different parts of the ADAS mental model were banned (e.g., the link between *d'-ACC* and *d'-LCC* was banned) in the ABN modelling process following [59].

E. Post-ABN Regression Analysis

The ABN structure was unable to directly inform the correlation between connected variables. Specifically, the link strength, as visualized as the width of the arcs in ABN, can only inform the uncertainty reduction of one variable if we know the value/level of another variable connected by an arc, but not the linear association among variables; the latter, however, is of more interest if we aim to explore the influential factors of specific variable of interests. Therefore, we conducted post-ABN regression analyses for all arcs in the graph model outputted by ABN.

As shown in **Fig. 1**, for each variable being directed by other variables in an ABN model, we fitted a regression model with the variable itself as the dependent variable and all variables directing to it as independent variables. More specifically, we fitted linear regression models when the dependent variable was continuous and binomial logistic regression models when the dependent variable was categorical. Besides, in the model with 2 or more independent variables, we considered potential two-way interaction effects between independent variables and only the statistically significant (p<.05) interaction effects were reported. It should be noted that the regression models for marriage status, household income, and education were omitted here, as they were not connected to the variables of interests (i.e., *d'*, c, and trust) directly or indirectly. All regression analysis and visualization in this study were conducted in R [51].

IV. RESULTS

A. ABN Modelling Analysis

For better visualization, we visualized the association among variables and *d'* & *c* for each part of the mental model separately, leading to four BN structures, i.e., **Fig. 2** for ACC-related functions**, Fig. 3** for LCC-related functions**, Fig. 4** for functions beyond ACC & LCC**,** and **Fig. 5** for ADAS limitations. In all BNs, rectangles represent binary nodes, and ovals represent continuous nodes. Further, demographic-related variables are highlighted in blue, and driving-related variables are in green. The percentage values in **Fig. 2 - Fig. 5** are the PLS of each link and the width of the arrows is proportional to the PLS between two nodes.

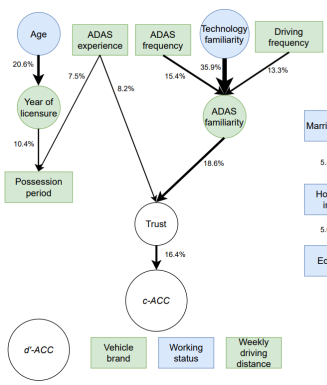

FIG. 2. THE FINAL DAG FOR ACC PART

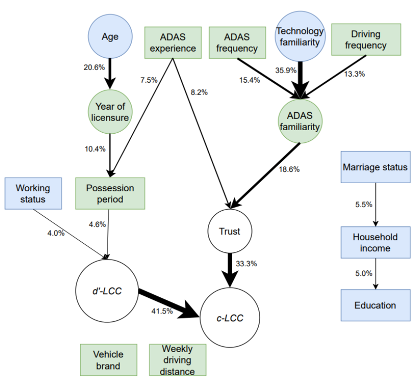

FIG. 3. THE FINAL DAG FOR LCC PART

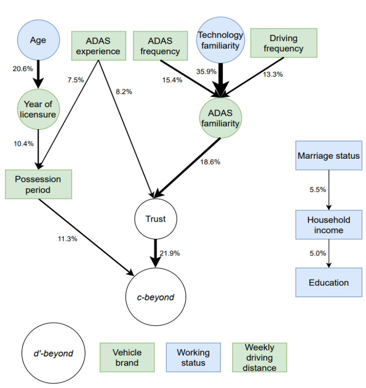

FIG. 4. THE FINAL DAG FOR BEYOND PART

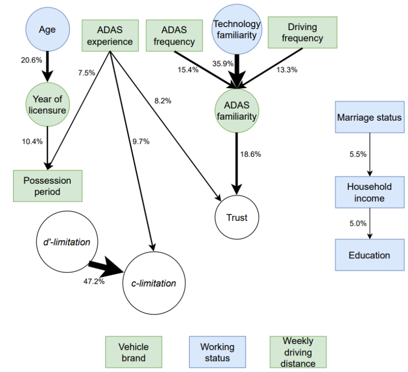

FIG. 5. THE FINAL DAG FOR LIMITATION PART

B. Post-ABN Regression Analysis

Based on the four DAGs from the ABN analysis, we fitted 9 regression models, including 8 linear regression models (with years of licensure, ADAS familiarity, trust, c-ACC, d'-LCC, c-LCC, c-beyond, and c-limitation as dependent variables) and 1 logistic regression model (with possession period as the dependent variable). Post-hoc contrasts were conducted if the independent variables were significant ($p < .05$) in the model. Results of all fitted models along with the significant post-hoc contrasts are as follows:

1) Year of licensure

It was found that the age was associated with the years of licensure, $F(1,285) = 135.77$, $p < .0001$. Each year of increase in age led to a 0.42-year (95%CI: [0.35, 0.49]) increase in the years of licensure.

2) Possession period

We found that both ADAS experience ($\chi^2(1) = 15.38$, $p < .0001$) and years of licensure ($\chi^2(1) = 14.45$, $p = .0001$) were associated with the possession period. Specifically, drivers with more ADAS experience (i.e., ≥ 6 months) had a higher probability of having a possession period over 1 year, with an odds ratio (OR) of 2.93 (95%CI: [1.71, 5.01]). Further, every 1-year increase in the years of licensure led to a 0.89 multiplicative decrease in the odds of a possession period over 1 year, with a 95% confidence interval (CI) of [0.84, 0.94].

3) ADAS familiarity

It was found that drivers' ADAS frequency ($F(1,283) = 13.12$, $p = .0003$), driving frequency ($F(1,283) = 15.68$, $p < .0001$), and technology familiarity ($F(1,283) = 59.63$, $p < .0001$) were all significant predictors of drivers' ADAS familiarity. Specifically, each 1-unit increase in technology familiarity led to a 0.35-unit (95%CI: [0.25, 0.45]) increase in the ADAS familiarity. Drivers with higher driving frequencies (i.e., almost every day) had higher ADAS familiarity compared to those who drove less frequently ($\Delta = 0.46$, 95%CI: [0.19, 0.73]). Drivers with higher ADAS frequencies (i.e., ⩾ several times per week) had higher ADAS familiarity compared to those who used ADAS less frequently ($\Delta = 0.50$, 95%CI: [0.22, 0.78]).

4) Trust

ADAS experience ($F(1,284) = 23.44$, $p < .0001$) and ADAS familiarity ($F(1,284) = 50.86$, $p < .0001$) both significantly affected drivers' trust in ADAS. Specifically, drivers with ADAS experience less than 6 months had lower trust in ADAS compared to drivers with more ADAS experience ($\Delta = -0.23$, 95%CI: [-0.38, -0.08]). Each 1-unit increase in ADAS familiarity led to a 0.21-unit (95%CI: [0.15, 0.27]) increase in the trust.

5) c-ACC

Drivers' trust in ADAS was significantly associated with drivers' response bias of ACC functions (i.e., *c-ACC*), F(1,285) = 42.15, $p < .0001$. Each 1-unit increase in trust was associated with a 0.23-unit (95%CI: [0.16, 0.30]) decrease in the c-ACC.

6) d'-LCC

Drivers' working status (F(1,284) = 16.61, $p < .0001$) and possession period (F(1,284) = 20.65, $p < .0001$) were significantly associated with their sensitivity of LCC functions (i.e., *d'-LCC*). Drivers with a possession period of less than 1 year had higher *d'-LCC* compared to drivers with a longer possession period ($\Delta = 0.35$, 95%CI: [0.20, 0.50]). Drivers who had full-time work had higher *d'-LCC* compared to drivers with other working status ($\Delta = 0.27$, 95%CI: [0.13, 0.41]).

7) c-LCC

Drivers' trust (F(1,284) = 26.46, $p < .0001$) and *d'-LCC* (F(1,284) = 101.13, $p < .0001$) were significantly associated with drivers' *c-LCC*. Each 1-unit increase in trust was associated with a 0.20-unit (95%CI: [0.12, 0.28]) decrease in the *c-LCC*. Each 1-unit increase in *d'-LCC* was associated with a 0.38-unit (95%CI: [0.30, 0.46]) increase in the *c-LCC*.

8) c-beyond

Drivers' trust (F(1,284) = 71.75, $p < .0001$) and possession period (F(1,284) = 14.27, $p < .0001$) were significantly associated with drivers' *c-beyond*. Each 1-unit increase in trust was associated with a 0.34-unit (95%CI: [0.26, 0.42]) decrease in the *c-beyond*. Drivers with a possession period of less than 1 year had higher *c-beyond* compared to drivers with longer possession periods ($\Delta = 0.24$, 95%CI: [0.12, 0.36]).

9) c-limitation

Drivers' ADAS experience (F(1,284) = 9.21, $p = .003$) and *d'-limitation* (F(1,284) = 62.69, $p < .0001$) were significantly associated with drivers' *c-limitation*. Each 1-unit increase in *d'-limitation* was associated with a 0.42-unit (95%CI: [0.31, 0.53]) decrease in the *c-beyond*. Drivers with ADAS experience of less than 6 months had higher *c-beyond* compared to drivers with more ADAS experience ($\Delta = 0.22$, 95%CI: [0.07, 0.37]).

V. Discussion

A. Factors Related to Drivers' ADAS Mental Model

First, no demographic factors nor driving-related factors considered in this study were found to be associated with drivers' objective knowledge of ACC functions (i.e., *d'-ACC*), functions beyond ACC & LCC (i.e., *d'-beyond*), and limitations of ADAS (i.e., *d'-limitation*); a shorter possession period (i.e., less than 1 year) was even associated with higher *d'-LCC*. This is surprising but somewhat echoes the findings in [13], which stated that users of traditional ADAS did not have better knowledge of ADAS even compared to non-ADAS users. Our results expanded this finding to emerging ADAS users. This is potentially because while driving, drivers rarely encounter ADAS failures [42] and have few chances to correct their ADAS mental model. In contrast, their perceived reliability of the ADAS may potentially have reinforced their "first impression" of ADAS. This explanation can be partially supported by the positive correlation between ADAS experience and trust in ADAS, both observed in our study and also found in many previous studies (e.g., [27]).

Further, the ABN suggested factors that are potentially associated with the possession period, which may help us better understand the relationships between the possession period and users' objective knowledge of LCC functions (i.e., *d'-LCC*). We found that more ADAS experience and a shorter year of licensure were associated with a longer possession period of the ADAS. The former is straightforward – those who owned their current vehicle (with ADAS) for a longer period are those who had more experience with ADAS. The association between the year of licensure and the possession period, however, is counter-intuitive. A possible explanation is that the drivers with shorter years of licensure are relatively younger and tend to be earlier adopters of new technologies [60], which may indirectly explain the negative association between the possession period and the *d'-LCC*, i.e., earlier technology adopters may seek information regarding new technologies more actively.

Further, it is interesting to notice that the *d'* and *c* of LCC function and ADAS limitation are coupled, while *d'* and *c* are uncoupled in ACC and beyond parts. Theoretically, the *d'* and *c* should be decoupled in SDT [61]. However, the participants may adjust their response bias when they have different levels of knowledge of a system. In our case, knowing better about the ADAS (i.e., higher *d'*), the users were more conservative regarding the ADAS capabilities (i.e., more conservative regarding the existence of LCC functions, and more likely to believe the existence of ADAS limitations). In other words, drivers tend to overestimate what the system can do if they know little about the system. This is alerting. Given that drivers have few chances to calibrate their mental model when using the ADAS in daily life (as ADAS failures are rare [62]), they may have less than ideal ADAS mental model and thus may tend to over-estimate the capabilities of the systems, which may further confirm their inaccurate ADAS mental model. The association between the *c-beyond* and the possession period has supported this "vicious circle" - drivers with a longer possession period had lower *c-beyond* compared to drivers with a shorter possession period, indicating that drivers become more inclined to believe the existence of ADAS functions after they gain more experience of the system. However, this association between the *d'* and *c* has not been observed for ACC and functions beyond ACC & LCC (potentially because drivers know ACC well and were very unfamiliar with functions beyond ACC & LCC [43]) – future research is needed to investigate how subjective and objective parts of the mental model can interact with each other and further influence drivers' behaviors when using the ADAS.

B. Factors Related to Drivers' Trust in ADAS

We further investigated the influential factors of trust. We found that objective experience with ADAS (i.e., ADAS experience and ADAS frequency), self-confidence in knowing ADAS (i.e., ADAS familiarity), and the technology in general (i.e., technology familiarity) can increase users' trust in ADAS, although the influences of some factors are indirect (i.e., technology familiarity is positively associated with trust via ADAS familiarity).

Further, the sensitivity (*d'*) of any parts of the ADAS was not associated with trust; but trust in ADAS was negatively associated with response bias of ADAS functions. Previous studies found an association between mental models and trust [16]. Our results are in line with this finding but provide a higher resolution. Specifically, trust in ADAS was more closely related to users' belief in the existence of the functions, but less with users' objective knowledge of the systems (i.e., *d'*), though as mentioned previously, the *d'* and *c* can be coupled.

By differentiating the ADAS functions and limitations, we further found that trust was associated with response bias of ADAS functions (i.e., *c-ACC*, *c-LCC*, and *c-beyond*) but not with limitations (i.e., *c-limitation*). It seems that trust was positively related to users' tendency to believe in what the ADAS can do (existence of the functions), but not what the ADAS cannot do (limitations of the systems). Previous research found that drivers' trust in ADAS may drop dramatically after experiencing unexplained ADAS failures [63], potentially because experiencing failure can influence one's ADAS mental model [27]. Our results, together with the previous finding regarding the relationship between ADAS failure and trust, provide a higher resolution on the relationship between trust and the ADAS mental model.

C. Limitations

It should be noted that this study only explored limited factors related to drivers' mental model of and trust in ADAS, future research may take other factors (e.g., advertising strategies for different vehicle brands) into consideration, which could provide more insights on how drivers' mental model of and trust in ADAS are formed. Further, it should be noted that neither ABN nor regression analysis was able to inform the causality between the variables. For example, it is also possible that those who trusted more in ADAS tended to use the ADAS more frequently, or higher trust in ADAS led to lower *c*, instead of the other way around. Experiments with better-controlled variables are needed to investigate the directions of the relationships observed in our study so that certain strategies can be proposed to calibrate drivers' mental models of and trust in ADAS. It is also important to note that we used a limited number of questions in the survey, although the Fuzzy SDT can potentially increase the resolution of the calculation, the estimation of d' and c can still be biased to some extent [64]. Future research with a larger number of questions is needed to further validate our conclusions.

Finally, although our study targeted an under-investigated population, the Chinese ADAS users, we could not compare the ADAS mental models of Chinese users in this study with other previous studies, as different questions were used to assess ADAS mental models. However, we compared Chinese drivers' trust in ADAS with other studies that used the same scale for measuring trust, and we found that Chinese drivers generally had higher trust (Mean = 4.1, SD = 0.7) in ADAS compared to drivers in the U.S. (Mean = 3.67, SD = 0.80) [65] and Canada (Mean = 3.4, SD = 0.8) [16]. Different cultures and different advertising strategies of vehicle manufacturers may lead to the difference in trust in ADAS among drivers from different countries, which may further affect their use of ADAS.

VI. Conclusions

Utilizing 287 drivers' responses obtained from a survey study, we investigated the factors that are associated with drivers' mental models of and trust in emerging ADAS beyond the traditional ADAS (i.e., ACC and LCC). Being different from prior research that relied on proportional correctness-based scores to evaluate drivers' mental models of ADAS, our study adopted the *d'* and *c* from signal detection theory and provided a more nuanced assessment of drivers' objective knowledge and subjective bias towards ADAS. Further, using a mixed approach combining ABN and linear regression, we found that users' objective knowledge of ADAS (*d'*) and inclination to believe the existence of ADAS functions and limitations (*c*) might be affected by different factors. In general, drivers' objective knowledge of ADAS mental models may not improve with accumulated experience with the ADAS; instead, drivers even became more inclined to believe in the existence of ADAS functions when they have more experience. Moreover, we revealed the underlying mechanisms of how the mental model can influence users' trust in ADAS – the objective knowledge might have limited influence on trust, but users' bias towards the existence of ADAS capabilities (not limitations) might play a more direct role. These findings highlight the imperative need to provide training to improve drivers' ADAS mental models and calibrate their trust in ADAS [66]. Compared to ADAS limitations, the training targeting ADAS functions should be designed to avoid drivers being over-confident in the ADAS, for example, using responsibility-focused strategies [67].

Acknowledgment

We would like to thank all participants who participated in the survey. This work was supported by the National Natural Science Foundation of China (Grant No. 52202425), the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083), and the Guangzhou Municipal Science and Technology Project (No. 2023A03J0011).

The authors would also like to express sincere thanks to Prof. Reinhard Furrer for discussing the methodology.

References

[1] SAE International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016_202104)," 2021. https://www.sae.org/standards/content/j3016_202104/ (accessed Feb. 17, 2022).

[2] V. A. Banks and N. A. Stanton, "Keep the driver in control: Automating automobiles of the future," *Appl. Ergon.*, vol. 53, pp. 389–395, Mar. 2016, doi: 10.1016/j.apergo.2015.06.020.

[3] T. Litman, "Autonomous vehicle implementation predictions: Implications for transport planning," 2020.

[4] A. B. McDonald, D. V. McGehee, S. T. Chrysler, N. M. Askelson, L. S. Angell, and B. D. Seppelt, "National Survey Identifying Gaps in Consumer Knowledge of Advanced Vehicle Safety Systems," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2559, no. 1, pp. 1–6, Jan. 2016, doi: 10.3141/2559-01.

[5] J. M. Carroll and J. R. Olson, "Mental models in human-computer interaction," *Handb. human-computer Interact.*, pp. 45–65, 1988.

[6] W. B. Rouse and N. M. Morris, "On looking into the black box: Prospects and limits in the search for mental models.," *Psychol. Bull.*, vol. 100, no. 3, p. 349, 1986.

[7] R. Rossi, M. Gastaldi, F. Biondi, F. Orsini, G. De Cet, and C. Mulatti, "A Driving Simulator Study Exploring the Effect of Different Mental Models on ADAS System Effectiveness," 2020, pp. 102–113. doi: 10.1007/978-3-030-58465-8_7.

[8] D. A. Dickie and L. N. Boyle, "Drivers' understanding of adaptive cruise control limitations," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2009, vol. 53, no. 23, pp. 1806–1810.

[9] A. M. Noble, M. Miles, M. A. Perez, F. Guo, and S. G. Klauer, "Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems," *Accid. Anal. Prev.*, vol. 151, p. 105959, Mar. 2021, doi: 10.1016/j.aap.2020.105959.

[10] J. C. F. de Winter, R. Happee, M. H. Martens, and N. A. Stanton, "Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 27, pp. 196–217, Nov. 2014, doi: 10.1016/j.trf.2014.06.016.

[11] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 46, no. 1, pp. 50–80, Jan. 2004, doi: 10.1518/hfes.46.1.50_30392.

[12] A. F. L. Larsson, "Driver usage and understanding of adaptive cruise control," *Appl. Ergon.*, vol. 43, no. 3, pp. 501–506, May 2012, doi: 10.1016/j.apergo.2011.08.005.

[13] C. A. DeGuzman and B. Donmez, "Drivers Still Have Limited Knowledge About Adaptive Cruise Control Even When They Own the System," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2675, no. 10, pp. 328–339, Oct. 2021, doi: 10.1177/03611981211011482.

[14] J. G. Gaspar, C. Carney, E. Shull, and W. J. Horrey, "Mapping drivers' mental models of adaptive cruise control to performance," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 81, pp. 622–638, Aug. 2021, doi: 10.1016/j.trf.2021.07.012.

[15] C. Carney, J. G. Gaspar, and W. J. Horrey, "Longer-term exposure vs training: Their effect on drivers' mental models of ADAS technology," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 91, pp. 329–345, Nov. 2022, doi: 10.1016/j.trf.2022.09.017.

[16] C. A. DeGuzman and B. Donmez, "Knowledge of and trust in advanced driver assistance systems," *Accid. Anal. Prev.*, vol. 156, p. 106121, Jun. 2021, doi: 10.1016/j.aap.2021.106121.

[17] G. Pai *et al.*, "Frequency and Quality of Exposure to Adaptive Cruise Control and Impact on Trust, Workload, and Mental Models," *Accid. Anal. Prev.*, vol. 190, p. 107130, Sep. 2023, doi: 10.1016/j.aap.2023.107130.

[18] M. Zahabi, A. M. A. Razak, R. K. Mehta, and M. Manser, "Effect of advanced driver-assistance system trainings on driver workload, knowledge, and trust," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 76, pp. 309–320, Jan. 2021, doi: 10.1016/j.trf.2020.12.003.

[19] Tesla, "Navigate on Autopilot," 2020. https://www.tesla.com/autopilot (accessed Dec. 01, 2022).

[20] XPeng Motors, "XPeng Unveils Beta Version Navigation Guided Pilot Function," 2021.

[21] H. Stanislaw and N. Todorov, "Calculation of signal detection theory measures," *Behav. Res. methods, instruments, Comput.*, vol. 31, no. 1, pp. 137–149, 1999.

[22] G. Kratzer, F. Lewis, A. Comin, M. Pittavino, and R. Furrer, "Additive Bayesian Network Modeling with the R Package abn," *J. Stat. Softw.*, vol. 105, no. 8, 2023, doi: 10.18637/jss.v105.i08.

[23] Ministry of Industry and Information Technology, "China's L2 ADAS market penetration rate reached 30%," 2022. https://baijiahao.baidu.com/s?id=1739975916659839317&wfr=spider&for=pc (accessed Dec. 01, 2022).

[24] Autoevolution, "Xpeng P7 on NGP Doesn't Detect a Truck, Crashes in China," 2021. https://www.autoevolution.com/news/xpeng-p7-on-ngp-doesn-t-detect-a-truck-crashes-in-china-170224.html (accessed Dec. 01, 2022).

[25] P. M. Greenwood, J. K. Lenneman, and C. L. Baldwin, "Advanced driver assistance systems (ADAS): Demographics, preferred sources of information, and accuracy of ADAS knowledge," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 86, pp. 131–150, Apr. 2022, doi: 10.1016/j.trf.2021.08.006.

[26] A. P. Hungund and A. K. Pradhan, "A survey on knowledge and perceptions of advanced driver assistance systems in Massachusetts drivers," *Traffic Inj. Prev.*, pp. 1–3, Oct. 2022, doi: 10.1080/15389588.2022.2125228.

[27] M. Beggiato, M. Pereira, T. Petzoldt, and J. Krems, "Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 35, pp. 75–84, Nov. 2015, doi: 10.1016/j.trf.2015.10.005.

[28] M. Beggiato and J. F. Krems, "The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 18, pp. 47–57, May 2013, doi: 10.1016/j.trf.2012.12.006.

[29] K. A. Hoff and M. Bashir, "Trust in Automation," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 57, no. 3, pp. 407–434, May 2015, doi: 10.1177/0018720814547570.

[30] I. Solís-Marcos, C. Ahlström, and K. Kircher, "Performance of an Additional Task During Level 2 Automated Driving: An On-Road Study Comparing Drivers With and Without Experience With Partial Automation," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 60, no. 6, pp. 778–792, Sep. 2018, doi: 10.1177/0018720818773636.

[31] O. Carsten, F. C. H. Lai, Y. Barnard, A. H. Jamson, and N. Merat, "Control task substitution in semiautomated driving: Does it matter what aspects are automated?," *Hum. Factors*, vol. 54, no. 5, pp. 747–761, 2012.

[32] J. Radlmayr, C. Gold, L. Lorenz, M. Farid, and K. Bengler, "How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving," in *Proceedings of the human factors and ergonomics society annual meeting*, 2014, vol. 58, no. 1, pp. 2063–2067.

[33] S. Hergeth, L. Lorenz, R. Vilimek, and J. F. Krems, "Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving," *Hum. Factors*, vol. 58, no. 3, pp. 509–519, 2016.

[34] B. D. Seppelt and J. D. Lee, "Making adaptive cruise control (ACC) limits visible," *Int. J. Hum. Comput. Stud.*, vol. 65, no. 3, pp. 192–205, 2007.

[35] C. Gold, M. Körber, C. Hohenberger, D. Lechner, and K. Bengler, "Trust in automation–before and after the experience of take-over scenarios in a highly automated vehicle," *Procedia Manuf.*, vol. 3, pp. 3025–3032, 2015.

[36] M. Li, B. E. Holthausen, R. E. Stuck, and B. N. Walker, "No risk no trust: Investigating perceived risk in highly automated driving," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2019, pp. 177–185.

[37] J. W. Jenness, N. D. Lerner, S. Mazor, J. S. Osberg, and B. C. Tefft, "Use of advanced in-vehicle technology by young and older early adopters," *Surv. results Adapt. cruise Control Syst. Rep. no. DOT HS*, vol. 810, p. 917, 2008.

[38] J. W. Jenness, N. D. Lerner, S. D. Mazor, J. S. Osberg, and B. C. Tefft, "Use of advanced in-vehicle technology by young and older early adopters. survey results on sensor-based backing aid systems and rear-view video cameras," United States. Department of Transportation. National Highway Traffic Safety …, 2007.

[39] T. Aziz, Y. Horiguchi, and T. Sawaragi, "An empirical investigation of the development of driver's mental model of a lane departure warning system while driving," *IFAC Proc. Vol.*, vol. 46, no. 15, pp. 461–468, 2013.

[40] H.-Y. W. Chen, B. Donmez, L. Hoekstra-Atwood, and S. Marulanda, "Self-reported engagement in driver distraction: An application of the Theory of Planned Behaviour," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 38, pp. 151–163, Apr. 2016, doi: 10.1016/j.trf.2016.02.003.

[41] H. Abraham *et al.*, “Autonomous vehicles and alternatives to driving: trust, preferences, and effects of age,” in *Proceedings of the transportation research board 96th annual meeting*, 2017, pp. 8–12.

[42] A. McDonald, C. Carney, and D. V McGehee, “Vehicle owners’ experiences with and reactions to advanced driver assistance systems,” 2018.

[43] C. Huang, D. He, X. Wen, and S. Yan, “Beyond adaptive cruise control and lane centering control: drivers’ mental model of and trust in emerging ADAS technologies,” *Front. Psychol.*, vol. 14, Aug. 2023, doi: 10.3389/fpsyg.2023.1236062.

[44] C. D. Wickens, W. S. Helton, J. G. Hollands, and S. Banbury, *Engineering psychology and human performance*. Routledge, 2021.

[45] R. Parasuraman, A. J. Masalonis, and P. A. Hancock, “Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance,” *Hum. Factors*, vol. 42, no. 4, pp. 636–659, 2000.

[46] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, “Foundations for an Empirically Determined Scale of Trust in Automated Systems,” *Int. J. Cogn. Ergon.*, vol. 4, no. 1, pp. 53–71, Mar. 2000, doi: 10.1207/S15327566IJCE0401_04.

[47] J. Ayoub, X. J. Yang, and F. Zhou, “Modeling dispositional and initial learned trust in automated vehicles with predictability and explainability,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 77, pp. 102–116, Feb. 2021, doi: 10.1016/j.trf.2020.12.015.

[48] M. M. Rahman, L. Strawderman, M. F. Lesch, W. J. Horrey, K. Babski-Reeves, and T. Garrison, “Modelling driver acceptance of driver support systems,” *Accid. Anal. Prev.*, vol. 121, pp. 134–147, Dec. 2018, doi: 10.1016/j.aap.2018.08.028.

[49] M. J. Hautus, “Corrections for extreme proportions and their biasing effects on estimated values ofd′,” *Behav. Res. Methods, Instruments, Comput.*, vol. 27, no. 1, pp. 46–51, Mar. 1995, doi: 10.3758/BF03203619.

[50] M. J. Hautus, N. A. Macmillan, and C. D. Creelman, *Detection theory: A user’s guide*. Routledge, 2021.

[51] R Development Core Team, “R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.,” *R Foundation for Statistical Computing*. 2019.

[52] H.-Y. W. Chen and B. Donmez, “What drives technology-based distractions? A structural equation model on social-psychological factors of technology-based driver distraction engagement,” *Accid. Anal. Prev.*, vol. 91, pp. 166–174, Jun. 2016, doi: 10.1016/j.aap.2015.08.015.

[53] F. I. Lewis and M. P. Ward, “Improving epidemiologic data analyses through multivariate regression modelling,” *Emerg. Themes Epidemiol.*, vol. 10, no. 1, p. 4, Dec. 2013, doi: 10.1186/1742-7622-10-4.

[54] J. F. Hair, “Multivariate data analysis,” 2009.

[55] K. Hornik, F. Leisch, A. Zeileis, and M. Plummer, “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling,” in *Proceedings of DSC*, 2003, vol. 2, no. 1.

[56] M. Scutari, “Learning Bayesian Networks with the bnlearn R Package,” *J. Stat. Softw.*, vol. 35, no. 3, 2010, doi: 10.18637/jss.v035.i03.

[57] M. A. Babyak, “What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models,” *Psychosom. Med.*, vol. 66, no. 3, pp. 411–421, 2004.

[58] N. Friedman, M. Goldszmidt, and A. Wyner, “Data Analysis with Bayesian Networks: A Bootstrap Approach,” Jan. 2013, [Online]. Available: http://arxiv.org/abs/1301.6695

[59] A. Comin, A. Jeremiasson, G. Kratzer, and L. Keeling, “Revealing the structure of the associations between housing system, facilities, management and welfare of commercial laying hens using Additive Bayesian Networks,” *Prev. Vet. Med.*, vol. 164, pp. 23–32, Mar. 2019, doi: 10.1016/j.prevetmed.2019.01.004.

[60] R. L. Huang *et al.*, “Features and predictors of problematic internet use in Chinese college students,” *Behav. Inf. Technol.*, vol. 28, no. 5, pp. 485–490, 2009, doi: 10.1080/01449290701485801.

[61] D. M. Green and J. A. Swets, *Signal detection theory and psychophysics*, vol. 1. Wiley New York, 1966.

[62] S. Shen and D. M. Neyens, “Assessing drivers’ response during automated driver support system failures with non-driving tasks,” *J. Safety Res.*, vol. 61, pp. 149–155, Jun. 2017, doi: 10.1016/j.jsr.2017.02.009.

[63] A.-K. Kraft, C. Maag, M. I. Cruz, M. Baumann, and A. Neukum, “Effects of explaining system failures during maneuver coordination while driving manual or automated,” *Accid. Anal. Prev.*, vol. 148, p. 105839, Dec. 2020, doi: 10.1016/j.aap.2020.105839.

[64] J. Miller, “The sampling distribution of d′,” *Percept. Psychophys.*, vol. 58, no. 1, pp. 65–72, 1996.

[65] D. G. Kidd, J. B. Cicchino, I. J. Reagan, and L. B. Kerfoot, “Driver trust in five driver assistance technologies following real-world use in four production vehicles,” *Traffic Inj. Prev.*, vol. 18, no. sup1, pp. S44–S50, May 2017, doi: 10.1080/15389588.2017.1297532.

[66] H.-Y. W. Chen, Z. Guo, and M. Ebnali, “Is text-based user manual enough? A driving simulator study of three training paradigms for conditionally automated driving,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 95, pp. 355–368, May 2023, doi: 10.1016/j.trf.2023.05.006.

[67] C. A. DeGuzman, S. Ayas, and B. Donmez, “Limitation-Focused versus Responsibility-Focused Advanced Driver Assistance Systems Training: A Thematic Analysis of Driver Opinions,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2677, no. 7, pp. 122–132, Jul. 2023, doi: 10.1177/03611981221151022.

Chunxi Huang received his bachelor’s degree in industrial engineering from Zhejiang University, Hangzhou, China in 2018 and his master’s degree in industrial and systems engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2020. He is currently a Ph.D. candidate at the Hong Kong University of Science and Technology. His research interests include human factors, driver behavior, and traffic safety.

Jiyao Wang received the B.Eng. degree in Software Engineering from the Sichuan University, Chengdu, China in 2021, and M.Sc. degree in Big Data Technology from the Hong Kong University of Science and Technology, Hong Kong S.A.R., China, in 2022. Currently, he is a Ph.D. student in the Intelligent Transportation Thrust at the Hong Kong University of Science and Technology (Guangzhou).

Song Yan received his bachelor’s degree in automotive engineering from Zhejiang University, China, and master’s degree in mechanical engineering from the University of Tokyo, Japan. He is currently a PhD student in Intelligent Transportation Thrust at the Hong Kong University of Science and Technology (Guangzhou). His research interests include automated driving systems, human factors, and driver behavior.

Dengbo He received his bachelor’s degree from Hunan University in 2012, M.S. degree from the Shanghai Jiao Tong University in 2016 and Ph.D. degree from the University of Toronto in 2020. He is currently an assistant professor from the Intelligent Transpiration Trust and Robotics and Autonomous Systems Thrust, the HKUST (Guangzhou). He is also affiliated with the Department of Civil and Environmental Engineering, HKUST, Hong Kong SAR.