Shifting Gears: Trust and Expectation Dynamics in Automated Vehicles

by

Qiaoning Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2023

Doctoral Committee:

Professor Lionel P. Robert Jr., Co-Chair
Associate Professor X. Jessie Yang, Co-Chair
Assistant Professor Paramveer Dhillon
Professor Mark Newman
Assistant Professor Feng Zhou

Qiaoning Zhang
qiaoning@umich.edu
ORCID iD: 0000-0002-2905-9853

DEDICATION

To my parents, mentors, and friends, for their unconditional love and support.

ACKNOWLEDGEMENTS

As I commenced my PhD journey at UMSI, I likened myself to an explorer embarking on a treasure hunt—an intricate, challenging, and mysterious expedition, filled with potential for profound fulfillment and intrigue. Standing now on the precipice of achieving one of the most significant milestones of my academic life, I am awash with gratitude for the bounty of intangible treasures I have amassed: curiosity, bravery, a diverse skill set, and, most importantly, confidence. Yet, this journey was not navigated alone. Scaling the daunting peaks of this academic endeavor would have been an insurmountable task without the guidance and support of numerous individuals. It is to these beacon lights of my academic path that I dedicate this acknowledgment.

My heartfelt gratitude goes foremost to my esteemed advisors, Dr. Lionel Robert and Dr. Xi Jessie Yang. Their unwavering support, astute guidance, and sagacious wisdom have anchored me throughout this journey. Through moments of frustration, confusion, and elation alike, their offices—whether physical or virtual—were sanctuaries of encouragement, expertise, and celebration. The inspiration derived from their dedication, amplified by the resilience of our relationship despite the pandemic's impact, propelled me towards academic and personal growth. Their contribution to my journey is an enduring treasure I will always cherish.

Secondly, I wish to express my appreciation to my committee members: Dr. Paramveer Dhillon, Dr. Mark Newman, Dr. Feng Zhou, Dr. Nadine Sarter, and Dr. Michael Nebeling. Their insightful advice and steadfast support played a vital role in molding my academic contributions and fortifying my determination to unravel sought-after truths. The wisdom gleaned from their suggestions and ideas reinforced my confidence in the importance of my work.

My sincere thanks also go out to my friends and colleagues at the lab, namely, Connor Esterwood, Suresh Kumaar Jayaraman, Kwame Porter Robinson, Sangmi Kim, and Rasha Alahmad. Additionally, I extend my gratitude to the wider community at the School of Information and my friends outside the PhD program. Special acknowledgement is due to Amy Eaton, the Senior Assistant Director of the UMSI Doctoral Program, for her invaluable guidance through numerous administrative and fellowship procedures.

Finally, but importantly, my heartfelt appreciation is dedicated to my family and friends. The relentless encouragement and unwavering love from my parents have formed the unshakeable foundation for my academic journey. Additionally, my gratitude extends to my best friend, Na Du. Her enduring love, support, and encouragement have been a source of joy and motivation. The moments spent discussing life, studies, family, and future aspirations with her have been invaluable, and her friendship is a gem that I will always cherish.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

LIST OF APPENDICES

LIST OF ACRONYMS

AVs Automated Vehicles

EDT Expectation Disconfirmation Theory

ECT Expectation Confirmation Theory

TAM Technology Acceptance Model

TTS Text-to-Speech

SAE Society of Automotive Engineers

CASA Computer Are Social Actors

UTAUT Unified Theory of Acceptance and Use of Technology

ABSTRACT

The research outlined in this dissertation provides a holistic exploration into the intricate factors influencing user expectations and trust in Automated Vehicles (AVs), both of which are central to the acceptance and adoption of such transformative technologies. AVs, with their potential to revolutionize transportation through enhanced safety, efficiency, and convenience, have generated widespread interest. However, persisting apprehensions around AV safety, performance, and reliability create a barrier to their widespread acceptance and utilization, suggesting a discrepancy between the theoretical advantages of AVs and public perceptions thereof. To unravel this discrepancy and effectively accelerate AV adoption, this dissertation undertakes a multifaceted investigation into the factors shaping public perceptions, specifically focusing on the formation of expectations and trust.

My research pivots around three core research questions: What individual factors shape people's expectations of AVs? How do these expectations impact their trust in AVs? How does the gender similarity between humans and AV explanation voices affect trust, and how is this moderated by gender-role congruity? The answers to these questions elucidate the intricacies of cognitive and affective trust development in the context of AV adoption.

Results from my dissertation highlight three major findings. One, individual characteristics such as demographic factors such as age, gender, and ethnicity, along with personality traits (e.g., extraversion, agreeableness, conscientiousness), significantly shape people's preconceived expectations of AVs. Two, expectations significantly mold the level of trust in AVs, influenced by the disconfirmation effect. Three, the study demonstrated that the impact of gender similarity between users and the AV's explanatory voice could be moderated by the expected role of the vehicle.

Overall, this dissertation embarks on a profound exploration of expectations, trust, and design elements, offering critical insights that will shape the forward path for AVs development and implementation. It dissects the intricate dynamics of expectation and trust formation, essential for the user acceptance and adoption of AVs. The study also underscores the powerful role of both user-centric and voice characteristic design in influencing these factors. By bringing these components to light, this research helps navigate the complexities and potentials of AVs, paving the way for an imminent paradigm shift in transportation.

CHAPTER 1

Introduction

1.1 Motivation

Automated vehicles (AVs) have the potential to revolutionize transportation by providing safer, more efficient, and convenient travel options. Despite the myriad of benefits AVs offer, recent studies reveal that the general public remains apprehensive about their safety and readiness for widespread adoption [37, 158, 178, 57]. Surveys indicate that over 50% of Americans perceive AVs as riskier than human-operated vehicles, with a significant 7% expressing reluctance to purchase a fully autonomous car [111]. Moreover, a substantial 78% of Americans report feeling uneasy about riding in an AV, while a mere 19% express confidence in their capabilities [171]. This reluctance to embrace AV technology could hinder the realization of its full potential, thus preventing society from fully benefiting from its many advantages. To accelerate AV acceptance and promote their integration into everyday life, it is crucial to explore the factors that influence the public's willingness to adopt these vehicles.

There is substantial empirical evidence that underscores the influence of expectations on technology adoption in general, while research focusing specifically on the role of expectations within the context of AVs is still in its infancy. Expectations refer to beliefs about the future performance of a particular technology [12]. Numerous studies have emphasized the vital role expectations play in shaping public attitudes toward technology, as well as influencing their willingness to adopt new innovations across various domains [77, 18, 17, 185]. Expectation Disconfirmation Theory (EDT) posits that a customer's satisfaction and future purchase intentions hinge on whether a product or service failed to meet, met, or exceeded their initial expectations [72, 126, 148]. If expectations are set too high or too low, users may experience disappointment, distrust, or be deterred from trying the technology [94, 96]. Therefore, establishing an appropriate level of expectation is crucial for encouraging individuals to adopt AVs and minimizing disappointment. To better understand the relationship between expectations and AV adoption, it is vital to identify factors shaping public expectations and recognize variations in AV expectations among specific individuals. Although individual differences have been demonstrated to affect technology adoption, there is limited knowledge regarding how these differences influence AV expectations.

Trust is another essential factor in determining AV adoption. Defined as the trustor's willingness to be vulnerable to the actions of the trustee, trust plays a crucial role in the acceptance of technology, including AVs [71, 76, 103, 120, 205]. Studies suggest that the degree of trust in AVs directly correlates with the willingness to adopt them. Recent surveys have found that a lack of trust in AV systems is the most frequently cited reason for drivers' reluctance to adopt this technology [1, 206]. Moreover, Choi and Ji (2015) applied the Technology Acceptance Model (TAM) to show that trust significantly affects perceived usefulness, and both factors determine behavioral intentions to adopt AVs [24]. Thus, trust is a key predictor of AV adoption, and fostering trust can encourage AV adoption. Although setting appropriate expectations has the potential to cultivate trust, few studies have examined the effect of expectations on trust in AVs. Research has shown that individuals are more likely to trust a system if its observed accuracy exceeds their expectations, suggesting that system accuracy expectations can significantly impact trust [199]. However, the specific role expectations play in nurturing trust in AVs remains unclear. Further research is necessary to better understand the relationship between expectations and trust in AVs and to investigate the underlying mechanisms through which expectations influence trust.

Providing explanations for AV actions is another approach to promoting trust in AVs. Explanations refer to the reasoning or logic behind actions, offering users essential information that often justifies decisions made by the AV [202]. This information helps users develop a mental representation of the system's functions and competencies, enabling them to take appropriate precautions in sudden takeover scenarios, understand future AV functions, and build confidence in the technology [37, 53, 92, 181]. Prior research has identified a significant positive relationship between providing AV explanations and trust in AVs [37, 53, 64, 165]. Although research on how explanations affect AV trust has progressed, there has been limited focus on AV explanation voice design and the multi-dimensional aspects of trust.

First, while numerous preceding studies have placed an emphasis on delivering explanations via auditory means in AVs, there remains a significant gap in understanding how different voice characteristics influence users' trust. The Computer Are Social Actor (CASA) paradigm posits that people tend to interact with technology in a similar fashion as they would in social interactions, essentially perceiving technology as a social entity [138, 54, 56, 104]. Notably, specific voice attributes such as gender can relay socially relevant cues, thereby potentially swaying individuals' attitudes and adoption of technology. Empirical studies suggest that the perceived similarity in demographic attributes between humans and voice agents can improve the quality of human-agent interactions. For instance, when users perceive similarity in attributes such as gender, personality, ethnicity, or age, the results are generally positive, manifesting as enhanced satisfaction and heightened emotional responses [42, 43, 99]. Furthermore, user preferences for voice characteristics may oscillate based on the perceived alignment between the voice's gender and the projected role of the technology, an inclination mirroring gender-role congruity and societal gender stereotypes. As such, female voices often find favor in scenarios where help is needed, while male voices are predominantly preferred in authoritative or knowledge-based contexts [106]. Yet, when a perceived incongruity arises between the gender and role, it can trigger a sense of "lack-of-fit," often resulting in negative performance evaluations and potential constraints [69, 70, 68, 116]. Despite these insights, the relationship between similarity in human and AV explanation voices, gender-role congruity, and trust formation within the context of AVs remains largely uncharted. Consequently, further detailed research is required to demystify these dynamics, potentially facilitating the design of more effective AV systems that engender trust and align with user preferences and needs.

Moreover, although numerous studies have been conducted on trust in the realm of AVs, much of this research assumes that trust is uni-dimensional rather than multi-dimensional [37, 53, 165, 64]. Lewis and Weigert's (1985) trust theory posits that trust is a multifaceted concept, consisting of two essential types: cognitive trust and affective trust [109]. Cognitive trust arises from a rational process that distinguishes between trustworthy, distrusted, and unknown agents, involving the identification of reasons for trust and the accumulation of evidence demonstrating trustworthiness [109, 121]. In contrast, affective trust is based on an emotional connection between all parties involved in the relationship. Individuals invest emotionally in trust relationships, show genuine care and concern for their partners' well-being, recognize the inherent value of such relationships, and believe that these feelings are mutual [109, 121, 156]. While research on interpersonal trust has emphasized the importance of differentiating between cognitive and affective trust, the application of this distinction in the AV domain remains unexplored. It is unclear whether AV-related antecedents will have varying effects on cognitive and affective trust. Future research could help address this gap and deepen our understanding of the multi-dimensional nature of trust in AVs.

The primary objective of this dissertation is to foster AV adoption by extensively examining AV expectations and their influence on trust development in these vehicles. Moreover, this research seeks to provide recommendations regarding AV explanation voice design to establish trustworthy human-AV interactions that strengthen trust from both cognitive and affective perspectives. To achieve this objective, a series of studies will be conducted, integrating insights from the domains of human-AV interaction, interpersonal relationships, and information systems.

RQ 1: What are the individual factors that shape people's expectations of AVs?

RQ 2: How do people's expectations of AVs impact their trust in AVs?

RQ 3: How does the gender similarity between humans and AV explanation voices affect the effectiveness of AV explanations in fostering both cognitive and affective trust in AVs? Additionally, to what extent does gender-role congruity moderate this effect?

1.1.1 Individual Factors and AV Expectations

The Society of Automotive Engineers (SAE) has developed a classification system for driving automation, consisting of six levels ranging from 0 to 5 (Figure 1.1). As the levels increase, the need for driver involvement decreases. At level 3, the automated driving system may prompt human driver intervention when needed. However, at levels 4 and 5, the automated driving system assumes full responsibility for all driving tasks under certain and all conditions, respectively [169]. In the context of this study, AVs pertain to vehicles at level 4 and above. AVs hold the potential to substantially enhance vehicle safety by eradicating human-related errors and optimizing traffic flow, thereby reducing carbon dioxide emissions and fuel consumption [8, 170, 16]. Furthermore, they have the potential to improve mobility access for elderly and physically impaired individuals by enabling them to safely navigate traffic [141]. Given that public opinion plays a pivotal role in the adoption of AVs, comprehending the factors that influence people's perspectives is of utmost importance [4, 38, 24].

Expectations significantly influence users' technology adoption decisions [24, 9, 180]. These expectations, which are beliefs regarding the future performance of technology [12], pertain to system attributes such as the ability to enhance task performance, boost efficiency, and improve work quality [186]. Initial expectations about technology are shaped by an individual's existing knowledge and the communication channels or information sources they utilize [201]. Prior research indicates that customer satisfaction and subsequent future purchase intentions are influenced by whether a technology or service fails to meet, meets, or surpasses initial expectations [72, 126, 148]. Excessively high expectations can lead to disappointment after using the technology, resulting in distrust and rejection [94]. On the other hand, overly low expectations may deter people from using the technology in the first place [96]. Consequently, establishing appropriate expectations is essential for encouraging

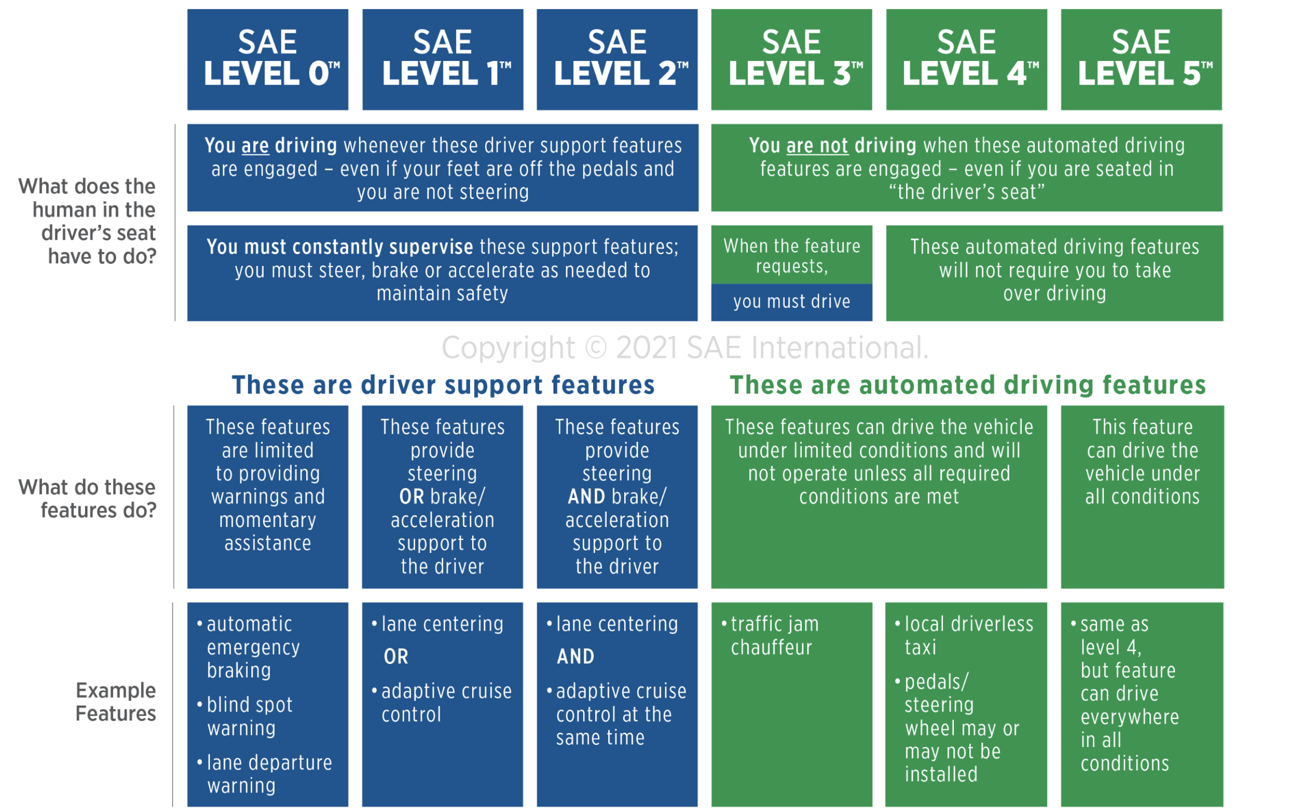

Figure 1.1: SAE J3016 levels of driving automation [27]

individuals to utilize AVs and minimizing disappointment by aligning or surpassing expectations with actual experiences. By comprehending the variations in AV expectations among specific individuals, we can set suitable expectations and design AVs to meet or exceed those expectations for each user.

Expectations related to certain technologies have been demonstrated to vary significantly among individuals; however, we possess limited knowledge about if or how such differences are manifested in the context of AVs. The concept of "individual differences" encompasses attributes such as demographics and personality [63, 161]. Previous research has explored the effect of individual differences on technology adoption [49, 2, 168, 125, 117]. Employing the technology acceptance model, Agarwal and Prasad (1999) [2] discovered that individual differences influence technology acceptance through their impact on personal beliefs about technology. Similarly, AV research has indicated that individual differences affect AV adoption. For instance, multiple studies have revealed that women exhibit greater concerns about AVs and are less inclined to adopt them compared to men [168]. Age has also emerged as a crucial individual difference in the adoption of AVs, with younger drivers being more likely to embrace AVs than older drivers [125, 117]. Nonetheless, the issue of whether individual differences shape AV expectations has garnered minimal attention. This lack of focus is unexpected, considering the significance of expectations in technology adoption overall and AV adoption specifically. Please refer to Chapter 2 Background section for a comprehensive review of the literature relevant to this study.

Considering the significance of individual differences in AV adoption and the sparse attention devoted to their impact on AV expectations, **Chapter 2** of this dissertation carries out an online survey to bridge this knowledge gap (Figure 1.2). Comprising a representative sample of 443 U.S. drivers, the survey collects data on age, gender, race and ethnicity, education, income, marital status, geographic region, driving frequency, driving experience, and personality. The study investigates the relationship between these essential characteristics and individual differences in AV expectations. The results hold considerable implications for both AV adoption research and AV design.

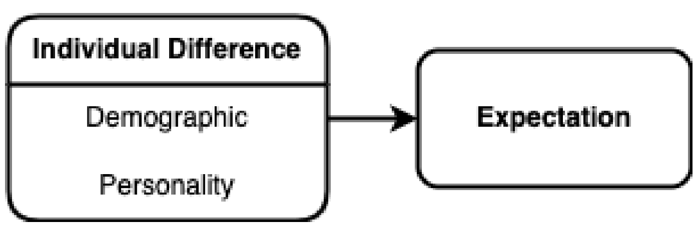

Figure 1.2: Research model in Chapter 2.

1.1.2 Expectations and Trust in AVs

Trust in AVs has been widely studied, concentrating on identifying its determinants and effects based on a variety of trust theories and models. Nevertheless, the role of expectations in shaping trust in AVs has been relatively underexplored. Expectations pertain to beliefs about a technology's future performance, often linked to the outcomes of current actions [12]. The Expectation Disconfirmation Theory (EDT) provides valuable insights into how users compare their expectations to actual experiences, forming disconfirmation that ultimately influences their attitudes and technology adoption [148, 36, 187]. Disconfirmation denotes the degree to which technology performance aligns with or deviates from initial expectations regarding specific attributes [10]. According to EDT, negative disconfirmation arises when experiences fail to meet expectations, resulting in a disappointment effect that deters users from further adoption and usage intentions [123, 176, 95]. Conversely, positive disconfirmation occurs when experiences surpass expectations, generating a positive surprise effect that boosts satisfaction and fosters future usage intentions [95]. EDT has also been applied to investigate the role of trust as an adoption predictor in relation to expectations. Prior research indicates that excessively high expectations can lead to disappointment, and ultimately, distrust and rejection of technology following its use [94]. On the other hand, low expectations may dissuade people from utilizing technology altogether [96]. Despite this, limited literature delves into the role of expectations in the cognitive appraisal process and how they impact the development of trust in AVs from an EDT perspective.

Developing trust in AVs is closely tied to individuals' risk perception while driving. Risk perception is a blend of uncertainty and the severity of potential outcomes involved [128]. Studies have shown that trust and risk are connected to uncertainty and vulnerability. As a result, cultivating trust in AVs is linked to individuals' risk perception [71]. Furthermore, trust is predicated on risk, as parties need to be open to vulnerability and perceive a certain degree of risk to establish trust [48, 198, 22]. For instance, Zhang et al. (2019) [204] discovered a significant negative correlation between risk and trust, a finding also supported by Verberne et al. (2012)'s study, which reported an increase in AV trust as risk decreased [188]. Additionally, previous research indicates that both external factors, such as weather, and internal vehicle characteristics, like driving speed, significantly affect perceived risk, which in turn influences trust in AVs [4, 61]. To gain a deeper understanding of the interplay between expectations, disconfirmation, and trust in AVs, it is vital to examine how individuals' risk perception of both internal and external factors moderates the relationship between disconfirmation and trust. By considering the role of risk perception, we can develop a more comprehensive understanding of the factors shaping individuals' trust in AVs and ultimately foster the adoption of this technology. For a comprehensive review of the relevant literature utilized in this study, please refer to the Background section in Chapter 3.

Chapter 3 of this dissertation seeks to examine the impact of disconfirmation between expectations and perceived performance on trust in AVs, and how the perceived risk of driving situations may potentially modify this relationship (as illustrated in Figure 1.3). To accomplish this, an online survey was conducted with 443 drivers in the US. The study evaluated two types of weather conditions as external environmental factors: sunny and snowy, and two types of driving behavior of AVs as internal factors: normal and aggressive driving. The findings of the study uncovered a complex interplay between expectations, disconfirmation, and trust in AVs, emphasizing the significant role of expectations in the cognitive appraisal process and their impact on building trust in AVs. Furthermore, the study highlights the moderating effects of both internal and external risk factors on the relationship between disconfirmation and trust in AVs. Therefore, understanding these factors is essential for devising effective strategies to enhance trust in AVs and encourage their adoption.

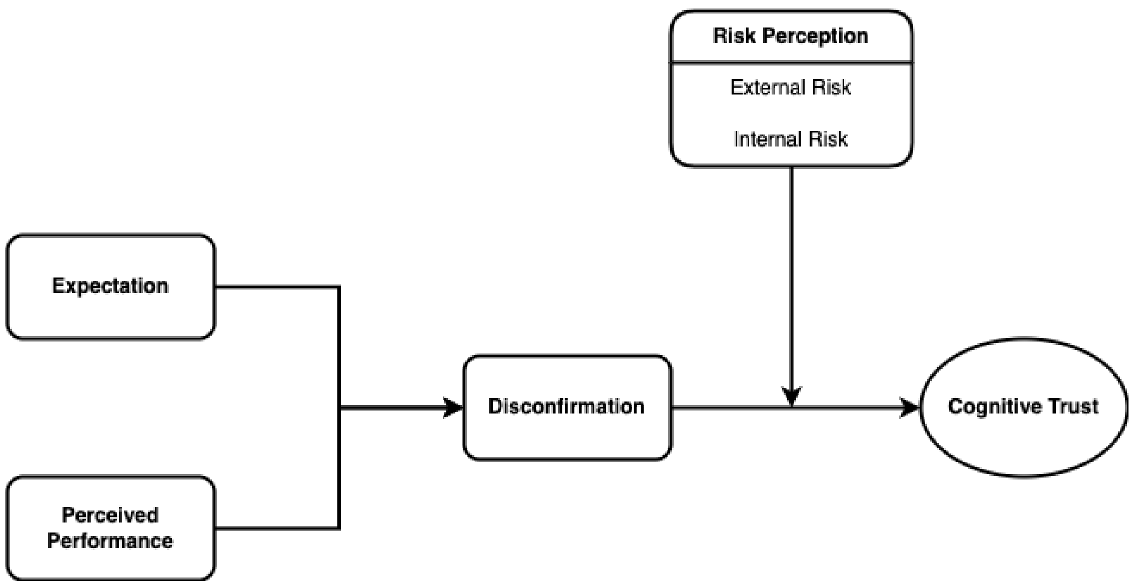

Figure 1.3: Research model in Chapter 3.

1.1.3 AV Design for Building Cognitive and Affective Trust

Disconfirmation, the process of comparing initial expectations with actual AV performance, can substantially impact trust in AVs. To bolster trust, it is vital to not only comprehend users' expectations but also enhance AV performance. Numerous studies have explored ways to develop AV features that foster improved human-AV interaction, including the implementation of AV explanations. These explanations offer passengers information about AV actions to augment understanding and transparency [202]. By rendering AV actions predictable and intelligible, AV explanations allow users to develop accurate mental models of the AV system, ultimately boosting confidence and trust [92]. Furthermore, AV explanations assist drivers in forming a rough representation of the system's functions and capabilities, enabling them to take suitable precautions in sudden takeover situations, comprehend future AV functions, and build confidence in them [37, 53, 181].

Prior research on AV explanations has primarily utilized two modalities: auditory and visual. Auditory explanations have generally been conveyed via a neutral-toned male or female voice with a standard American accent, while visual explanations have been presented in text form [37, 92, 85, 86, 165, 142, 64, 193]. Despite the emphasis on auditory explanations, the influence of voice characteristics on users' trust in AVs has not been adequately explored, despite their significant implications for human-technology interactions.

The Computer Are Social Actor (CASA) theory suggests that individuals interact with technology in a similar manner to social interactions, treating technology as a social entity. As such, voice characteristics, including gender, serve as channels for conveying socially relevant cues, influencing attitudes and the extent of technology adoption [54, 56, 138, 104]. Importantly, individuals often unconsciously apply social theories and models, such as the similarity-attraction theory and role congruity theory, in their interactions with technology.

The similarity attraction theory postulates that individuals are naturally inclined towards those who share similar attributes [97, 129, 191]. It posits that people are more likely to form relationships and perceive others more positively when they believe them to share similar beliefs, attitudes, values, personality traits, or interests [59, 190, 60]. This principle extends to technology design; for instance, users often perceive recommendation agents matching their ethnicity as more sociable and useful [155]. Moreover, Lee et al. (2007) [105] discovered that people tend to trust synthesized speech more and learn more effectively when the voice gender aligns with their own.

Beyond the impact of similarity on user preferences, the role congruity theory underlines how perceived role congruity can affect outcomes when existing stereotypes align with the characteristics perceived as necessary for success in a particular role [82]. For example, roles associated with male stereotypes often include high levels of assertiveness and dominance, while roles related to female stereotypes generally involve nurturing and deference [82]. This concept of gender-role stereotype and congruity extends to technology as well, with users typically preferring a male voice for delivering informational content and a female voice for more relational, social communication [106].

Implementing the principles of the similarity attraction theory and role congruity theory in technology design can greatly enhance user experiences and nurture trust in human-technology interactions. It is increasingly acknowledged that the gender similarity between human users and voice agents, along with gender-role congruity, can shape user preferences and thus improve the dynamics of interaction. Yet, research investigating these factors' influence on trust formation in the context of AVs is notably limited. In particular, there is a significant knowledge gap surrounding how gender similarity and the alignment between voice gender and the perceived AV role influence trust formation in human-AV interactions, particularly with respect to cognitive and affective trust. This gap highlights the need for more comprehensive research to guide the design of AV agents that effectively promote trust. The Background section in Chapter 4 provides an in-depth examination of the literature relevant to this study.

Chapter 4 of this dissertation (Figure 1.4) seeks to examine the impact of gender similarity between human users and AV explanation voice, along with the potential moderating effect of gender-role congruity on the development of cognitive and affective trust in AVs.

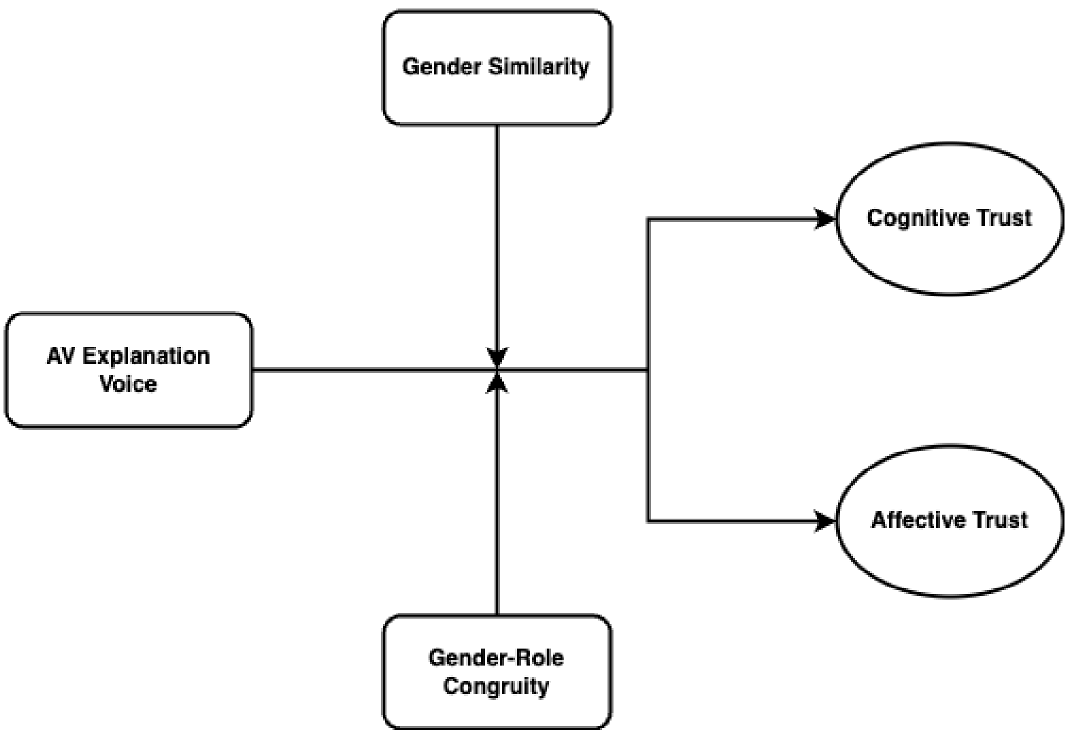

Figure 1.4: Research model in Chapter 4.

1.2 Dissertation Overview

This dissertation embarks on an in-depth exploration of trust in AVs, divided into three distinct studies. Each one delves into diverse research questions associated with the complex relationship between trust and expectations in AVs, while also shedding light on the cognitive and emotional mechanisms at play during trust formation. The core objective of the dissertation is to broaden the understanding of AV adoption by critically analyzing the interplay between trust and expectations regarding this groundbreaking technology.

AVs hold the potential to revolutionize transportation by enhancing road safety, promoting independence, reducing traffic congestion, and curbing pollution [37, 178, 57]. Yet, the challenges related to the public's acceptance and adoption of this technology are substantial. Recent surveys reveal considerable apprehensions among US drivers about riding in an AV and discomfort at the idea of sharing the road with them [111]. However, a smaller proportion of the population displays positive attitudes towards this technology [171]. To fully capitalize on the potential benefits of AVs for society, boosting their acceptance and adoption is of paramount importance.

Expectations, which significantly influence users' technology adoption decisions [24, 9, 180], relate to system attributes like task performance and efficiency enhancement [186]. These are molded by individual's knowledge and information sources [201]. Misaligned expectations can result in disappointment or deter usage [94, 96], underscoring the necessity of understanding variations in AV expectations to set appropriate ones. Although individual differences in technology expectations are recognized [63, 161, 49, 2], understanding about how these differences emerge in AVs is limited. Previous research has shown that individual differences like gender and age influence AV adoption [168, 125, 117], but studies on how these differences shape AV expectations are scarce. This knowledge gap emphasizes the importance of potential research in the role of expectations in AV adoption. Our investigation of these differences and their impact on AV expectations will yield invaluable insights into the process of AV adoption.

Further, the correlation between trust and adoption is compelling, especially concerning AVs. Despite substantial research into the determinants and effects of trust, the role of expectations in trust formation remains under-researched. Applying Expectation Disconfirmation Theory (EDT) provides insights into how users' comparison of expectations with actual experiences leads to disconfirmation that influences adoption [148, 36, 187]. Disappointment discourages users when experiences fall short of expectations [123, 176, 95]. On the other hand, when experiences surpass expectations, a positive surprise effect amplifies satisfaction and usage intentions [95]. While prior research connects excessively high or low expectations to rejection or non-use of technology [94, 96], few studies explore the role of expectations in shaping trust in AVs from an EDT perspective. Trust in AVs also connects to an individual's risk perception [71], a mix of uncertainty and potential severity of outcomes [128]. External factors such as weather and internal vehicle characteristics like driving speed significantly affect perceived risk, thereby influencing trust in AVs [4, 61]. Several studies found negative correlations between risk and trust [204, 188]. Therefore, understanding the interplay between expectations, disconfirmation, and trust in AVs necessitates considering individuals' risk perception of both internal and external factors. This approach can yield a more holistic understanding of factors shaping trust in AVs, thereby promoting AV technology adoption.

Recognizing that disconfirmation, or the comparison of expectations to AV performance, significantly influences trust in AVs is crucial. Improving AV performance and understanding user expectations is key to building trust. Research highlights AV explanations as a means to enhance human-AV interaction, making AV actions more predictable and understandable, thereby fostering trust [202, 92]. They assist users in comprehending the AV system's capabilities, particularly in abrupt takeover situations [37, 53, 181]. The majority of research concentrates on auditory and visual explanations, with the impact of voice characteristics on trust largely unexplored, despite potential implications. The Computers are Social Actors (CASA) theory suggests that users treat technology as a social entity, with voice characteristics influencing attitudes and technology adoption [54, 56, 138, 104]. Similarity attraction and role congruity theories also affect technology interaction. Similarity attraction theory posits that individuals tend to be drawn to those with similar attributes, influencing technology design [97, 129, 191, 59, 190, 60, 155, 105]. Role congruity theory emphasizes how perceived role congruity, swayed by gender-role stereotypes, influences outcomes [82, 106]. Incorporating these theories into technology design improves user experience and trust. The gender similarity between users and voice agents, along with gender-role congruity, shapes user preferences, enhancing interaction dynamics. However, research on their influence on trust in AV interactions, particularly cognitive and affective trust, is limited, signifying a need for further investigation to effectively guide AV agent design.

Overall, in a comprehensive endeavor to deepen our understanding of the current knowledge landscape in the field and identify pertinent research gaps, we meticulously conducted an extensive literature review. This dissertation, divided into three distinct but interconnected studies, aims to enrich the scholarly dialogue on the subject.

1.2.1 Chapter 2

Chapter 2 focuses on an encompassing nationwide survey that investigates the preliminary expectations of AVs among a diverse set of user groups, each with distinct individual characteristics. Given the pivotal role of expectations in the acceptance and adoption of new technology, this study seeks to unpack the distinctive levels of AV expectations that materialize before users have firsthand AV experiences. The findings are listed below:

Chapter 2 Finding Highlights:

- Drivers' expectations of AVs show considerable variation by demographic characteristics and personality traits.
- Younger, educated, unmarried, White non-Hispanic men with frequent but less experienced driving habits generally have higher AV expectations.
- Drivers exhibiting high levels of extraversion, agreeableness, conscientiousness, and emotional stability tend to have higher AV expectations.

1.2.2 Chapter 3

Chapter 3 delves deeper into understanding the implications of these diverse initial expectations on the formation of trust in AVs. It employs a survey-based methodology to scrutinize how cognitive appraisals of discrepancies between initial expectations and AV performance influence trust formation, under the theoretical umbrella of Expectation Disconfirmation Theory (EDT). Additionally, the chapter sheds light on the potential moderating role of internal and external risks in the relationship between disconfirmation and cognitive trust. The findings can be summarized as follows:

Chapter 3 Finding Highlights:

- Disconfirmation, the discrepancy between expectations and performance, significantly impacts trust in AVs.
- Negative disconfirmation, where performance falls short of expectations, results in diminished trust
- Positive disconfirmation, where performance exceeds expectations, promotes increased trust.
- Perceived risk, influenced by both internal and external factors, significantly impacts the relationship between disconfirmation and trust in AVs.

1.2.3 Chapter 4

Chapter 4 shifts the spotlight onto an essential facet of AVs – the design – underscoring its significance in nurturing trust and facilitating effective human-AV interactions. This chapter argues for the inclusion of unique voice characteristics in AV explanations, a feature aimed at not only bolstering cognitive trust but also fostering affective trust, anchored in the principles of similarity attraction theory and role congruity theory. The ultimate goal is to devise an AV that is deemed trustworthy and user-friendly by virtue of features that resonate with users at both cognitive and emotional levels. The key findings are outlined as follows:

Chapter 4 Finding Highlights:

- A higher level of cognitive and affective trust is fostered when the AV voice aligns with the user's gender, as opposed to when interacting with a dissimilar voice.
- The effect of gender similarity on affective trust is negatively influenced when the gender of the AV voice contradicts its gender-stereotypically expected role.

The research framework illustrated in Figure 1.5 summarizes the objectives of this dissertation. This work aims to contribute significantly to the rapidly evolving research field of AVs, proposing practical design recommendations that facilitate the formation of optimal trust in and adoption of AVs.

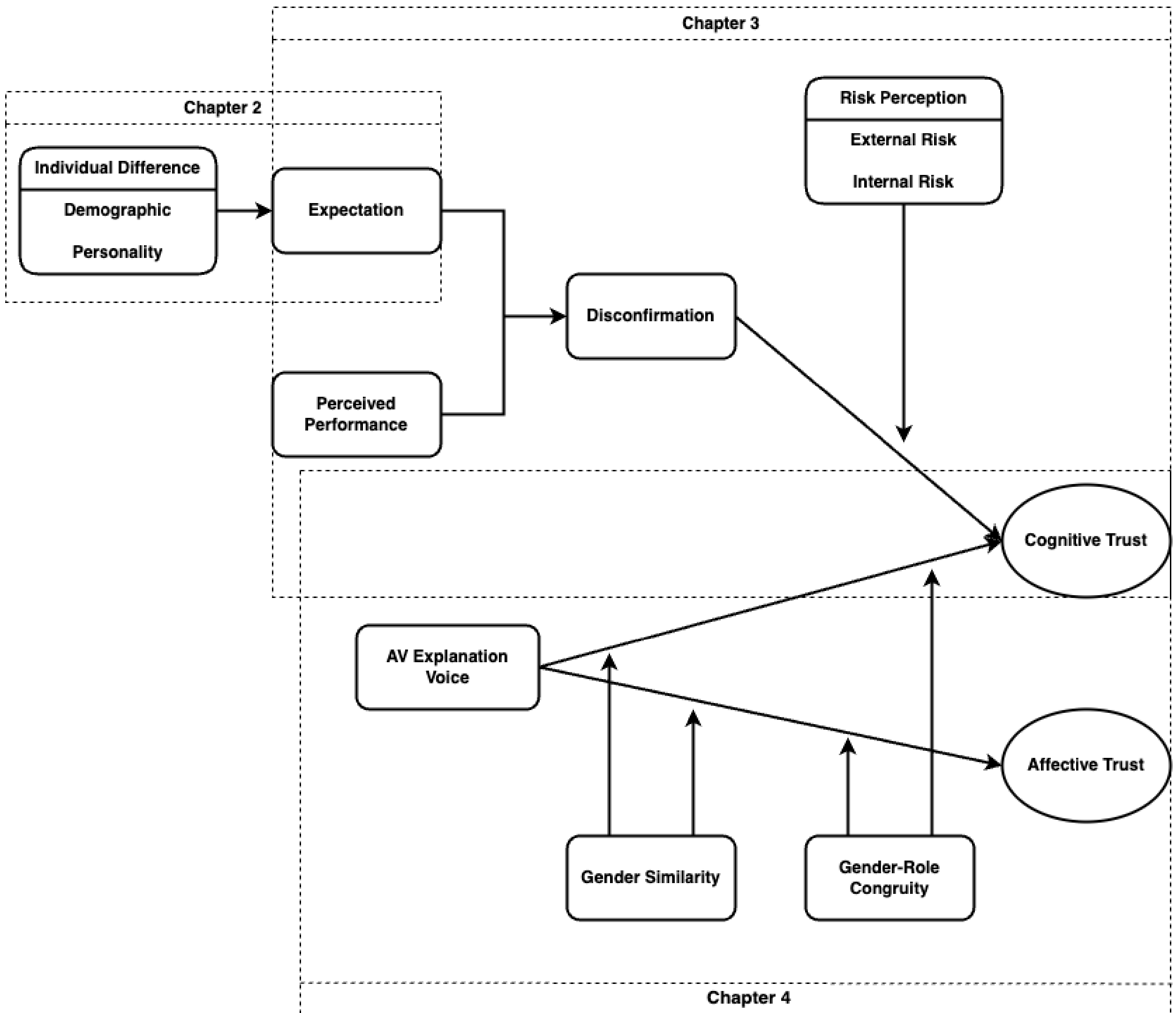

Figure 1.5: Research framework.

CHAPTER 2

Individual Differences and Expectations of Automated Vehicles

2.1 Introduction

Automated Vehicles (AVs) hold the promise to transform transportation radically, with potential benefits including reduced automobile accidents, preservation of human lives, and decreased fuel consumption and environmental pollution. Despite these advantages, public adoption of AVs remains unpredictable, with a substantial proportion of people expressing hesitancy towards using AVs. Given this situation, public sentiment is crucial for the broader acceptance of AVs, necessitating a deep understanding of the factors shaping these opinions.

Among the many influencing factors, expectations stand out as pivotal determinants for technology adoption decisions. Setting realistic and appropriate expectations is therefore vital to motivate individuals towards AV usage. While previous studies have highlighted the influence of individual differences on technology acceptance, the impact of these differences on AV-specific expectations remains underexplored. To bridge this knowledge gap, we conducted an online survey with 443 U.S. drivers to examine the link between individual differences and expectations surrounding AVs.

Our research furnishes valuable insights into how individual traits shape expectations about AVs and carry significant implications for research on AV adoption and the design of AVs. By identifying the relationship between individual characteristics and expectations, this study contributes to the efforts in facilitating a smoother transition towards AV adoption, thereby harnessing the transformative potential of this groundbreaking technology [1].

[1]This chapter is published in the International Journal of Human–Computer Interaction 2022, VOL. 38, NO. 9, 825–836; https://doi.org/10.1080/10447318.2021.1970431

2.2 Background

In this section, we provide an overview of the literature that shaped and inspired our research. Initially, we examine literature that highlights the importance of expectations in technology adoption, with a particular emphasis on how it impacts the adoption of AVs. Subsequently, we review previous research on individual differences, such as personality and demographics, and their connection to AV adoption. Specifically, we delve into research that has explored the impact of individual differences on the public's concerns regarding the adoption of AVs.

2.2.1 Expectations and Technology Adoption

The importance of expectations has been greatly emphasized concerning consumer satisfaction and the adoption of technology. In this paper, expectations are defined as beliefs about the future performance of a given technology [12]. The Expectation-Confirmation Theory (ECT), also known as Expectation-Disconfirmation Theory (EDT), provides an explanation of the impacts of expectations on technology adoption. The ECT was initially used to understand consumer satisfaction with a given product [72]. Generally, customers were more satisfied with a product when it met or exceeded their expectations and were less satisfied with the product when it failed to meet their expectations. One implication of this finding is that setting initial expectations is vital to promoting consumer satisfaction [126, 148, 180]. Research on technology adoption has found similar results: when technology performance exceeds expectations, users are much more likely to adopt that technology [17, 187, 119, 167].

Researchers have also examined the impacts of expectations on the intention to adopt AVs. For example, Tussyadiah, Zach, and Wang (2017) conducted a survey study with 325 U.S. residents and found that the likelihood of using AVs (e.g., self-driving taxis) was positively associated with their expectations of the AVs' reliability, functionality, and helpfulness [184]. Similarly, Ro and Ha (2019) examined 1,506 survey responses from South Korea to identify relationships among expectations, attitudes, and behavioral intentions [157]. Ro and Ha found that expectations are positively associated with attitude toward using an AV, which in turn is positively associated with intention to use. Körber, Baseler, and Bengler (2018) examined and found that AV expectations are associated with trust in an automated driving system; they also found that trust in an automated driving system is positively correlated with reliance on the automated driving system [92].

Expectations have also been employed in broader technology adoption theories including the technology adoption model and the unified theory of acceptance and use of technology (UTAUT). For example, Kaur and Rampersad (2018) surveyed 101 responses to examine the effects of key factors (e.g., performance expectancy, reliability, security, and privacy) and found that expectations positively influence the adoption of AVs [78]. Similar to Kaur and Rampersad, researchers Madigan, Louw, Wilbrink, Schieben, and Merat (2017) surveyed 315 respondents from the city of Trikala, Greece, and applied UTAUT to investigate the factors that influence users' acceptance of AVs (i.e., automated road transport systems) [118]. Results provided evidence that expectations regarding performance have a significant impact on the intention to use AVs. The results also indicated that effort expectations was not a pivotal factor to impact the intention to use, suggesting that AV adoption is unlikely to be influenced by the effort required to operate AVs. In all, the existing literature has consistently found a strong link between AV expectations and attitudes toward and the adoption of AVs.

2.2.2 Individual Differences and Automated Vehicles

Individual differences are the enduring psychological characteristics that distinguish one person from another and help define a person's individuality [28]. Individual differences have been linked to the adoption of various technologies across many settings [25, 63, 100, 124]. Demographics and personality traits have been used to represent individual differences [100, 161]. Various studies have assessed the impact of individual differences on AV acceptance. Next, we present and discuss the literature on individual differences and AV adoption.

Age is among the most important individual differences in predicting AV acceptance. The AV literature found that older drivers generate more negative attitudes and reluctance to adopt AVs. For example, Schoettle and Sivak (2014) conducted a survey with 1,533 respondents from the U.S., the U.K., and Australia to understand their opinions and concerns about accepting AVs [168]. The results indicated that younger drivers are less concerned about AVs, more interested in having AV technology on their vehicle, and more likely to ride in AVs than older drivers. Older drivers also tend to distrust AVs, while younger drivers have shown higher trust in AVs and higher intention to use AVs [49, 55, 88]. Driver's age has an important role in understanding the impact of AV explanation on AV trust. Zhang et al. (2021) found that older drivers had higher trust in an AV when it asked for permission to take action, while for younger and middle-aged drivers this actually lowered trust and increased their anxiety [203]. They also found that younger drivers have the lowest anxiety when the AV provided explanations after it took action. On the contrary, this condition produced the highest level of anxiety for middle-aged and older drivers [203].

Gender is a prominent factor in predicting whether someone accepts AV. For instance, Nordhoff et al. (2018) obtained 7,755 survey responses from 116 countries to investigate the determinants of acceptance of driverless shuttles [145]. Their results revealed that men are much more willing to accept driverless shuttles than women. Esterwood et al. (2021) conducted an online survey with 428 participants to understand the impact of demographic differences on the acceptance of autonomous buses. The results showed that males expressed a higher intention to ride an autonomous bus than females. They also found that females were more concerned about self-driving vehicles and are less likely to believe in their potential benefits than men [49].

Race and ethnicity have also been investigated as a determinant of AV acceptance. Prior research has been conducted to examine the relationship between race/ethnicity and AV-related attitudes and adoption. For example, Howard and Dai (2014) found that individuals who self-identified as Hispanic and Asian tend to value AVs' potential to improve mobility for people with driving impairments [73]. Asian Americans had a significantly higher positive attitude towards autonomous buses than White Americans. Another study found that Asian Americans also had a significantly higher positive attitude toward and expressed greater intention to ride autonomous buses than those identifying as White Americans [49].

Education level is closely associated with AV acceptance. For instance, Schoettle and Sivak (2014) found that higher education levels are associated with the intention to adopt AV technology [168]. Individuals with higher education levels were found to be more likely to have advanced driving technology on their vehicle, more likely to say they would read or work while using AVs, and less likely to say that they would not ride in an AV [168].

Prior research also explored the role of income in influencing AV acceptance. Howard and Dai's (2014) survey study found that people with lower income are more concerned with safety issues and giving up control while higher-income drivers pay more attention to liability. Marital status also has an impact on attitudes toward AVs [73]. Results of Howard and Dai's (2014) study suggested that married people are less concerned with cost and amenities (e.g., the ability to text message or multitask while driving) but place high importance on safety.

Driving experience and frequency are two individual differences that play roles in understanding the acceptance of advanced driving-related technology like AVs. Koul and Eydgahi (2018) found a negative relationship between driving experience and AV adoption [88]. In Koul and Eydgahi, drivers' intention to use an AV decreased slightly as their years of driving experience increased. Driving frequency has also been found to be negatively associated with AV acceptance. Rödel et al. (2014) conducted a survey and found that frequent drivers prefer unassisted driving to technology-assisted driving. Frequent drivers described using the assisted vehicle technology as more challenging and less controllable than people who drove less often [166]. Frequent bus riders had a higher positive attitude and intention to ride autonomous buses than in-frequent riders [49].

Geographic region, which has often been used as a proxy for differences in prevailing values and belief systems, has also been identified as a factor that impacts the acceptance of driving technology. Carley et al. (2013) conducted a survey study in large U.S. cities to examine the intention to purchase plug-in electric vehicles [21]. The results noted a significant difference in electric vehicle adoption across major cities in the United States. Also, the purchase intention of alternative-fuel vehicle technologies differed by geographic regions due again to potential differences in culture, values and beliefs [153, 175]. However, geographic region's impact on attitudes toward and adoption of AVs still needs further investigation.

Another class of individual differences—personality—is also associated with AV adoption. Personality is defined as "generalized and personalized determining tendencies—consistent and stable modes of an individual's adjustment to his environment," which can be used as a label to describe traits that represent an individual's predisposition toward behavior or objects [3]. The Big Five is the most popular set of personality traits used across many domains, providing a comprehensive taxonomy of individual differences. The Big Five personality model includes openness to experience, conscientiousness, extraversion, agreeableness, and emotional stability [33]. Openness to experience represents the flexibility of thought and tolerance of new ideas [122]. Conscientiousness reflects self-control and a need for achievement and order [149]. Extraversion is the extent to which an individual is assertive, outgoing, talkative, and sociable [33]. Agreeableness represents the extent to which someone is kind, considerate, likable, and cooperative [152]. Emotional stability is the degree to which someone is well-adjusted, emotionally stable, and secure [152].

Kyriakidis, Happee, and de Winter (2015) investigated the relationship between personality traits and concerns over fully AVs [91]. They found that respondents who scored high on emotional stability or lower on agreeableness were more likely to believe that automation was less silly and were also less comfortable with AV data transmission. In addition, T. Zhang et al. (2020) investigated the role of social and personal factors in AV acceptance using a questionnaire survey with 647 drivers in China. Results indicated that drivers with an openness to new experiences are more likely to trust and accept AVs [205]. On the contrary, drivers low in emotional stability tended to have distrustful attitudes and lower intentions to use AVs.

To summarize, individual differences (i.e., demographics and personality) have significant impacts on attitudes toward and adoption of AVs. However, the literature offers little insight into the role of individual differences associated with AV expectations and AV adoption although a strong link between the two has been found [184, 92, 157].

2.3 Method

To examine whether individual differences impact AV expectations, we conducted a nationwide representative survey with 443 participants using an online survey platform. This research complied with the American Psychological Association code of ethics and was approved by the university's institutional review board. All participants provided informed consent.

2.3.1 Survey Instrument and Respondents

We conducted a survey using an online survey platform in the period of July–September 2018. We developed and distributed a questionnaire for the target study population on www.qualtrics.com. Each respondent's key demographics and individual differences including age, gender, race and ethnicity, education, income, marital status, census region, frequency of driving, driving experience, and personality, along with his/her expectations of AVs, were collected for further analysis. All recorded information was anonymous.

The study population targeted U.S. drivers because the United States is one of the countries expected to be an early adopter of AVs in hopes of achieving greater safety and environmental benefits [37, 159]. To select a representative sample of U.S. drivers, we used the Qualtrics Online Sample tool to recruit participants [11]. There were four steps to obtain a research sample that could represent the characteristics of U.S. drivers. The first step, according to the 2014–2015 statistics of the U.S. Department of Transportation and the AAA Foundation [182], was to collect the percentages of subpopulations based on demographics involving age (18+), gender, region, and ethnicity. For the next step, we calculated the study sample size based on a 95% confidence level, 5% confidence interval, 50% population proportion, and total number of U.S. drivers (225 million), which is driven by the population's size reported by the U.S. Census Bureau. The third step was to calculate the numbers of qualified U.S. drivers based on each studied demographic percentage and sample size and to provide it to the Qualtrics sample service. Finally, the Qualtrics online sample service selected and included participants within the bounds of the criteria by the embedded screener and collected data from a random and representative sample.

A total of 443 respondents filled out the survey completely. The Qualtrics Online Sample service filtered out partial responses and participants who declined to consent to the study process. Each qualified respondent was paid $3 for their participation and responses. A demographic breakdown for the respondents is presented in Table 2.1.

Table 2.1: Demographic information on study participants

Demographic Characteristics		Population Percentage	Sample Percentage	Number
Age	Younger Driver (18-54)	63.8%	63.7%	282
	Older Driver (55+)	36.2%	36.3%	161
Gender	Male	49.0%	49.0%	217
	Female	51.0%	51.0%	226
Region	Northeast	17.0%	16.5%	73
	West	24.0%	23.5%	104
	South	38.0%	40.1%	178
	Midwest	21.0%	19.6%	87
Ethnicity	White non-Hispanic	65.3%	65.2%	289
	Black/African American non-Hispanic	14.1%	14.0%	62
	Hispanic	14.8%	14.9%	66
	Other	5.8%	5.8%	26
$Education^a$	High school/GED	39.0%	23.6%	97
	Some college	21.0%	24.3%	100
	College graduate	40.0%	52.1%	214
$MaritalStatus^b$	Never married	23.5%	32.3%	143
	Married	51.7%	42.0%	186
	Living with partner	6.5%	5.4%	24
	Widowed	7.5%	4.5%	20
	Divorced/separated	10.8%	15.8%	70
Income	Low Income ($35k/yr)	N/A	25.1%	111
	Medium Income ($35k-$100k/yr)	N/A	62.8%	278
	High Income ($100k/yr)	N/A	12.2%	54
Driving Frequency	Drives almost every day	68.5%	84.0%	372
	Drives sometimes or rarely	18.1%	16.0%	71
$DrivingExperience^c$	Low (12 yrs)	N/A	21.0%	93
	Medium (13-45 yrs)	N/A	58.0%	257
	High (46 yrs)	N/A	20.3%	90
Extraversion	Solitary/reserved	N/A	57.8%	256
	Outgoing/energetic	N/A	42.2%	187
Agreeableness	Challenging/detached	N/A	43.6%	193
	Friendly/compassionate	N/A	56.4%	250
Conscientiousness	Efficient/organized	N/A	40.6%	180
	Easy-going/careless	N/A	59.4%	263
Emotional Stability	Sensitive/nervous	N/A	48.3%	214
	Secure/confident	N/A	51.7%	229
Openness to Experience	Consistent/cautious	N/A	53.7%	238
	Inventive/curious	N/A	46.3%	205

Note. a: Education showed only for respondents age 24+. b: Marital status is shown only for respondents age 18+. c: Driving experience excluded four invalid answers.

2.3.2 Dependent Variable

To measure people's expectations of AVs, we used a questionnaire developed by Van Ryzin (2004) with a 7-point Likert rating scale ranging from 1 (low) to 7 (high) [185]. There were three items in the questionnaire: (1) How would you rate your overall expectations regarding the driving of a self-driving car?; How would you rate your expectations regarding the effectiveness of a self-driving car?; and (3) How would you rate your expectations regarding the safety of a self-driving car?

2.4 Results

Our overall objective was to use AV expectations to help identify potential barriers to the adoption of AVs by specific subgroup populations. Therefore, our analysis was designed to detect subgroup population differences based on demographics across our representative sample. Contrary, we did not attempt to build a general predictive model which might fail to detect subgroup differences in smaller subgroups. Nonetheless, our results may help others explore and build valid predictive models.

To accomplish our overall objective, a statistical analysis was performed using IBM SPSS Statistics software. One-way analysis of variance (ANOVA) was used to examine the potential differences in the expectation of AVs based on individual differences. The alpha level was set at 0.05 for all statistical tests. All post hoc comparisons utilized a Bonferroni alpha correction. The construct reliability of the initial expectation, 0.94, was above the acceptable threshold of 0.70. Also, we applied convergent validity through exploratory factor analysis to determine whether this measurement construct was valid. All three items loaded above the 0.7 thresholds. The summary of responses is shown in Table **??**.

2.4.1 Age and Expectations of AVs

Respondents were divided into two age groups: younger (18–54 years) and older drivers (55 years and older) based on previous age categorizations [98, 189]. ANOVA revealed a statistically significant effect of people's age on their expectations of AVs ($F = 208.856$, $p < 0.001$, $\eta^2 = 0.106$). Compared to older drivers (mean = 3.04, standard deviation [SD] = 1.86), younger drivers (mean = 4.40, SD = 1.93) tended to have higher expectations of AVs.

Table 2.2: Study results

Demographic Characteristics		Number	Expectations			
			Mean	St. dev.	F	*p value*
Age*	Younger Driver (18-54)	282	4.40	1.93	208.856	0.001
	Older Driver (55+)	161	3.04	1.86		
Gender*	Male	217	4.31	1.98	69.818	0.001
	Female	226	3.52	1.96		
Region	Northeast	73	3.93	2.16	0.657	0.578
	West	104	3.79	2.03		
	South	178	3.96	1.94		
	Midwest	87	3.91	2.02		
Ethnicity*	White non-Hispanic	289	3.68	1.99	14.944	0.001
	Black/African American non-Hispanic	62	4.40	1.94		
	Hispanic	66	4.32	2.07		
	Other	26	4.21	1.83		
Education*	High school or GED	97	3.66	2.00	5.192	0.001
	Some college	100	3.58	2.09		
	College graduate	214	3.98	1.95		
Marital Status*	Never married	143	4.47	1.84	20.319	0.001
	Married	186	3.74	2.06		
	Living with partner	24	3.92	2.21		
	Widowed	20	3.25	1.80		
	Divorced/separated	70	3.37	1.94		
Income	Low Income ($35k/yr)	111	3.73	2.01	2.659	0.070
	Medium Income ($35k-$100k/yr)	278	3.94	2.01		
	High Income ($100k/yr)	54	4.07	1.98		
Driving Frequency*	Drives almost every day	372	3.97	2.02	10.411	0.001
	Drives sometimes or rarely	71	3.55	1.93		
Driving Experience*	Low (12 yrs)	93	4.94	1.68	119.065	0.001
	Medium (13-45 yrs)	257	3.89	2.03		
	High (46 yrs)	90	2.79	1.64		
Extraversion*	Solitary/reserved	256	3.76	1.96	12.328	0.001
	Outgoing/energetic	187	4.10	2.06		
Agreeableness*	Challenging/detached	193	3.64	2.13	22.256	0.001
	Friendly/compassionate	250	4.10	1.89		
Conscientiousness*	Efficient/organized	180	3.59	2.16	44.699	0.001
	Easy-going/careless	263	4.12	1.88		
Emotional Stability*	Sensitive/nervous	214	3.79	2.09	5.846	0.016
	Secure/confident	229	4.02	1.93		
Openness to Experience	Consistent/cautious	238	3.88	1.90	0.250	0.617
	Inventive/curious	205	3.93	2.13		

Note. Asterisks mark the demographic variables that have significant effects on expectations of AVs.

2.4.2 Gender and Expectations of AVs

Gender was significant (F = 69.818, $p < 0.001$, $\eta^2 = 0.038$). Male drivers (mean = 4.31, SD = 1.98) had higher expectations of AVs than female drivers (mean = 3.52, SD = 1.96).

2.4.3 Geographic Region and Expectations of AVs

Expectations were not significantly different among drivers of different regions (e.g., Northeast, West, South, Midwest) (F = 0.657, p = 0.578, $\eta^2 = 0.001$). Drivers who lived in the West had the lowest mean (mean = 3.79, SD = 2.03) compared to drivers in the other three regions.

2.4.4 Ethnicity and Expectations of AVs

Ethnicity was significant (F = 14.944, $p < 0.001$, $\eta^2 = 0.025$). As illustrated in Figure 2.1, post hoc analysis indicated that White non-Hispanic participants tended to have lower expectations than people in the other three ethnic groups: White non-Hispanic (mean = 3.68, SD = 1.99) vs. Black/African American non-Hispanic (mean = 4.40, SD = 1.94, $p < 0.001$); White non-Hispanic vs. Hispanic (mean = 4.32, SD = 2.07, $p < 0.001$); White non-Hispanic vs. Other (mean = 4.21, SD = 1.83, p = 0.047). There was no significant difference among the other three groups.

2.4.5 Education and Expectations of AVs

Educational level was also significant (F = 5.192, p = 0.001, $\eta^2 = 0.009$). As illustrated in Figure 2.2, post hoc analysis revealed that college-educated drivers had higher expectations of AVs (mean = 3.98, SD = 1.95) than drivers with some college education (mean = 3.58, SD = 2.09, p = 0.002) and those with high school/GED or less education (mean = 3.66, SD = 2.00, p = 0.022). However, there was no difference between high school graduates and those with some college.

2.4.6 Marital Status and Expectations of AVs

As shown in Figure 2.3, there was a significant effect of drivers' marital status on their expectations of AVs (F = 20.319, $p < 0.001$, $\eta^2 = 0.044$). Post hoc comparisons revealed that drivers who had not married had higher AV expectations (mean = 4.47, SD = 1.84) than married drivers (mean = 3.74, SD = 2.06, $p < 0.001$), widowed drivers (mean = 3.25,

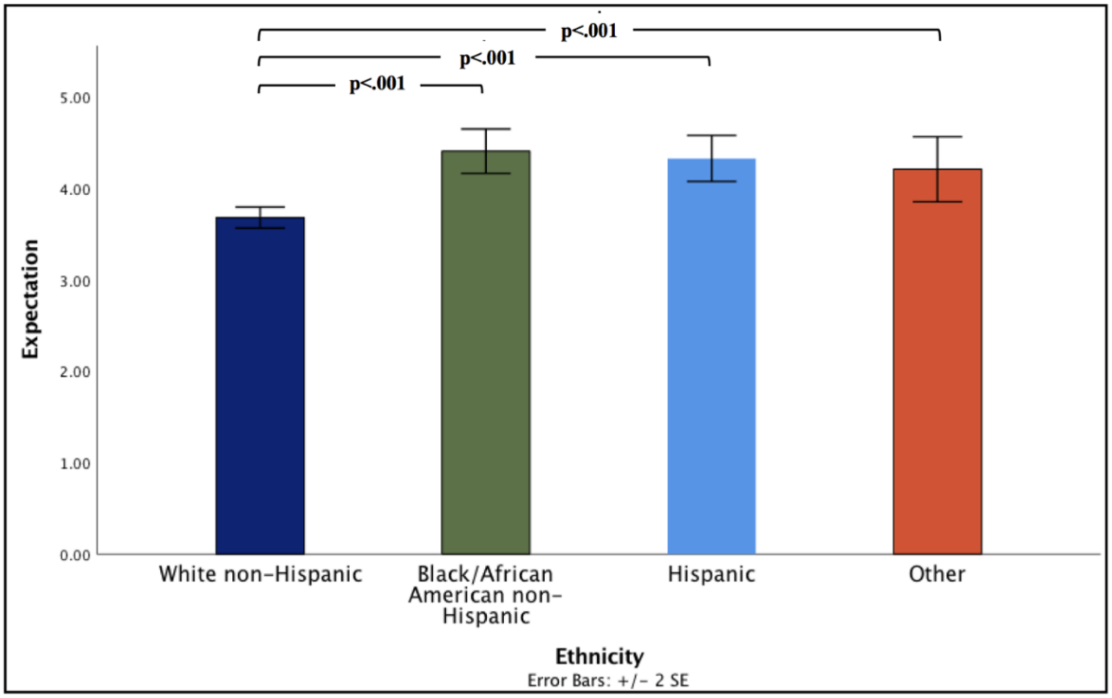

Figure 2.1: Summary of the responses, by ethnicity, to people's expectations of AVs

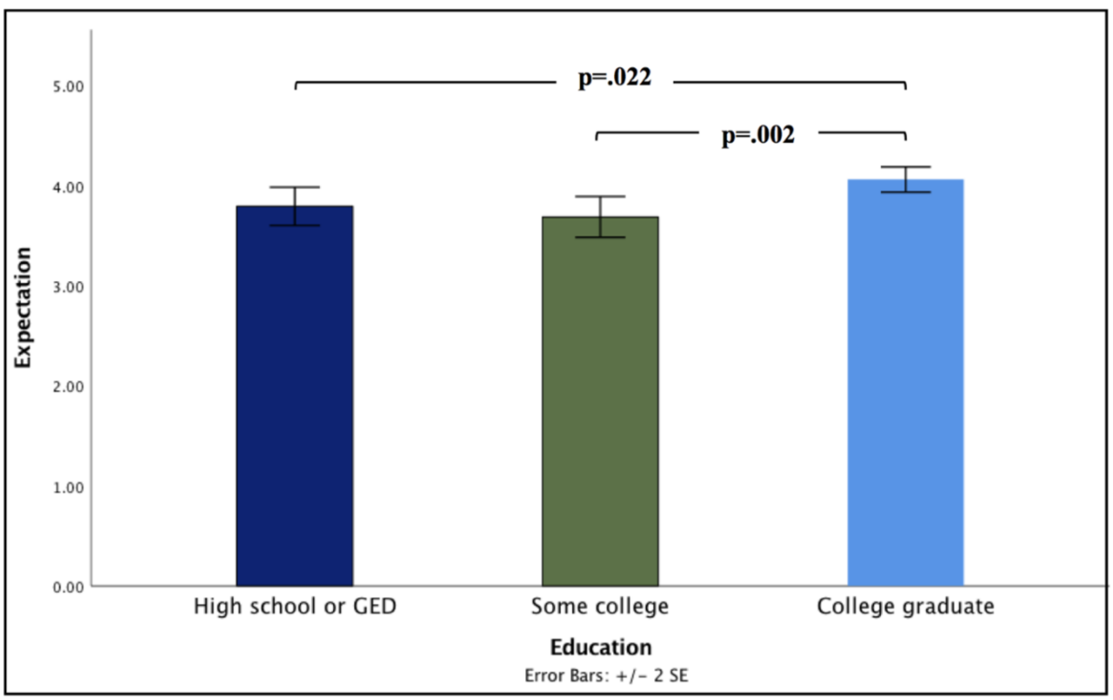

Figure 2.2: Summary of the responses, by education, to people's expectations of AVs.

SD = 1.80, $p < 0.001$), and divorced/separated drivers (mean = 3.37, SD = 1.94, $p < 0.001$). There was no difference between never-married drivers and drivers who lived with a partner.

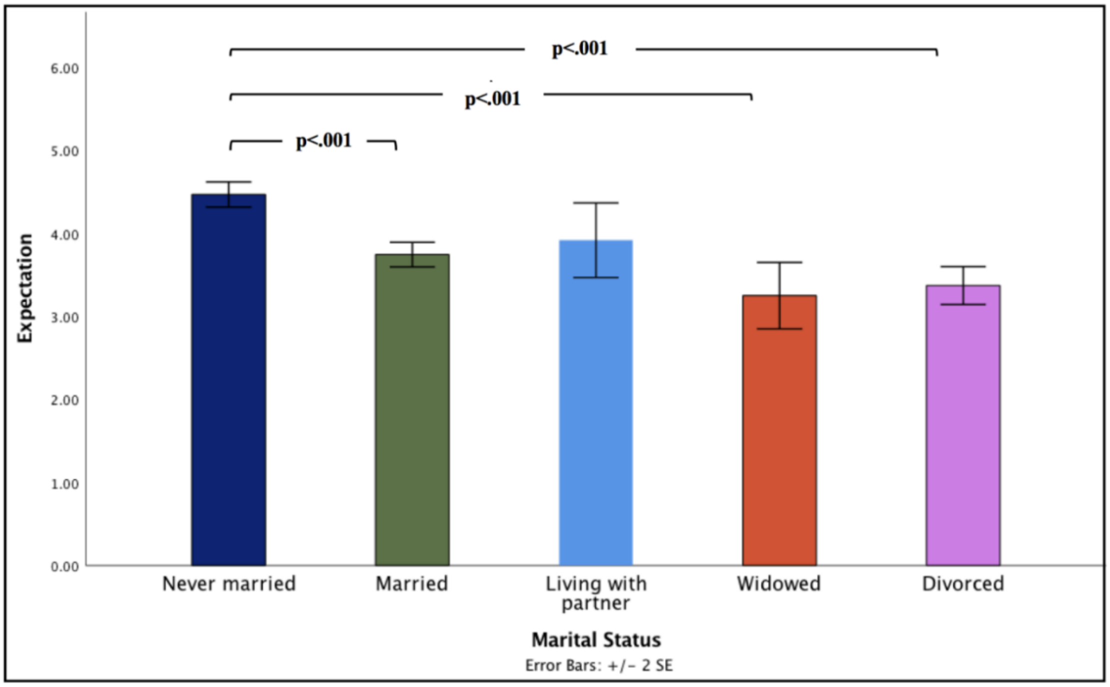

Figure 2.3: Summary of the responses, by marital status, to people's expectations of AVs.

2.4.7 Income and Expectations of AVs

Participants were divided into three groups by income (i.e., low: \$0–\$34,999; medium: \$35,000–\$99,999; high: \$100,000 or more). There were no significant differences based on income (F = 2.659, p = 0.070, $\eta^2 = 0.003$). However, lower-income drivers had the lowest expectations of AVs (mean = 3.73, SD = 2.01) and high-income drivers had the highest expectations of AVs (mean = 4.07, SD = 1.98).

2.4.8 Driving Frequency and Expectations of AVs

Participants were divided into two groups according to driving frequency (i.e., sometimes or rarely drive; drive almost every day). There was a significant effect of driving frequency on expectations (F = 10.411, p = 0.001, $\eta^2 = 0.006$). Drivers who sometimes or rarely drove had lower expectations of AVs (mean = 3.55, SD = 1.93) than people who drove almost every day (mean = 3.97, SD = 2.02).

2.4.9 Driving Experience and Expectations of AVs

Drivers were divided into three groups based on the mean and one standard deviation of their driving experience. The three groups were low (have driven equal to or less than 12 years), medium (have driven more than 12 years and less than 45 years), and high (have driven more than 45 years). Drivers with less than 12 years of driving experience had higher expectations of AVs than the other two groups ($F = 119.065$, $p < 0.001$, $\eta^2 = 0.119$): low (mean = 4.94, SD = 1.68) as shown in Figure 2.4.

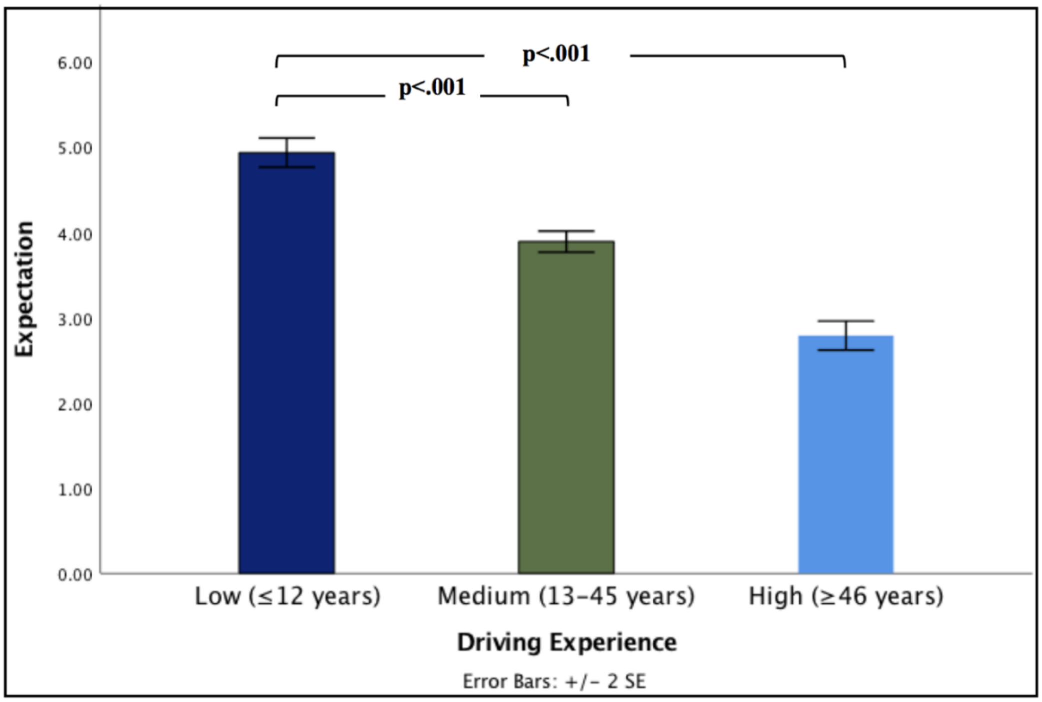

Figure 2.4: Summary of the responses, by driving experience, to people's expectations of AVs.

2.4.10 Personality and Expectations of AVs

The Big Five personality traits scores were divided into two groups consisting of high or low scores based on their means. Scores above the mean were classified as high and those below the mean were classified as low. ANOVA indicated that there were significant effects of extraversion ($F = 12.328$, $p < 0.001$, $\eta^2 = 0.007$), agreeableness ($F = 22.256$, $p < 0.001$, $\eta^2 = 0.012$), conscientiousness ($F = 44.699$, $p < 0.001$, $\eta^2 = 0.025$), and emotional stability ($F = 5.846$, $p = 0.016$, $\eta^2 = 0.003$) on expectations. Drivers who were high in each of those personality traits had higher expectations than those who were lower in each personality trait. However, there was no significant difference between people who were high and low in openness to experience (F = 0.250, p = 0.617, $\eta^2 = 0.000$).

2.4.11 Summary of the Results

The findings of this paper can be organized into two overarching results. One, we found significant effects of demographic factors. Results revealed that drivers' expectations of AVs differ greatly by age, gender, ethnicity, education level, marital status, driving frequency, and driving experience. More specifically, drivers who were younger, male, non-White non-Hispanics, more highly educated, never married, with a higher drive frequency and with less driving experience were prone to have higher expectations of AVs. Two, this study provides evidence that personality traits do impact AV expectations. In general, drivers who were high in extraversion, agreeableness, conscientiousness, and emotional stability revealed higher expectations of AVs. The next section provides a detailed discussion of the findings and their contributions to the literature.

2.5 Discussion

The goal of this research was to understand whether AV expectations differ by individual differences. Results of this study highlight the significant differences in AV expectations based on age, gender, race and ethnicity, education level, marital status, driving frequency, driving experience, and personality traits. Results of this study can also provide future research with a rich set of factors to explore when predicting AV expectations.

Our results contribute to the literature in the following ways. First, our findings that AV expectations differ among individuals in a representative sample of U.S. drivers highlight the importance of and extend the literature on individual differences in AV expectations. Prior research discussed individual differences related to public concerns and opinion regarding AVs, and to the best of our knowledge, only one paper discussed expectations related to individual differences; in that paper, men and drivers with higher educational levels had higher expectations of AVs [168]. Results of our study confirm this assertion that male drivers and those with higher education levels tend to have higher expectations. Further, our study adds to the literature by uncovering other key individual difference factors. More specifically, drivers who were older, female, White non-Hispanic, or rarely involved in driving did not have higher expectations of AVs, whereas drivers who were younger, male, more highly educated, never married, had a higher driving frequency, had less driving experience, and were high in extraversion, agreeableness, emotional stability and conscientiousness were prone to have higher expectations of AVs. In all, the results of this study provide new insights into how individual differences can influence expectations, which act as a predictor of attitudes and behaviors around AVs, including trust, satisfaction, and adoption intention.

Second, the results of this study provide new insights into the relationships between individual differences and AV acceptance. Based on the technology adoption theories (i.e., TAM and UTAUT) and expectation-confirmation theory (ECT), expectations are one determinant of technology adoption [72, 78, 118]. Expectations that are too low can discourage individuals from ever using a technology, which leads to low technology adoption. However, expectations that are too high can lead to disappointment by setting the bar too high and creating a greater discrepancy between expectations and performance, which will also lead to low technology adoption [94, 96]. In other words, people are less likely to adopt AVs if their expectations are too low but they are also less likely to continue to use them if their expectations are too high and the AV fails to meet those expectations. By understanding which groups are likely to have low or high expectations, we can begin to design interventions to address these issues.

Results of this study found that older drivers tend to have lower expectations compared to younger drivers. This aligned with and helped explain prior literature that suggested that older drivers have negative attitudes toward AV adoption [168, 65, 130]. Because older adults have lower expectations and less interest in driving with AVs, they are less likely to adopt AVs. Similarly, prior literature revealed that male drivers and those who have higher educational levels have a higher acceptance of AVs [145, 168]. Our results support and explain this assertion by highlighting the higher expectations that men and people with a higher education level have of AVs. Our results could also explain the relationships between driving experience and AV adoption. Prior research suggested a negative relationship between driving experience and AV adoption [88]. Our study's findings align with prior research by suggesting that less experienced drivers have higher expectations of AVs.

Our results also found significant effects of extraversion, agreeableness, conscientiousness, and emotional stability on AV expectations. Specifically, our findings suggest that people who are high in these personality traits are prone to have higher expectations of AVs. Previous literature found that drivers with low emotional stability tend to have negative attitudes and lower intentions to use AVs [91, 205]. Our findings aligned with this assertion by suggesting a positive relationship between emotional stability and expectations. To summarize, expectations that are too high or too low can prevent people from having positive attitudes and adopting AVs. Individual differences, including demographics and personality traits, are critical factors to consider because of their impacts on drivers' expectations of AVs. That being said, understanding the effects of individual differences on expectations can help predict and support AV design and adoption.

Finally, this study emphasizes the importance of finding influential factors that can impact AV expectations in terms of expectation calibration. This study provides evidence and examples of how groups with different traits (i.e., individual differences) react differently in terms of their initial expectations, which serves as the baseline for calibration. For example, results showed that men have higher expectations of AVs compared to women. To encourage both men and women to drive with an AV and decrease disappointment, some effective measures could be done to calibrate their expectations of AVs based on their different initial expectations. Therefore, the results of this study demonstrate the need to account for individual differences in AV expectations.

2.6 Limitations and Future Research

Our study has several limitations. First, this study focused on examining drivers' demographics and personality traits. However, individual differences could be found in all psychological characteristics, physical and mental abilities, knowledge, habits, personality, and character traits [196, 160]. These include, for instance, cultural differences, religious differences, and motor ability differences. Future studies should examine the relationships between these attributes and expectations of AVs. Second, this study targeted U.S. drivers. Future research could focus on different populations from various countries and investigate whether the results can be replicated. Third, there were large differences in the sample size across groups, which might limit our interpretations of the comparisons across groups. Fourth, this was a cross-sectional observational study, which allowed us to examine whether there were differences but not why there were differences. Future experimental studies could examine causal relationships between individual traits and AV acceptance along with other potential mediation mechanisms. Researchers might also wish to investigate the relationship between AV expectations and AVs' actual adoption, and the relationship between individual differences and different aspects of expectation (i.e., safety and effectiveness expectations). Fifth, individual difference variables were grouped to summarize and test for differences in AV expectations among groups. Despite reaching statistical significance, the effect sizes of education, income, driving frequency, extraversion, emotional stability were quite small. In addition, although detecting subgroups' differences is empirically loosely related to developing a predictive model of AV expectations, it is not clear that the data set we have collected allows us to make strong causal inferences to build valid predictive model. Further research is needed to make causal claims regarding just why these particular demographics and traits were significant. We hope that the results of this paper highlight future directions that others can explore to build valid predictive models. Finally, this study did not consider the participants' previous experiences with AVs when investigating the relationship between individual differences and AV expectations. Although the level 5 fully automated vehicles are not available on the market, participants might have experienced AV-related technology (e.g., AV simulator and virtual AV platforms) before participating in the survey. Individuals who had prior experiences could have different AV expectations from those who had no experience. While we assume all subjects had roughly similar AV experience, future research could investigate whether previous AV-related technology experience influences people's expectations of AVs. In all, more research is needed to investigate AV expectations.

2.7 Conclusion

This study examined U.S. drivers' expectations of AVs from the perspective of individual differences. The findings in this study emphasize the importance of individual differences, including demographics and personality, on understanding expectations of AVs. More specifically, higher expectations are more often generated by drivers who are younger, male, non-White non-Hispanic, with higher education, never married, with a higher frequency of driving, less driving experience, and who are high in extraversion, agreeableness, emotional stability, and conscientiousness. The results of this study provide a basis for conducting future research related to expectations and AVs. Our results have important implications on the future design of AVs.

CHAPTER 3

Expectations and Trust in Automated Vehicles

3.1 Introduction

Trust, or lack thereof, in Automated Vehicles (AVs) presents a significant barrier to their widespread acceptance and smooth integration into existing transportation systems. While research into trust in AVs is extensive, there remains a dearth of understanding regarding the influence of expectations in shaping this trust. Expectation Disconfirmation Theory (EDT) sheds light on this dynamic, suggesting that users compare their expectations with actual experiences, leading to a disconfirmation that influences their attitudes and technology acceptance decisions. Further, trust in AVs is intimately linked with individuals' perception of risk while driving, which encompasses elements of uncertainty and the potential severity of outcomes.

In an effort to more comprehensively understand the intricate relationship between expectations, disconfirmation, and trust in AVs, and to further assess the potential moderating role of risk perception, we implemented an online survey that engaged 443 US drivers. Our research approach included an examination of an external risk factor - weather, and an internal risk factor - driving behaviors of AVs. Our results reveal a complex interaction between expectations, disconfirmation, risk perception, and trust in AVs. They underscore the significance of expectations in the cognitive appraisal process and illuminate how these expectations can either foster or erode trust in AVs. If we aim to accelerate the acceptance and adoption of AVs, it is of paramount importance to comprehend these interactions and devise effective strategies to build trust in this innovative technology. The interwoven dynamics of expectations, experience, and risk perception provide fertile ground for such trust-building strategies.

3.2 Background

3.2.1 Trust in AVs and Expectation Disconfirmation Theory

3.2.1.1 Trust in AVs

Trust is a crucial factor in the realm of automation, as evidenced by extensive prior research. It refers to the willingness of one party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action that is essential to the trustor, irrespective of the ability to monitor or control that other party [120, p. 712]. In the context of automation, when operators trust automated systems, they are more likely to assign control to them [39, 103, 131]. Moreover, trust is a vital component in the relationship between operators and new technology, as people may struggle to accept and utilize novel technologies due to a lack of trust. However, excessive trust or mistrust in new technology can also be problematic, as it may lead to failure to intervene when the technology falters [102].

The successful integration of AVs into transportation systems relies heavily on trust. Various models and theories have been developed to understand how trust affects people's intentions to use AVs. For example, the Technology Acceptance Model (TAM) demonstrates how trust significantly influences perceived usefulness, with both factors impacting the behavioral intention to adopt AVs [32]. Research has also shown that trust positively affects drivers' attitudes towards using AVs [205]. The formation and development of trust in AVs are complex, and researchers have conducted extensive studies to identify the factors that contribute to its basis. For instance, Oliveira (2020) [147] evaluated drivers' trust in various configurations of interfaces within a fully automated vehicle that communicated driving status and intended behavior. The study found that animate representations and augmented reality interfaces scored higher on three trust bases: familiarity, reliability, and confidence. These trust bases align with the trust theory defined by Sheridan (1989) [173]. Similarly, Gold et al. (2015) [58] found that driving experience increased self-reported trust in AVs, and older participants rated AVs more highly than younger drivers in terms of all trust components. These components include reliability, competence, responsibility, predictability, dependability, and faith and align with the trust dimensions proposed by Barber (1983) and Rempel et al. (1985) [6, 156].

In conclusion, trust is a critical component in AV acceptance and utilization. Various models and theories have been developed to understand how trust influences people's intention to use AVs, and multiple factors contribute to trust formation and development in AVs. Further research is necessary to improve drivers' trust in and acceptance of AVs by investigating the factors that affect trust in AVs.

3.2.1.2 Expectation and Expectation Disconfirmation Theory (EDT)

Expectations significantly impact attitudes toward and adoption of technology. Expectations refer to individuals' beliefs about the future performance of technology [12]. These beliefs are shaped by the quality and reliability of the information received from various sources, such as advertising, media reports, media interviews, and interpersonal communication [201, 183]. Research has shown that expectations are a critical factor in the acceptance and adoption of AVs, as individuals are more likely to accept and drive with AVs if their expectations are higher [78, 184]. Positive expectations also lead to positive attitudes and reliance on AVs [92, 157].

The Expectation-Disconfirmation Theory (EDT) is a psychological theory that explains how expectations influence individuals' attitudes and behaviors towards a particular product or service by comparing their expectations with their actual experiences [36, 148, 187]. Performance, in the context of EDT, refers to an individual's beliefs about the technology's performance during a usage period [95, 20]. Disconfirmation refers to the extent to which technology performs either better or worse than initially expected on those attributes [10]. Disconfirmation can be positive when a technology's performance exceeds expectations or negative when a technology's performance falls short of expectations [148, 72].

EDT has been applied in various fields, including marketing and consumer behavior [123, 87], psychology [154], information systems [176], and service quality [79]. For example, Mellers et al. (1997) [123] found that individuals were more satisfied with a positive outcome when a negative outcome was expected and more dissatisfied with a negative outcome when a positive outcome was expected. Staples et al. (2002) [176] discovered that individuals with unrealistically high expectations had lower perceptions of system effectiveness and satisfaction compared to those with accurate or low expectations.

In summary, EDT suggests that individuals form expectations, compare them with actual experiences, and generate disconfirmation. Attitudes and behaviors towards a particular product or service are further influenced based on this comparison. However, the impact of such cognitive appraisals on trust in the context of AVs remains to be explored.

3.2.1.3 Expectation Disconfirmation Theory (EDT) and Trust

Trust is a crucial variable in many domains, yet its connection to the Expectation-Disconfirmation Theory (EDT) has not been extensively examined. Trust-building in trust research is often described as an expectation-disconfirmation process [94, 107, 108], and the use of terms such as "expect" and "expectations" in defining trust further supports this link. For example, Zaheer [200] defines trust as the expectation that an actor is trustworthy, predictable, and fair, while Zucker [207] defines it as a set of expectations shared by all those involved in an exchange. Trust theory suggests that trust grows when it is positively confirmed but declines when it is negatively disconfirmed [132, 107, 108]. As these definitions and theories indicate, trust progression is determined by disconfirmation of trust-related expectations, making it closely linked to EDT [107, 162, 164].

Lankton's (2012) research [95] applied EDT to understand the trust people have in technology. The study discovered that disconfirmation enhances trust as it becomes more positive, while negative disconfirmation decreases trust. Furthermore, negative disconfirmation has a more significant negative impact on trust than positive disconfirmation has a positive effect. However, additional research is needed to explore how such cognitive appraisal, through the lens of EDT, affects trust in the context of AVs.

3.2.1.4 Analytical Advances in Expectation Disconfirmation Theory (EDT)

The concept of disconfirmation is a critical aspect of the EDT and is determined by two key components: expectations and performance perception. There are two ways to calculate the level of disconfirmation: the first is by computing the difference between expectation and performance (Difference scores), while the second is by directly observing the degree of disconfirmation. However, the first approach has some limitations, such as individuals' unawareness of whether the calculated disconfirmation is associated with expectation and perceived performance or only one of them. Moreover, the results simply reveal the three-dimensional relationship and untested constraints are imposed on the congruence equations [45, 46, 187]. To avoid the issues associated with differences in scores, a direct measurement of disconfirmation was developed. However, this method also has its limitations. The problem of oversimplification is still unresolved since the direct measure and outcome measure are related in a two-dimensional way, unlike the three-dimensional relationship revealed by the first approach. Additionally, the direct measurement of disconfirmation produces a unidirectional measure that does not differentiate between positive and negative disconfirmation [45, 46].

Previous research has proposed the polynomial model and response surface methodology as effective means to understand the complex three-dimensional relationship between expectations, perceived performance, and outcomes (e.g., satisfaction and trust). These methods address the limitations associated with difference scores and direct measurements in examining disconfirmation levels in the EDT. The polynomial model, based on the theoretical model $Z = f(X,Y)$, can detect the relationship between component and outcome measures by examining curvilinear terms [45, 46]. However, the results generated by polynomial modeling can be difficult to interpret. Therefore, response surface methodology is used to present a set of visual and statistical tests to better demonstrate how these results describe the surfaces they imply. Response surface methodology emphasizes three primary characteristics of surfaces generated from polynomial models: a stationary point, principal axes, and slopes along various lines of interest. This approach provides a more comprehensive understanding of the contours and details of the plotted surface, helping researchers to better comprehend the three-dimensional relationship between expectations, perceived performance, and outcomes in the EDT [45, 46, 187].

3.2.2 Perceived Risk and Trust in AVs

Perceived risk is the subjective belief that pursuing a desired outcome carries a chance of experiencing negative consequences [23]. In the context of driving, perceived risk is shaped by evaluating the likelihood and severity of potential accidents, with factors such as adverse weather, high-speed driving, and abnormal vehicle behavior influencing a driver's risk perception [110].

In the realm of trust, perceived risk is a fundamental aspect, as trust inherently involves elements of vulnerability and uncertainty arising from unpredictable circumstances and associated risks [103]. This interplay between vulnerability, uncertainty, and perceived risk makes the latter a crucial component of trust. Numerous studies propose a causal chain in which perceived risk informs trust, which subsequently impacts acceptance [204, 112, 120, 128]. Mayer et al. (1995) [120] posits that trust is tied to the perceived risk of a specific outcome, and Mitchell (1999) [128] contends that risk predicts trust. Trust hinges on the willingness of the involved parties to embrace vulnerability and acknowledge varying degrees of risk. Empirical evidence corroborates this relationship, as research has revealed that perceived safety risks negatively affect the acceptance of AVs [204, 4]. Hebert et al. (2021) [4] further underscore the substantial impact of both internal and external risks on trust in automated driving systems (ADS).

In summary, perceived risk is a critical factor in understanding trust in automation, as it significantly influences trusting behavior and the acceptance of automated systems. Recognizing the importance of both internal and external risks is essential for fostering trust over time.

3.3 Hypothesis Development

Our hypothesis development draws upon the Expectation Disconfirmation Theory (EDT) and trust theory. EDT suggests that individuals form initial performance expectations, serving as a reference point for evaluating perceived performance after an experience. This cognitive comparison is known as disconfirmation. EDT predicts that when disconfirmation is more positive (performance exceeds expectations), individuals experience greater pleasure, leading to increased positive attitudes and intention to use. Conversely, when disconfirmation is more negative (performance falls short of expectations), individuals experience displeasure, resulting in negative attitudes.

Trust theory posits that the cognitive appraisal of disconfirmation affects trusting intention. When AV performance exceeds expectations, positive disconfirmation occurs, potentially encouraging individuals to trust the system more. Previous research has shown that users are more likely to explore and utilize an information system when initial expectations are significantly surpassed [75]. This tendency may stem from users investigating various features to enhance productivity beyond their initial expectations, thereby achieving equilibrium between expectations and performance through increased trust [95, 17]. Conversely, negative disconfirmation arises when perceived AV performance falls short of expectations, which may lead to a reduction in trusting intention. Prior studies have demonstrated that users are less likely to trust e-commerce websites when performance does not meet expectations, as the resulting disappointment diminishes the perception of trustworthiness [17]. Such negative trust disconfirmation can render relationships unsustainable and expensive to mend. Consequently, we predict that positive disconfirmation will promote trust in AVs, while negative disconfirmation will hinder it.

H1A: *Positive disconfirmation will positively influence trust in AVs.*
H1B: *Negative disconfirmation will negatively influence trust in AVs.*

Operating a vehicle inherently involves risks stemming from both external and internal factors, especially in diverse driving situations. External risks are tied to uncertainties, such as adverse weather conditions, which significantly influence risk perception. In the United States, weather-related risks contribute considerably to high traffic accident rates and disrupted traffic flow [80]. Factors like slippery roads and reduced visibility heighten these risks, making safe navigation of roads challenging for drivers [47]. Moreover, internal characteristics within the vehicle also affect the perceived risk of driving. Aggressive driving styles can impair a driver's performance and raise the likelihood of accidents. Such behavior encompasses unsafe and hostile driving practices, disregarding other drivers, and includes actions like frequent or unsafe lane changes, failure to signal or yield, tailgating, and ignoring traffic controls [127].

Extensive research has established a sequential relationship between trust and perceived risk, showing that trust formation is influenced by the level of perceived risk in a specific environment or interaction. Generally, as perceived risk increases in technology engagement, trust decreases. For instance, Corbitt et al. (2003) found that users' perceived risk negatively affected their trust in online banking systems [29]. Various risk factors, including privacy, security, product pricing, and customer service risks, have been identified as trust determinants. Lower levels of these risks led to increased trust in e-commerce platforms [81]. Moreover, prior studies confirmed that perceived privacy risks negatively impact trust in electronic transactions, emphasizing the role of perceived risk as a precursor to trust [34]. In the context of AVs, individuals become vulnerable to potential performance issues when they allow an AV to assume control of driving tasks. Risk is an intrinsic factor that influences trust. As risk levels decrease, higher trust has been observed in AV research [4, 81, 146, 151]. In other words, the AV literature supports the negative relationship between risk perception and trust formation. Therefore, it is crucial to examine factors related to risks in AVs when assessing trust. In conclusion, including risk perception is vital for a comprehensive evaluation of trust in human-vehicle interaction.

We propose that both external and internal risk factors, like weather and AV driving behavior, can affect the relationship between disconfirmation and trust in AVs. Severe weather conditions can significantly impact trust in AV systems, as drivers may doubt the AV's ability to accurately assess the situation and make appropriate decisions. This uncertainty leads to a decrease in their willingness to be vulnerable and trust the AV system during inclement weather compared to fair weather. Internal risk factors, such as AV driving behavior, can also substantially impact trust. For instance, an AV exhibiting aggressive driving behavior may be perceived as less capable and safe by drivers, resulting in a decline in confidence in the system and further diminishing trust. Therefore, we hypothesize that both external and internal risk factors affect the relationship between disconfirmation of expectation and performance and trust in AVs. Specifically, we suggest that severe weather conditions will lead to reduced trust in AVs. Similarly, AVs displaying aggressive driving behavior will also undermine trust in AVs.

H2*: The impact of disconfirmation on driver's trust in AVs can be moderated by weather and AV driving behavior, with snowy weather and aggressive driving resulting in reduced trust compared to sunny weather and normal driving behavior, respectively.*

3.4 Method

We obtained data from 443 participants through a nationwide survey using the web-based survey tool Qualtrics. This study adhered to the ethical guidelines outlined by the American Psychological Association and was approved by the Institutional Review Board of the University of Michigan. All participants provided informed consent before participating.

3.4.1 Survey Instrument and Respondents

In July and September 2018, we conducted an online survey via www.qualtrics.com to investigate the attitudes and perceptions of U.S. drivers towards AVs. To ensure a representative sample, participants were screened for demographic characteristics associated with the driver population in the United States. The collected data was anonymous.

We followed a four-step process using Qualtrics Online Research Service to select a representative sample of U.S. drivers. First, we collected the percentages of subpopulations by demographic variables such as age (18+), gender, region, and ethnicity, based on 2014–2015 statistics from the U.S. Department of Transportation and the AAA Foundation [182]. Second, we calculated the study sample size using a confidence level of 95%, a confidence interval of 5%, a 50% population proportion, and the U.S. Census Bureau's population size of 225 million. Third, we calculated the number of qualified U.S. drivers based on the studied demographic percentage and sample size and provided this information to Qualtrics. Finally, Qualtrics' online sample service collected a random and representative sample of respondents based on the criteria selected by an embedded screener to include participants that met the criteria.

Out of the participants who completed the survey, 443 respondents provided complete responses and consented to participate. We excluded participants who declined to consent or provided incomplete responses. Each qualified respondent received a $3 compensation for their participation and responses. Table 2.1 presents a breakdown of the respondents' demographics, including age, gender, race, ethnicity, education, income, marital status, census region, frequency of driving, driving experience, personality, and expectations about AVs.

3.4.2 Study Design

The study utilized a within-subjects design with four conditions, as shown in Table 3.1. Two independent variables were manipulated, which were related to risk factors influencing AVs, namely external factor (i.e., weather) and internal factor (i.e., AV driving behavior). As part of the study, four videos were created to depict the different experimental conditions.

Table 3.1: Experimental design.

		AV Driving Behavior	
		Normal(N)	Aggressive(A)
Weather	Sunny(N)	NN	NA
	Snowy(W)	WN	WA

Screenshots from each of these videos, illustrating the conditions, can be seen in Figures 3.1.

(a) Sunny weather & Normal driving behavior

(b) Sunny weather & Aggressive driving behavior

(c) Snowy weather & Normal driving behavior

(d) Snowy weather & Aggressive driving behavior

Figure 3.1: Video screenshots of four conditions

3.4.2.1 Independent Variables

Weather Conditions: In this study, we manipulated weather conditions as an external factor by using two levels: sunny (N) and snowy (W). We altered the environment in which the AVs operated to manipulate the weather conditions, adjusting road visibility and video brightness accordingly. Specifically, videos depicting sunny weather conditions had higher road visibility and video brightness compared to those showing snowy weather conditions.

AV Driving Behaviors: We identified two types of AV driving behaviors: normal (N) and aggressive (A). We operationalized these behaviors by adjusting the speedometer reading and the frequency of car shakes in the video clips. Video clips presenting aggressive driving behavior had a higher number of car shakes and a higher average speed, whereas those depicting normal driving behavior had fewer car shakes and a lower average speed.

AV Expectation: To measure participants' expectations of AVs before experiencing them in the study, we administered a questionnaire with a 7-point Likert rating scale ranging from 1 (low) to 7 (high) [185]. The questionnaire consisted of three items designed to assess participants' overall expectations, effectiveness, and safety regarding self-driving cars. The items were as follows: (1) How would you rate your overall expectations regarding the driving of a self-driving car?; (2) How would you rate your expectations regarding the effectiveness of a self-driving car?; and (3) How would you rate your expectations regarding the safety of a self-driving car? Our goal in using this questionnaire was to measure participants' pre-existing beliefs and attitudes towards AVs, which may have influenced their subsequent perceptions and behaviors in the study.

Perceived Performance: We administered a questionnaire to assess participants' realistic perceptions of the capabilities and performance of AVs. The questionnaire consisted of three items, rated on a seven-point Likert scale (1 = low, 7 = high), which aimed to evaluate participants' perceptions of the AVs they had just experienced. The items were as follows: (1) How would you rate the driving of the self-driving car you just experienced?; (2) How would you rate the effectiveness of the self-driving car you just experienced?; and (3) How would you rate the safety of the self-driving car you just experienced? The questionnaire was designed to assess participants' subjective evaluations of the AVs they had just encountered and to compare these evaluations with their pre-existing expectations, as assessed in an earlier questionnaire. This comparison aimed to provide insight into the potential impact of experience on participants' perceptions of AVs.

3.4.2.2 Dependent Variable

The level of trust in AVs was measured using the Scale of Trust in Automated Systems, developed by Jian et al. (2000) [76]. This validated 12-item scale is commonly used to collect data on trust in both human and automated systems. We made appropriate revisions to the trust questionnaire to suit the context of AVs. To derive the composite subjective trust score, we averaged the responses to all 12 items in the questionnaire.

3.4.2.3 Manipulation Check of Risk Conditions

Participants were presented with four videos, each representing varied AV driving behaviors and weather conditions. Our aim was to determine whether participants could discern the differences between these scenarios in terms of their internal and external influences. Following the viewing of each video, participants were requested to assess their perceived risk associated with different weather conditions, utilizing a 7-point Likert scale. This measure was adapted from the questionnaire used in Hayes' 1998 study [66]. Subsequently, we evaluated whether participants could distinguish between two distinct AV driving behaviors, employing a perceived unsafety scale. This scale, consisting of eight items, was also adapted from Hayes' 1998 study [66], and utilized a 5-point Likert scale for responses.

3.4.3 Data Analysis Approach

This section covers the use of polynomial regression, response surface methodology, and the analytical model applied in this study.

3.4.3.1 Polynomial Regression

Previous research has used difference scores as predictors of certain outcomes [44, 45]. Difference scores can be algebraic, absolute, or squared differences between two components. Polynomial regression is an empirically-driven statistical technique that employs equations to test the relationship between difference scores and outcomes [45]. For instance, Equation 3.1 can be used to assess how the change in outcome (Z) is related to the squared difference score between two measures (X and Y) as these measures increase or decrease:

$$Z = b_0 + b_1(X - Y)^2 + e \tag{3.1}$$

e represents a random distribution term, and the positive sign on b_1 indicates that Z increases as the difference between X and Y changes in either direction [45, 44]. The equation can be expanded as follows:

$$Z = b_0 + b_1X^2 - 2b_1XY + b_1Y^2 + e \tag{3.2}$$

Equation 3.2 can be further relaxed and added with the low-order terms, which generate Equation 3.3:

$$Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + e \tag{3.3}$$

When combining Equations 3.2 and 3.3, four constraints are established: (1) the coefficient on X equals zero, (2) the coefficient on Y equals zero, (3) the coefficient on X^2 and Y^2 is identical, and (4) the sum of coefficients on X^2, Y^2, and XY equals zero. In prior research using polynomial regression equations, these constraints were usually rejected. Therefore, response surface methodology is necessary to correctly interpret the coefficients of unconstrained quadratic equations in such cases.

3.4.3.2 Response Surface Methodology

Response surface methodology is a useful tool for testing and describing the key surface features that correspond to quadratic equations. It resolves the challenge of interpreting coefficients in unconstrained quadratic equations [45, 44, 89, 13]. The response surface is interpreted by three main features: (1) Stationary points, which correspond to those at which the slope of the surface is zero (minimal, maximal, or saddle point of the surface); (2) Principal axes, which are the lines that run perpendicular to each other and intersect at the stationary point. The slopes and intercepts of both principal axes can be computed using the coefficients in Equation 3.3; (3) Slopes along lines of interest, such as the principal axis and confirmation/disconfirmation axis, that are relevant to testing and interpreting the hypothesis [45, 18].

3.4.4 Proposed Study Model

The study model is based on the expectation disconfirmation theory. Hypothesis 1 proposes that people will trust AVs more when their performance exceeds expectations, while trust will be negatively impacted when performance falls short of expectations. This hypothesis can be tested and explained using the following equation:

$$Z = b_0 + b_1 E + b_2 P + b_3 E^2 + b_4 EP + b_5 P^2 + e \tag{3.4}$$

Z represents AV trust, E represents expectations towards AVs, and P represents perceived performance.

To test Hypothesis 2 (H2), we propose that two moderators, weather conditions and AV driving behavior, play a significant role in the relationship between AV trust and expectation-disconfirmation. The equation representing this hypothesis is as follows:

$$Z = b_0 + b_1 E + b_2 P + b_3 E^2 + b_4 EP + b_5 P^2 + b_6 V + b_7 EV + b_8 PV + b_9 E^2 V + b_{10} EPV + b_{11} P^2 V + e \tag{3.5}$$

Z represents trust, E represents AV expectations, P represents perceived performance, and V represents the moderator (i.e., weather conditions or driving behavior of the AV). The terms EV, PV, E^2V, EPV, and P^2V are included to test the moderation effect. The increment in R^2 generated by these terms will be used to test the moderation effect of the two variables.

3.5 Result

3.5.1 Construct Reliability and Validity

To reduce multicollinearity and improve the interpretation of the polynomial equation coefficients, the expectation and perceived performance scales were centered by subtracting the scale midpoints from the actual score [46]. Construct validity and reliability were evaluated using a sequential approach suggested by Wille (1996) [194]. Firstly, Cronbach's alpha was computed, and the results showed that the alpha values for all constructs were 0.80 or higher, indicating acceptable levels of construct reliability. Secondly, the validity of the measurement constructs was assessed using principal component analysis with varimax rotation, resulting in a five-factor solution. As displayed in Table 3.2, all items had factor loadings of 0.7 or higher, indicating acceptable construct validity. An overview of the means, standard deviations, and correlations is provided in Table 3.3.

3.5.2 Hypotheses Testing

Our study involved an exploratory analysis that utilized polynomial regression and response surface analysis. Through our analysis, we discovered that the higher-order equation explained a more substantial amount of variance compared to the first-order equation, as evidenced by the results in Table 3.4. Consequently, we rejected the linear model and chose the quadratic model instead. Furthermore, the F-test results demonstrated that the R^2 value of the quadratic equation was significantly higher than that of the linear equation when predicting trust in AVs based on expectations of AVs and perceived performance.

3.5.2.1 Manipulation Check

The results from this portion of our study illustrated that participants were significantly capable of differentiating various weather conditions. Snowy weather was associated with a higher perceived risk (mean = 4.722) compared to sunny weather (mean = 4.116, $p < 0.001$). In terms of AV driving behaviors, aggressive driving behavior led to a heightened

Table 3.2: Factor loadings and items.

Variable	Items	Cronbach's Alpha	Component 1	Component 2	Component 3
Trust	The just shown self-driving car was deceptive.	0.93	0.81		
	The just shown self-driving car behaved in an underhanded manner.		0.87		
	I was suspicious of the just shown self-driving car's intent, action, or output.		0.92		
	I was wary of the just shown self-driving car.		0.89		
	I worried that the just shown self-driving car action will have a harmful or injurious outcome.		0.84		
	I was confident in the just shown self-driving car.		0.92		
	The just shown self-driving car provided security.		0.93		
	The just shown self-driving car had integrity.		0.91		
	The just shown self-driving car was dependable.		0.93		
	The just shown self-driving car was reliable.		0.93		
	I can trust the just shown self-driving car		0.94		
	I am familiar with the just shown self-driving car.		0.76		
Expectation	How would you rate your overall expectations regarding the driving of a self-driving car?	0.96		0.96	
	How would you rate your expectations regarding the effectiveness of a self-driving car?			0.96	
	How would you rate your expectations regarding the safety of a self-driving car?			0.95	
Perceived Performance	How would you rate the driving of the just shown self-driving car?	0.95			0.84
	How would you rate the effectiveness of the just shown self-driving car?				0.84
	How would you rate the safety of the just shown self-driving car?				0.83

Table 3.3: Descriptive statistics and correlation matrix.

Variable	Means	Std.Dev	1	2	3
Expectation	-0.25	2.10	1		
Perceived performance	-0.29	2.05	0.57**	1	
Trust	3.93	1.78	0.22**	0.47**	1
*Note: **p < 0.001*					

Table 3.4: Predicting trust using expectation and perceived performance.

Dependent Variable	**Independent Variables**	**First-Order Linear Equation**		**Second-Order Quadratic Equation**	
		R^2		R^2	
Trust	Expectation	0.48**	-0.057*	0.49*	-0.059*
	Perceived Performance		0.442**		0.447**
	$Expectation^2$				-0.031*
	Expectation x Perceived Performance				0.025*
	$Perceived\ Performance^2$				0.018
*Note: *p < 0.05, **p < 0.001*					

perception of unsafety (mean = 3.797) relative to normal driving behavior (mean = 3.289, $p < 0.001$). These findings underscore the sensitivity of participants to both environmental and behavioral variations in AV operation.

3.5.2.2 Expectation-Disconfirmation Theory and Polynomial Regression Analysis

Based on the EDT, we propose that as disconfirmation becomes more negative, trust in AVs will decrease (H1A), while as disconfirmation becomes more positive, the intention to trust will increase (H1B). To support these hypotheses, two conditions on the response surface need to be met: the slope of the disconfirmation axis should be negative, while the confirmation axis should be positive. Additionally, the slope of the disconfirmation axis should be higher than the slope of the confirmation axis, since positive disconfirmation is linked with a higher level of trust outcome [18]. Table 3.6 shows that the slope of the disconfirmation axis is -0.506, while the confirmation axis slope is 0.388, meeting both conditions. Therefore, these results provide empirical evidence that the discrepancy between experiences and expectations influences trust in AVs. Trust increases when expectations are met or exceeded and decreases when expectations fall short, supporting H1A and H1B.

Table 3.5: Stationary points and principal axes.

Dependent Variable	**Stationary Point**		**First Principal Axis**		**Second Principal Axis**	
	X_0	Y_0	p_{10}	p_{11}	p_{20}	p_{21}
Trust	-4.655	-9.184	10.182	4.160	-10.303	-0.240

To gain a better understanding of the relationship between disconfirmation and trust in AVs, we conducted a detailed response surface analysis. The key features and corresponding results of this analysis are presented in Tables 3.6 and 3.5. Our approach followed the

Table 3.6: Slopes along lines of interest.

Dependent Variable	**Y = X**		**Y = -X**		**Surface Along First Principal Axis**		**Surface Along Second Principal Axis**	
	a_x	a_x^2	a_x	a_x^2	a_x	a_x^2	a_x	a_x^2
Trust	0.388**	0.012	-0.506**	-0.038	3.580	0.385	-0.335	-0.036
*Note: **p < 0.001*								

steps outlined by Edwards and Harrison (1993) [46]. Specifically, we began by examining the location of the stationary point to determine whether the surface was centered at the origin of the XY plane. Our results showed that the surfaces for expectation and perceived performance predicting trust in AVs were saddle-shaped (see Figure 3.2). Furthermore, we found that the stationary point was located beyond the front edge of the XY plane, specifically at X_0 = -4.655 and Y_0 = -9.184. These findings provide valuable insights into the complex interplay between disconfirmation and trust in AVs.

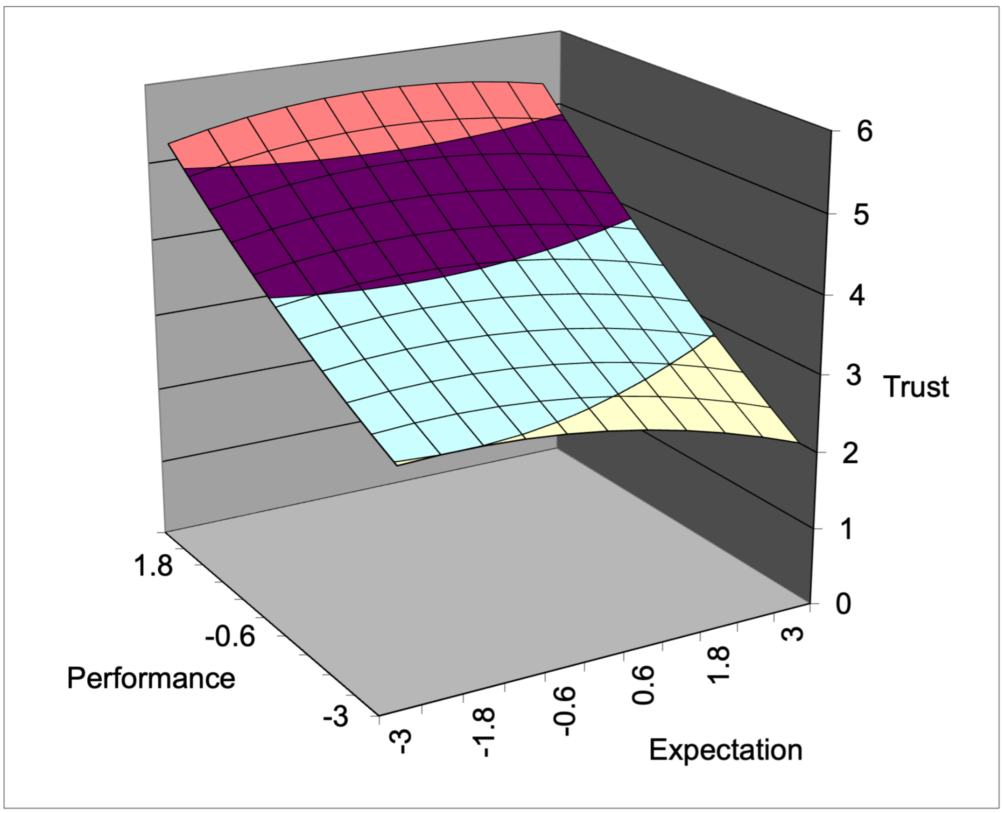

Figure 3.2: Response surface analysis for disconfirmation predicting trust in AVs

In the second step of our analysis, we examined the intercepts and slopes of the principal axes of the surface, which describe its orientation in the XY plane. The first principal axis, Y = $p_{10} + p_{11}$X, had an intercept and slope of 10.182 and 4.160, respectively, but neither was statistically significant. This indicates that the axis is not significantly different from the X = Y line. We found that the surface was flat along the first principal axis (a_x^2 = 0.385, not statistically significant; a_x = 3.580, not statistically significant). The second principal axis had an intercept and slope of -10.303 and -0.240, respectively, but neither was statistically significant, indicating that the axis is not significantly rotated or shifted from the X = -Y line. Along this axis, the surface had a flat contour and a slightly negative slope where it crossed the y-axis (a_x^2 = -0.036, not statistically significant; a_x = -0.335, not statistically significant). These findings provide further insights into the shape and orientation of the surface and how it relates to the complex interplay between disconfirmation and trust in AVs.

As a final step, we examined the surface along the Y = X and Y = -X lines. The Y = X axis surface manifested no notable curvature (a_x^2 = 0.012, not statistically significant) and exhibited positivity at the origin (a_x = 0.388, $p < 0.001$). In contrast, the surface along the Y = -X axis displayed a negative slope, without any observable curvature (a_x^2 = -0.038, not statistically significant; a_x = -0.506, $p < 0.001$).

Taken together, the response surface analysis revealed three key findings. First, the surface showed an upward surface along the Y = X line, indicating that drivers tend to have higher trust in AVs when both expectation and perceived performance are high compared to when both are low. Second, the surface exhibited a downward tendency along the Y = -X line, indicating that trust in AVs increases at a constant rate (0.506) as perceived performance increases toward expectation and continues to increase when expectation and performance are equal (X = 0, Y = 0). This means that when the AV's performance falls short of expectations, the greater the deviation from those expectations, the lower the level of trust in AVs. Third, the surface along the Y = -X line also revealed that trust in AVs increases when perceived performance exceeds expectations. Specifically, the greater the AV's performance exceeds expectations, the higher the level of trust in AVs. This comprehensive analysis sheds light on the intricate relationship among expectation, disconfirmation, and trust in AVs and provides support for both hypotheses H1A and H1B.

3.5.2.3 Moderated Polynomial Regression Analysis

To assess whether external factors such as weather and internal risk factors like AV driving behavior influence the relationship between disconfirmation and trust in AVs, we used moderated polynomial regression techniques [46, 44, 42]. We incorporated the variable V as a moderator into the quadratic regression equation for X, resulting in equation Y. We then measured the moderation effect by including the terms V, XV, YV, XYV, X^2V, and

Table 3.7: Results of moderated polynomial regression analysis for weather conditions.

Dependent Variable	Independent Variables	Coefficients	Model 1	Model 2
Trust	Expectation	b_1	-0.059*	-0.235**
	Perceived Performance	b_2	0.447**	0.673**
	$Expectation^2$	b_3	-0.031*	-0.072
	Expectation x Perceived Performance	b_4	0.025*	0.021
	$PerceivedPerformance^2$	b_5	0.018	0.068
	Weather	b_6		-0.325*
	Expectation x Weather	b_7		0.119*
	Perceived Performance x Weather	b_8		-0.162**
	$Expectation^2\ xWeather$	b_9		0.033
	Expectation x Perceived Performance x Weather	b_{10}		0.003
	$PerceivedPerformance^2 xWeather$	b_{11}		-0.037
R^2			0.478**	0.492**
$R^2 Change$				0.014**
*Note: *p < 0.05, **p < 0.001*				

Y^2V and examining the increase in R^2. This approach helps to determine whether the effects of disconfirmation on trust in AVs are contingent upon specific external or internal factors, providing a more nuanced understanding of the complex relationship between these variables.

The results presented in Table 3.7 show that including weather as a moderator in the model led to a statistically significant increase in R^2, with an F change value of 5.236 ($p < 0.001$). This finding suggests that weather, as an external risk factor, moderates the relationship between disconfirmation and trust in AVs, providing support for hypothesis H2A.

Our analysis also supports hypothesis H2B, which suggests that AV driving behavior acts as an internal risk factor that moderates the relationship between disconfirmation and trust in AVs. Including AV driving behavior as a moderator in the model leads to a significant increase in R^2, with an F change value of 5.461 and a p-value of less than 0.001, as shown in Table 3.8. These findings confirm that AV driving behavior influences the relationship between disconfirmation and trust in AVs, providing support for hypothesis H2B. The results suggest that both external factors, such as weather, and internal factors, such as AV driving behavior, play a critical role in shaping the complex relationship between disconfirmation and trust in AVs. To gain a clearer understanding of how trust varies with disconfirmation and risk perceptions, we performed a more detailed response surface analysis for each of the experimental conditions.

Moderated Polynomial Regression Analysis for Weather

Table 3.8: Results of moderated polynomial regression analysis for AV driving behavior.

Dependent Variable	Independent Variables	Coefficients	Model	
			1	2
Trust	Expectation	b_1	-0.059*	-0.177*
	Perceived Performance	b_2	0.447**	0.632*
	$Expectation^2$	b_3	-0.031*	-0.072
	Expectation x Perceived Performance	b_4	0.025*	0.050
	$PerceivedPerformance^2$	b_5	0.018	0.028
	AV Driving Behavior	b_6		-0.432*
	Expectation x AV Driving Behavior	b_7		0.091
	Perceived Performance x AV Driving Behavior	b_8		-0.144*
	$Expectation^2\ xAVDrivingBehavior$	b_9		0.027
	Expectation x Perceived Performance x AV Driving Behavior	b_{10}		0.010
	$PerceivedPerformance^2 xAVDrivingBehavior$	b_{11}		-0.012
R^2			0.478**	0.492**
$R^2 Change$				0.014**
*Note: *p < 0.05, **p < 0.001*				

Table 3.9: Coefficients for response surface analysis predicting trust in AVs at sunny versus snowy weather.

Dependent Variable	Independent Variables	Coefficients	Weather	
			Sunny	Snowy
Trust	Intercept	b_0	4.178	3.853
	Expectation	b_1	-0.116**	0.004
	Perceived Performance	b_2	0.511**	0.348**
	$Expectation^2$	b_3	-0.040*	-0.007
	Expectation x Perceived Performance	b_4	0.024	0.027
	$PerceivedPerformance^2$	b_5	0.031	-0.006
*Note: *p < 0.05, **p < 0.001*				

Sunny Weather: The association between expectation, perceived performance, and trust in Autonomous Vehicles (AVs) under sunny weather conditions is depicted in Figure 3.3. The coefficients used in the response surface analysis, which predict trust in AVs in sunny weather, are documented in Table 3.9. As found in the comprehensive data analysis, the surface portrayed a saddle shape with a stationary point residing at X_0 = 3.514, Y_0 = -6.88, as specified in Table 3.10. Table 3.11 showed that the slope of the first principal axis was 6.081 (not statistically significant), and the intercept of the axis was 14.490 (not statistically significant). The analysis of the Y = X line revealed a surface without curvature and a positive slope at the origin (a_x^2 = 0.015, not statistically significant; a_x = 0.395, p < 0.001). In contrast, the Y = -X line demonstrated a surface void of curvature and exhibited a negative slope (a_x^2 = -0.033, not statistically significant; a_x = -0.627, p < 0.001). Overall, these findings were consistent with those obtained from the analysis of all data surfaces.

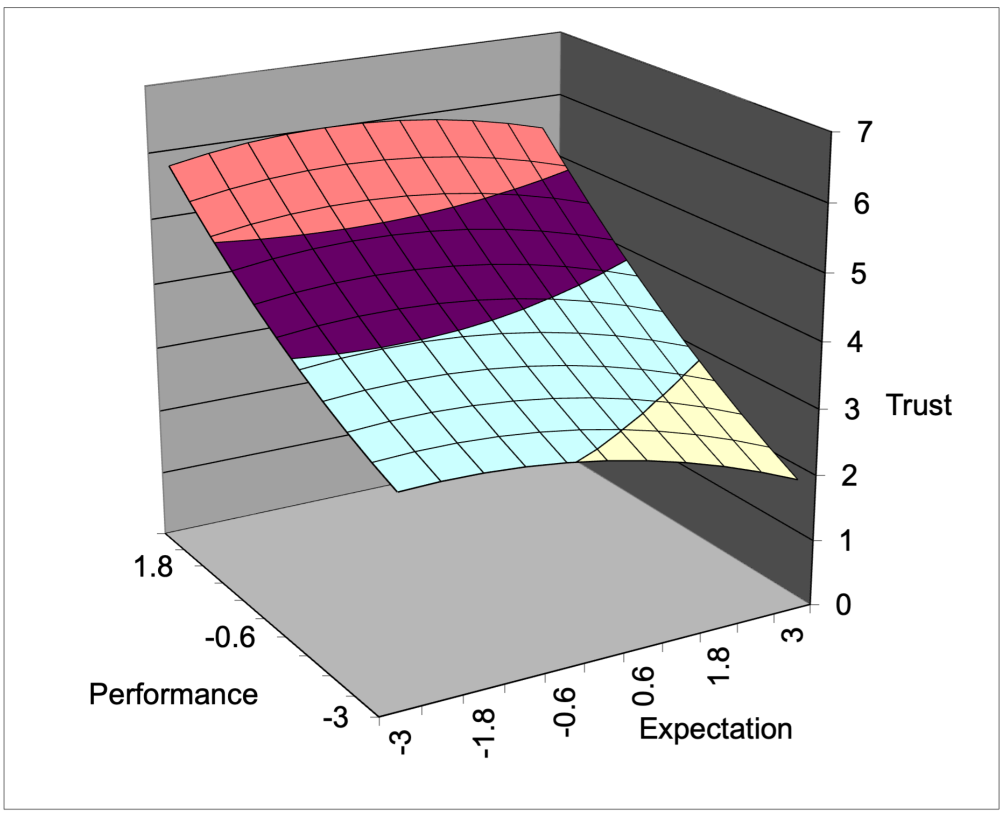

Figure 3.3: Response surface for sunny weather condition

Snowy Weather: In snowy weather conditions (Figure 3.4), we observed a saddle-shaped surface for expectation and perceived performance that predicted trust in AVs. The stationary point was located beyond the XY plane at X_0 = -16.834 and Y_0 = -8.87 (Table 3.10). The slope of the first principal axis was slightly greater than 1 (p_{10} = 1.038, p¡0.05), indicating a

Table 3.10: Stationary points and principal axes for sunny and snowy weather condition.

Weather	Stationary Point		First Principal Axis		Second Principal Axis	
	X_0	Y_0	p_{10}	p_{11}	p_{20}	p_{21}
Sunny	-3.514	-6.882	14.490	6.081	-7.459	-0.164
Snowy	-16.834	-8.877	8.592	1.038*	-25.099	-0.964*
*Note: *p < 0.05*						

Table 3.11: Slopes along lines of interest for sunny and snowy weather condition.

Weather	Y = X		Y = -X		Surface Along First Principal Axis		Surface Along Second Principal Axis	
	a_x	a_x^2	a_x	a_x^2	a_x	a_x^2	a_x	a_x^2
Sunny	0.395**	0.015	-0.627**	-0.033	8.802	1.252	-0.303	-0.043
Snowy	0.352**	0.014	-0.344**	-0.040	0.490*	0.015*	-1.299	-0.039
*Note: *p < 0.05, **p < 0.001*								

slight counterclockwise rotation of the surface off the Y = X line (Table 3.11). The surface was convex along the first principal axis (a_x^2 = 0.015, p < 0.05) and had a significant positive slope where it crossed the y-axis (a_x = 0.490, p < 0.05). The slope of the second principal axis was slightly lower than -1 (p_{20} = -0.964, p < 0.05), indicating a slight counterclockwise rotation of the surface off the Y = -X line. The surface along the Y = X line displayed a non-curved profile with a positive slope at its origin (a_x^2 = 0.014, not statistically significant; a_x = 0.352, p < 0.001), and the surface along the Y = -X line showed a lack of curvature and had a negative slope at its origin (a_x^2 = -0.037, not statistically significant; a_x = -0.035, p < 0.001).

Overall, the results from the analysis of the surface for snowy conditions were generally consistent with those for sunny conditions, with two exceptions. First, although trust increased as perceived performance increased toward and exceeded the initial expectation for both conditions, the rate of increase in trust was slower in snowy conditions than in sunny conditions (0.344 vs. 0.627). This indicates that drivers are more cautious in forming trust when facing high external risk conditions, such as snowy weather. Second, based on the curvature along the first principal axis and tendency along the Y = X surface, we can conclude that trust increases at a faster rate when both expectation and performance are high than when both are low.

Moderated Polynomial Regression Analysis for AV Driving Behavior

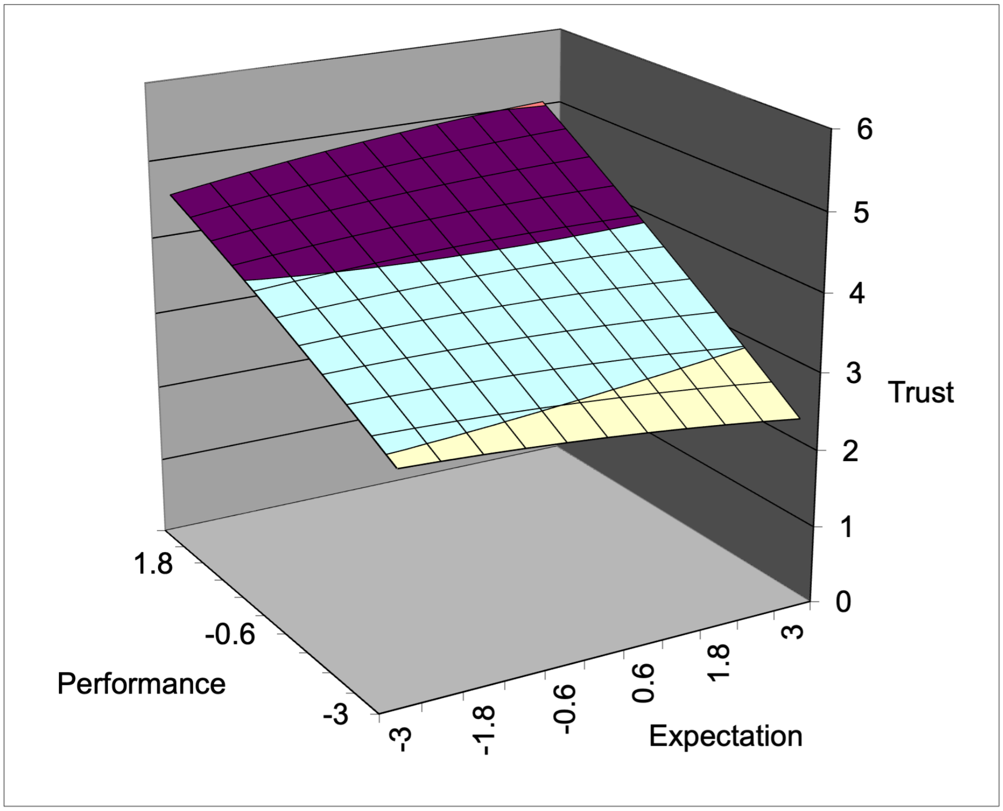

Figure 3.4: Response surface for snowy weather condition

Normal AV Driving Behavior: Figure 3.5 displays the corresponding surface for normal AV driving behavior conditions. The stationary point was located at X = -4.971, Y = -9.036 (Table 3.13). The first principal axis had an intercept and slope of 7.612 and 3.349, respectively, indicating that the axis is not significantly rotated or shifted from the X = Y line (Table 3.14). The second principal axis had an intercept and slope of -10.520 and -0.299 and is not significantly different from the X = -Y line. Along the Y = X line, the surface had an insignificant curvature of 0.011 and a slope of 0.402 at the points X = 0, Y = 0. Along the Y = –X line, the surface had a non-significant curvature of –0.069 and a slope of 0.574 at the point X = 0, Y = 0. These results provide further understanding of the intricate relationship between disconfirmation, risk perception, and trust in AVs in the context of normal AV driving behavior. The negligible curvature and positive slope along the Y = X line suggest that drivers have higher trust in AVs when their expectations and perceived performance are high. Conversely, the insignificant curvature and negative slope along the Y = –X line indicate that trust increases as the AV's performance exceeds expectations.

Aggressive AV Driving Behavior: Figure 3.6 illustrates the response surface analysis for normal AV driving behavior conditions. The stationary point was located at X = -8.475

Table 3.12: Coefficients for response surface analysis predicting trust in AVs at normal versus aggressive AV driving behavior.

Dependent Variable	**Independent Variables**	**Coefficients**	**AV Driving Behavior**	
			Normal	**Aggressive**
Trust	Intercept	b_0	4.227**	3.795**
	Expectation	b_1	-0.086*	0.006
	Perceived Performance	b_2	0.488**	0.343**
	$Expectation^2$	b_3	-0.045*	-0.018
	Expectation x Perceived Performance	b_4	0.040*	0.031
	$PerceivedPerformance^2$	b_5	0.016	0.004
*Note: *p < 0.05, **p < 0.001*				

Table 3.13: Stationary points and principal axes for normal AV driving behavior.

AV Driving Behavior	**Stationary Point**		**First Principal Axis**		**Second Principal Axis**	
	X_0	Y_0	p_{10}	p_{11}	p_{20}	p_{21}
Normal	-4.971	-9.036	7.612	3.349	-10.520	-0.299
Aggressive	-8.475	-10.035	6.371	1.936	-14.413	-0.517

Table 3.14: Slopes along lines of interest for normal AV driving behavior.

AV Driving Behavior	**Y = X**		**Y = -X**		**Surface Along First Principal Axis**		**Surface Along Second Principal Axis**	
	a_x	a_x^2	a_x	a_x^2	a_x	a_x^2	a_x	a_x^2
Normal	0.402**	0.011	-0.574**	-0.069	2.668	0.268	-0.552	-0.056
Aggressive	0.349**	0.017	-0.337**	-0.045	0.966	0.057	-0.558	-0.033
*Note: **p < 0.001*								

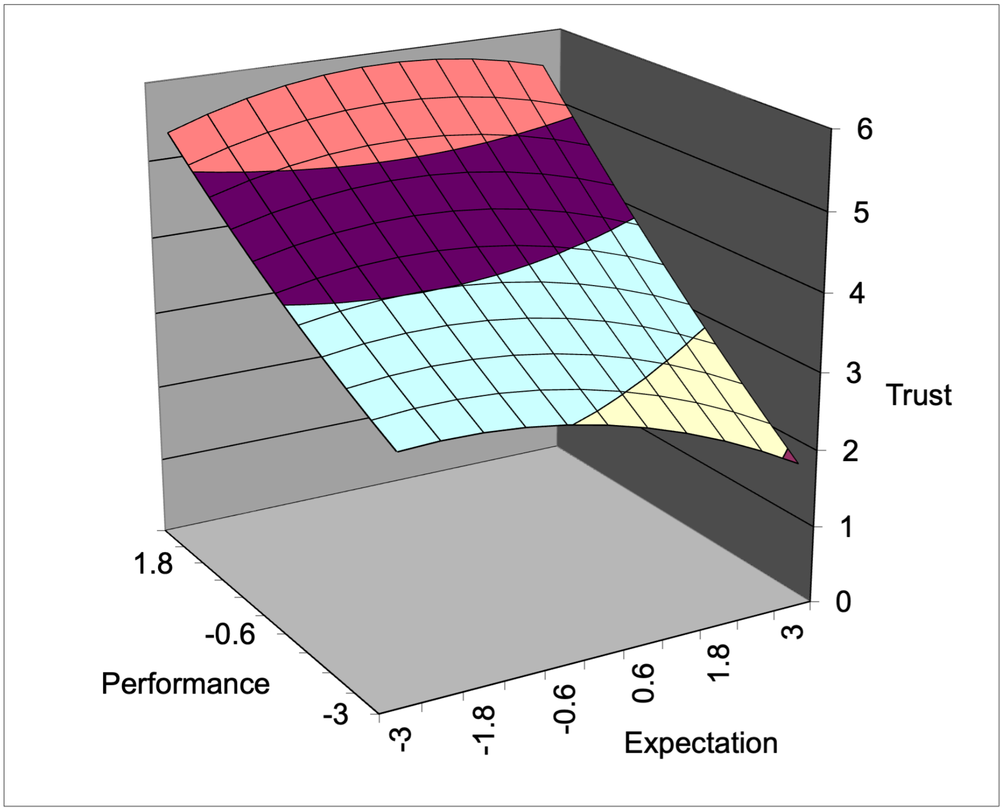

Figure 3.5: Response surface for normal AV driving behavior

and Y = -10.035 (Table 3.13). The first principal axis had an intercept of 6.371 and a slope of 1.936 (Table 3.14). The second principal axis had an intercept of -14.413 and a slope of -0.517, which indicates that it is not significantly different from the X=-Y line. Along the Y = X line, the surface demonstrated no curvature, with a curvature value of 0.017, and it presented a slope of 0.349 at the origin (X = 0, Y = 0). On the other hand, along the Y = -X line, the surface had a non-significant curvature of -0.045 and a negative slope of -0.337 at the point X = 0, Y = 0. Compared to AVs with normal driving behavior, those with aggressive behavior tend to elicit a slower rate of increase in trust as perceived performance moves towards, meets, and exceeds expectations.

3.5.3 Summary of the Results

This paper presents three overarching conclusions. First, the findings indicate that as disconfirmation becomes more negative, trust in AVs decreases. This is evidenced by the negative slope along the Y=-X surface, supporting H1A. Second, positive disconfirmation has a positive impact on trust in AVs. Trust continuously increases when perceived performance exceeds expectations, thus supporting H1B. Furthermore, moderation effect tests

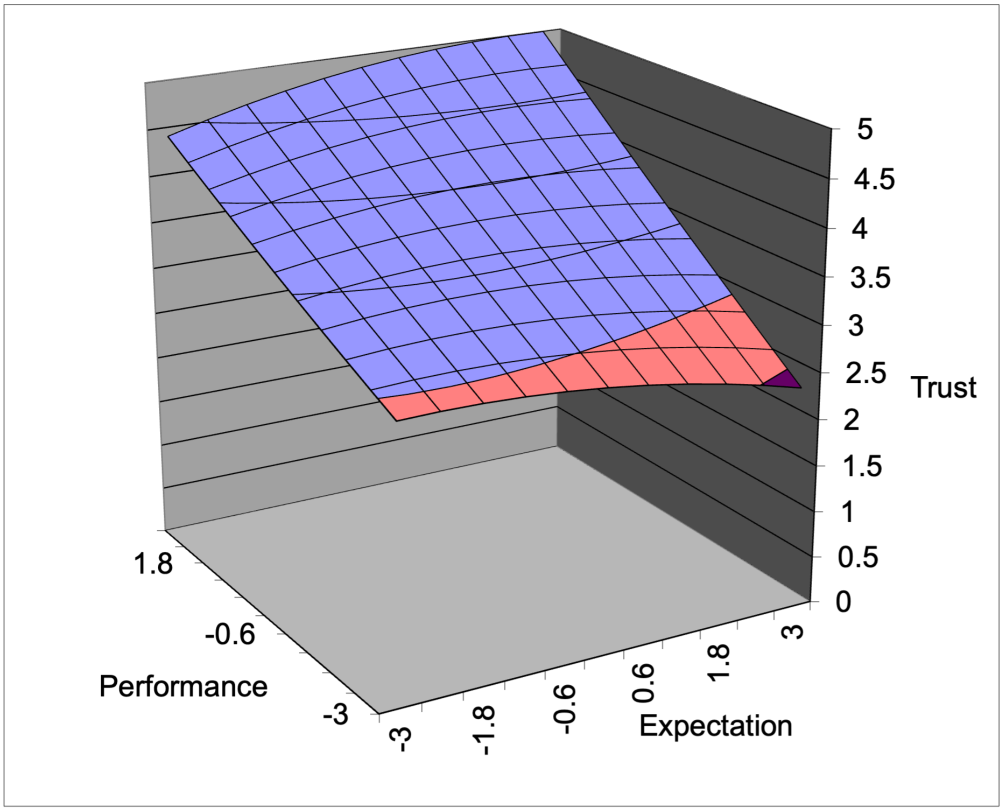

Figure 3.6: Response surface for aggressive AV driving behavior

for weather and driving behaviors reveal that these factors significantly influence the relationship between expectation disconfirmation and trust in AVs, supporting H2A and H2B. In addition, response surface methodology tests demonstrate that the rate at which trust increases varies depending on the risk conditions. For instance, in risky situations such as snowy weather and aggressive driving behavior, trust increases at a less steep rate than in sunny and normal AV driving behavior conditions, respectively. The following section describes the findings and their contributions to the literature.

3.6 Discussion

3.6.1 Expectation, Disconfirmation, and Trust in AVs

This study offers a significant contribution to the literature on trust in AVs by elucidating the role expectations play in shaping trust. While previous research has demonstrated that positive expectations towards automated driving systems increase individuals' subjective trust [92], this study delves deeper by revealing the mechanism through which expectations influence trust and emphasizing the crucial role of disconfirmation in determining trust in AVs.

The findings of the study show that AVs are perceived as more trustworthy when their performance exceeds expectations, leading to a positive disconfirmation effect that engenders a favorable response. Conversely, negative disconfirmation has a detrimental effect on trust, resulting in disappointment and distrust. When AVs do not perform as well as expected, it leads to a negative disconfirmation effect, which can undermine people's trust in them. Moreover, the study uncovered significant effects of varied levels of confirmation on trust. The results suggest that when the perceived performance of AVs aligns with high expectations, trust in AVs is greater than in cases where low expectations are met by low-performing AVs.

In summary, the study's findings support and explain previous research on the role of expectations in fostering trust in AVs. High-performing AVs generate more trust than lower-performing AVs when meeting or exceeding people's high expectations. However, when performance falls short of expectations, trust is negatively affected.

3.6.2 Perceived Risk, Disconfirmation, and Trust in AVs

This study underscores the crucial role of perceived risks, both internal and external, in influencing the impact of disconfirmation on trust in AVs. The findings suggest that perceived risk significantly affects the rate at which trust fluctuates, whether it decreases or increases. Notably, when AVs fail to meet expectations, trust declines more rapidly in low-risk conditions (e.g., sunny weather and normal driving behavior) compared to high-risk conditions (e.g., snowy weather and aggressive driving behavior). A possible explanation for this phenomenon is that people may recognize the challenging circumstances in high-risk situations and take into account the fact that AVs operate under more demanding and risky conditions. This understanding can enhance their perception of the technology's competence and reliability, reducing the negative impact of performance failures on trust. In contrast, in low-risk conditions, people may anticipate flawless AV performance, leading to a greater loss of trust when their expectations are not met. Moreover, people may have a higher baseline trust in AVs in low-risk conditions due to the perceived safety and predictability of the driving situation. As a result, when AVs fail to meet their expectations, individuals might experience a more profound sense of disappointment and betrayal, causing a more rapid decline in trust. In comparison, people may have a lower baseline trust in AVs in high-risk conditions because of the perceived complexity and uncertainty. Consequently, when AVs perform poorly in these situations, people may be less surprised and disappointed, resulting in a slower decline in trust.

Furthermore, the study reveals that trust in AVs increases more rapidly in low-risk situations than in high-risk ones when the AVs perform better than expected. This finding is counterintuitive, as we might anticipate trust to rise more quickly in high-risk situations when performance exceeds expectations, considering that people may be surprised and impressed by the AV's ability to successfully complete tasks under challenging and unfavorable conditions. A potential explanation for this counterintuitive phenomenon is that individuals may be emotionally averse to high-risk situations. When driving conditions are unfavorable, people might be more reluctant to drive and adopt a cautious approach in high-risk situations to protect themselves from potential harm. This natural inclination to avoid high-risk situations could result in a slower increase in trust toward AVs, even when they perform better than expected. Conversely, in low-risk conditions, people may be more open to trusting AVs, as they perceive a lower threat in the driving environment. As a result, their intention to trust AVs may increase more rapidly than in high-risk conditions. These findings align with previous research demonstrating that risk perception influences individuals' evaluations and trust in AVs. In conclusion, these findings emphasize the importance of considering both internal and external risk factors when cultivating trust in AVs. By addressing these factors, we can better understand the dynamics of trust development and work towards creating an environment where users can confidently rely on AVs in various driving situations.

3.6.3 Design Implications

The findings of this study provide valuable insights into designing AVs to enhance user trust and promote widespread adoption. It is crucial for AV manufacturers and designers to concentrate on meeting or exceeding users' expectations in order to instill confidence in the technology. To engender trust, AVs should operate efficiently and effectively across a wide range of driving conditions, including high-risk situations.

Effectively managing user expectations is essential for building trust in AVs. AV designers must transparently communicate the capabilities and limitations of their systems to users. By offering clear and concise information about the potential risks and benefits of using AVs, designers can manage user expectations and foster trust in the technology. Users should also be informed of situations in which AVs may not perform optimally, such as during high-risk conditions. Furthermore, AV designers should prioritize creating adaptable vehicles that can handle various driving conditions and cater to user preferences. Equipping AVs with features that provide personalized and comfortable driving experiences will help meet the unique needs and expectations of individual users.

In conclusion, by focusing on meeting and surpassing user expectations, managing those expectations, and designing adaptable AVs for different driving conditions and user preferences, designers can build trust among users and encourage the widespread adoption of this transformative technology.

3.6.4 Limitation and Future Research

While this study sheds light on the relationship between expectations, perceived performance, and trust in AVs, it is vital to acknowledge certain limitations that must temper our interpretation of the findings. To begin with, the study relied heavily on survey methods to ascertain variables, raising potential issues about method bias and external validity. For instance, the real-world conditions may not be accurately represented by the videos used in this study, possibly undermining the generalizability of the findings. To bolster the external validity and reduce these biases, it is advisable for future research to adopt field settings, offering a more authentic context for study.

Moreover, although the study collected demographic and driving-related information from participants, it left unexplored numerous influential factors, such as personality traits, the tendency to be an early adopter, and the influences of surrounding social systems. These elements can significantly shape an individual's perception and trust in AVs. For instance, personality traits can exert a significant impact on expectations, performance perception, and trust, making it important for future research to probe these interrelationships for a more layered understanding of trust dynamics in AVs.

Furthermore, an individual's inclination to be an early adopter of novel technologies can sway their acceptance and adoption of AVs. As suggested by the Diffusion of Innovations theory, potential adopters with an affinity for innovations are more likely to accommodate the necessary changes for the adoption of a new technology [163, 51]. Therefore, future research should take into account the early adopter tendencies of individuals when examining the acceptance and adoption of AV technology. The social system within which an individual operates can also exert substantial influence on their technology-related decisions and perceptions. A crucial factor in diffusion of innovations research, a social system includes external pressures (such as media influence, mandates from organizations or government) and internal factors (like social relationships, proximity to influential figures) [51]. A deeper consideration of these multifaceted roles within a social system can offer a more comprehensive view of the factors impacting an individual's decision to adopt AVs.

Finally, it's important to note that the results of this study are closely tied to the constructs measured, which constrains their generalizability. The investigation of other constructs beyond trust could yield diverse insights and contribute to a more complete understanding of the factors driving the adoption and usage of AVs.

In conclusion, while this study provides essential insights into the relationship between expectations, performance perception, and trust in AVs, these identified limitations call for further research. Future studies should strive to address these limitations to provide a more comprehensive and nuanced understanding of the factors influencing the adoption and acceptance of AVs.

3.7 Conclusion

This study investigates the challenges of incorporating AVs into transportation systems, specifically focusing on the prevalent lack of trust many people have towards them. Although trust in AVs has been extensively researched, the role of expectations in shaping trust remains underexplored. The Expectation Disconfirmation Theory (EDT) offers a valuable framework for comprehending how users compare their expectations to actual experiences, resulting in disconfirmation, which in turn affects their attitudes and adoption of technology. To better understand the interplay between expectations, disconfirmation, risk perception, and trust in AVs, an online survey involving 443 drivers in the United States was conducted. The survey's findings highlight the significant role of expectations in the cognitive appraisal process and their impact on building trust in AVs. This study emphasizes the necessity of ongoing research and development of AVs, as well as the implementation of effective strategies to enhance trust in this emerging technology. By gaining a deeper understanding of the factors influencing trust in AVs, we can work towards establishing a safer and more efficient transportation system that benefits everyone.

CHAPTER 4

Finding the Right Voice: Exploring the Impact of Gender Similarity Between Human and Automated Vehicle Voice and Gender-Role Congruity on the Efficacy of Automated Vehicle Explanations

4.1 Introduction

Existing research has primarily concentrated on the aspect of auditory explanations within the realm of human-AV interactions, with less emphasis on understanding how voice characteristics impact user trust in AVs. This paper endeavors to address this gap, delving into the influence of gender similarity between the human user and the AV's explanation voice, and the concept of gender-role congruity, on user preferences in relation to cognitive and affective trust towards AV explanation voices.

To advance our understanding in this domain, we conducted an online survey. The outcomes from this survey provide a nuanced comprehension of the demographic and perceptual elements at play. Our findings reveal that gender similarity between human users and the AV explanation voice significantly sways both cognitive and affective trust. When the gender of the human user and the AV voice align, trust in AVs is enhanced compared to groups where genders are dissimilar.

Additionally, the principle of gender-role congruity serves as a substantial moderating influence in the dynamic between gender similarity and affective trust. The profound impact of gender similarity is notably moderated when the gender of the AV's voice is incongruous with the societal role traditionally attributed to that gender. To put it more simply, if the AV voice matches the user's gender, but this gender doesn't align with what is traditionally expected for the role that the AV is performing (based on societal gender stereotypes), then the increase in affective trust is not as strong. On the other hand, when the gender of the AV voice is in line with both the user's gender and traditional gender-role expectations, the increase in trust is more noticeable. So, gender-role congruity — or how well the gender of the AV voice aligns with traditional gender-role expectations — can influence the level of trust users place in AVs.

Consequently, integrating these insights into future research and design practices can substantially enrich the advancement of AV technology. It can pave the way for fostering heightened levels of trust and confidence among users, which in turn can expedite the broader acceptance and adoption of AVs.

4.2 Background

4.2.1 Explanations and Automated Vehicles

According to the *Cambridge Advanced Learner's Dictionary*, explanations are given to make something clear or easy to understand. In the context of AVs, explanations refer to the rationale provided by the AV to make its actions comprehensible. Explanations are crucial for AV acceptance, as they supply users with essential information regarding automation decisions, leading to improved user-AV interactions [37]. AV explanations render the vehicle's actions predictable and intelligible, assisting drivers in forming accurate mental models [92]. These models serve as approximate representations of the AV's functions and capabilities, helping the driver understand appropriate actions in any given situation [53, 181]. Therefore, explanations are vital in ensuring users trust the AV's abilities and make informed decisions while operating it. Research on AV explanations has identified two key areas: explanation content and explanation timing [37, 61, 92, 53, 85, 86, 64, 193, 143].

Explanation content pertains to the information presented to the driver. Prior studies have investigated the influence of AV explanation content on driver reactions. AV explanation content can be categorized into three groups: (1) "what" (2) "why" or (3) "what" + "why". The "what" content describes the actions taken or to be taken by the AV [85, 193]. The "why" content specifies the reasons for a specific action taken or to be taken by the AV [85, 86]. The "what" + "why" provides information on both the action taken/will be taken and the underlying reasons [37, 85]. Previous research has demonstrated that varying content can impact drivers' attitudes and behaviors. The literature can be organized into three overarching findings. First, the why-only explanation content leads to the best driver outcomes, promoting positive attitudes, including acceptance, trust, preference, understandability, alertness, and a sense of control, reducing negative emotions such as anxiety, and assisting drivers in driving safely [85, 86, 193]. Second, the what-only explanation content is associated with the worst driver outcomes, resulting in more dangerous driving and reluctance to accept the AV [85]. Lastly, the what + why explanation content has mixed results. Although the what + why explanation induced positive emotional valence and safe driving performance, drivers experienced anxiety and annoyance when receiving the what + why explanation [53, 85, 143]. Furthermore, the effectiveness of the why + what explanation was contingent upon three factors: driving event, driving environment, and explanation point of view [61, 64].

The effectiveness of AV explanations largely depends on their timing. Research on AV explanations has highlighted two distinct timing approaches: providing explanations before or after the AV takes action. Offering explanations before AV actions is generally preferable, as it can elicit positive emotions such as trust and preference while mitigating negative feelings like anxiety and workload [37, 92, 53, 86, 165, 143, 172]. Conversely, presenting explanations after the AV has acted yields mixed outcomes. Although this approach does not enhance trust and preference for AVs [37, 92], it does improve the driver's comprehension of the situation, particularly for less aggressive drivers and following accidents [92, 172].

4.2.2 Explanations and Voice Characteristic

Previous research has predominantly employed auditory and visual modalities to convey explanations to users. Although auditory modality is the more prevalent approach, studies have presented auditory explanations with varied characteristics, potentially overlooking the influence of voice attributes on user perceptions. For instance, Du et al. (2019) [37] utilized a male voice with a standard American accent in a simulation setting, while Körber et al. (2018) [92] and Foster (2017) [53] employed a female voice actor's natural voice for delivering explanations. Moreover, Ruijten et al. (2018) [165] allowed participants to select a gender for the AVs they would operate, integrating a male or female voice into the interface accordingly. However, past research on explanations has insufficiently considered the potential effects of voice features, such as gender, which have been demonstrated to significantly impact perceptions and behaviors in human-technology interactions. This underscores the need for more comprehensive research to examine the role of voice characteristics in explanation design and to improve user experiences.

4.2.2.1 Voice Gender and Preference

The human tendency to anthropomorphize, attributing human-like characteristics to non-human entities, is pervasive, extending to computational devices as indicated by various studies [54, 138, 140]. Intriguingly, this behavior manifests even in the presence of minimal social cues, with individuals prone to ascribe attributes like gender, ethnicity, and age to machines [42, 134]. The critical role of voice in human communication and its increasing integration into technological devices [135] underscore its significance in the development of socially interactive technologies.

A significant element in voice design that has drawn considerable interest is the gender associated with the voice. Most voice assistants are either defaulted to a feminine voice or offer only female voices, which is reportedly based on companies' anecdotal evidence suggesting a general preference for female voices across various cultural and gender demographics [67, 113, 50]. Characteristics such as warmth, gentleness, cooperation, and assistance are routinely linked with female voices [84]. Dong's 2020 study supports this, revealing a general preference for female voices in automated vehicles (AVs) among drivers. This preference is often attributed to the familiarity of female voices frequently used in vehicular navigation systems [35].

Yet, it's debatable whether the universal adoption of female voices across all technological contexts is appropriate. Some studies suggest that the user's voice gender preference can be significantly swayed by the technology's perceived role [84, 115]. This echoes the inclination of users to engage with technology following the same social norms they use in human-human interactions, possibly including gender stereotypes. Factor-analytic research delineates gender stereotypes into two overarching categories: communal and agentic [15, 41, 83]. Communal traits, generally ascribed to women, embody a concern for others, with qualities such as helpfulness, kindness, emotional expressivity, and nurturing [40]. In contrast, agentic attributes, traditionally associated with men, convey assertiveness and control, and include traits like ambition, dominance, independence, and confidence [40].

These stereotype-laden characteristics notably influence user preferences in various contexts. For example, during purchasing decisions, users frequently favor voices that exude authority and confidence—traits linked to male stereotypes. However, when users seek assistance, they tend to prefer voices that are comforting and cooperative, which are traits often associated with women [84, 115]. This correlation mirrors the relationship between gender stereotypes and perceived roles, with women often perceived as adept helpers and problem-solvers, and men viewed as authoritative solution providers like CEOs or supervisors [133, 62, 31]. Interestingly, the perception of dominance extends to technology, where female-voiced computers are deemed less dominant and serious than their male-voiced counterparts when delivering evaluations [139]. Furthermore, regardless of the apparent gender of a robot, studies indicate that humans show a preference and respond more favorably to robots when their gender and occupational roles are congruent [179]. These findings extend to the realm of AVs, where the voice gender preference is influenced by the vehicle's perceived social role—whether it is mainly informative or socially interactive [106]. Users often find a male voice more appealing when delivering information about the AV's actions and environment, while a female voice is favored when the vehicle offers social communication, such as sharing personal anecdotes or addressing user concerns.

Additionally, user gender can significantly affect the preference for the gender of a voice assistant. Existing research indicates that male users tend to prefer male robotic voices, whereas female users lean towards female voices [42]. This gender-aligned preference was further corroborated by Lee et al. (2000), who found that female participants were more inclined to agree with a female Text-to-Speech (TTS) voice, while male participants exhibited a similar preference for male TTS voices [99]. Yet, the interplay between user gender, voice gender preference, and gender-role congruity within the context of AVs remains underexplored. Specifically, the impact of gender similarity, i.e., the matching of the user's gender with the AV's voice gender, and the gender-role congruity, i.e., the alignment between the gender of the AV voice and its role, has not been thoroughly investigated. This understudied area suggests a compelling direction for future research that would not only enhance our understanding of user preferences and behavior but could also provide valuable insights for the design and development of future AV systems.

4.2.3 Trust in Automated Vehicles and Explanation

Previous research has delved into the connection between trust in AVs and the provision of explanations. Drawing from various trust theories, researchers have identified numerous trust dimensions and constructs. For instance, Forster et al. (2017) [53] examined the impact of AV explanations on trust using Lee and See's (2004) [103] trust theory, which encompasses three dimensions: performance, process, and purpose. Their findings indicated that offering an explanation significantly enhanced trust across all dimensions. Similarly, Du et al. (2019) [37] investigated the influence of explanation timing on trust, employing Rempel et al.'s (1985) [156] and Barber's (1983) [6] six trust dimensions (i.e., competence, predictability, dependability, responsibility, reliability, and faith). Their study revealed that delivering explanations before the AV takes action bolstered trust more than providing no explanation or offering an explanation after the AV takes action. Other research, such as those conducted by Ruijten et al. (2018) [165] and Hatfield (2018) [64], has also explored trust in AVs using Sheridan's (1989) [173] trust theory, which centers on familiarity, reliability, and confidence. Ruijten et al. found that furnishing explanations effectively fostered trust in AVs compared to providing no explanation. Conversely, Hatfield's findings demonstrated that offering transparency through explanations did not impact trust during forced moral outcomes.

In summary, trust plays a pivotal role in AV-related research, and supplying explanations can considerably boost trust across multiple dimensions. Further investigation of trust in AVs is essential to enhance and encourage trust in these vehicles moving forward.

4.2.4 Cognitive and Affective Trust

Current research on AVs has yet to thoroughly examine the impact of explanations provided by AVs on trust from both cognitive and affective perspectives. This gap in research presents a significant theoretical challenge for several reasons. First, although trust is acknowledged as a multidimensional factor in AV research, it is predominantly approached from a cognitive perspective, neglecting the distinction between affective and cognitive trust. Existing literature primarily focuses on logical reasons for trusting AVs, overlooking the emotional investment and connection people may develop with AVs. Second, trust theory development has differentiated trust from various states, dimensions, and processes that are sometimes conflated. By distinguishing cognitive and affective trust, we can achieve a higher hierarchical perspective, better understanding and explaining trust.

Interpersonal relationship literature underscores the importance of differentiating between affective and cognitive trust [109, 121]. Cognitive trust in relationships is primarily founded on reasoning and evidence of trustworthiness. Individuals consciously choose whom to trust and why, based on sound reasons supporting the trustworthiness of the person or entity in question [109, p. 970]. Consequently, trust in relationships is a rational, experiential process requiring the identification of reasons to trust and evidence of trustworthiness. Conversely, affective trust, or emotional trust, is a complementary aspect of trust involving emotional connections between individuals. This emotional connection transcends cognitive evaluations of trustworthiness, encompassing a sense of emotional bonding among individuals. Trust relationships involve emotional investments, genuine care, and concern for partners' well-being. A belief in the intrinsic value of these relationships and an expectation of reciprocity in these feelings are present [156, 121, 109]. In summary, interpersonal relationships comprise both cognitive and affective dimensions of trust. While cognitive trust primarily relies on rational evaluations of trustworthiness, affective trust extends beyond reasoning to include emotional connections. Both cognitive and affective trust are essential for establishing and maintaining robust, healthy relationships.

Though trust in AVs has not been extensively studied within the context of Lewis and Weigert's (1985) trust theory, recent research by Lee et al. (2022) [101] aimed to investigate trust in AVs through both cognitive and affective lenses. The study explored the impact of politeness strategies on drivers' perceptions of trust in AVs. The research consisted of two studies. In the first study, video clips depicting various driving scenarios were used to compare the effects of a straightforward demand for required actions from the vehicle (e.g., "Keep paying attention to the road") with politeness strategies that incorporated requests for the driver's assistance (e.g., "Could you keep your hands on the wheel?"), providing a reason for the request (e.g., "School zone is ahead"), and expressing gratitude for the assistance (e.g., "I'd appreciate your help"). The study discovered that employing politeness strategies in AVs can enhance both cognitive and affective trust in the vehicle. The second study used a driving simulator to assess the impact of each politeness strategy on trust in AVs. The results indicated that requesting help and expressing gratitude can boost affective trust in AVs by activating politeness perception and social presence heuristics. However, providing reasons did not significantly influence either cognitive or affective trust in AVs. In conclusion, Lee et al.'s (2022) study emphasize the importance of considering both cognitive and affective trust in AVs. Implementing politeness strategies can improve trust in AVs, particularly regarding affective trust, which is crucial for fostering positive relationships between humans and automated driving systems.

4.3 Hypothesis Development

This study aims to investigate the effects of gender similarity between AV explanation voices' gender and the respective genders of human recipients. Furthermore, it explores the potential moderating impact of gender-role congruity on this relationship. Our primary focus is on understanding how these factors may shape cognitive and affective trust within the AV context. The theoretical model for our investigation is based on the "Computers Are Social Actors" (CASA) paradigm, the Similarity Attraction Theory, and the Role Congruity Theory. This theoretical framework provides us a nuanced perspective on the interplay among these factors and their potential implications on trust formation.

The CASA paradigm fundamentally reimagines the relationship between humans and computers. It pushes beyond the simplistic perspective that views computers as mere lifeless tools, positioning them instead as influential social entities [54, 56, 138, 104]. In this innovative perspective, computers become dynamic constituents within social systems, possessing essential attributes such as agency [138], personality [136, 137], and a sense of social presence [197, 114, 14]. These attributes profoundly reconfigure how humans perceive and interact with digital entities. In the realm of these interactions, voice has emerged as a crucial feature across various technologies. Its role is not peripheral, but central to shaping user experiences and perceptions [42, 140, 195, 93]. Much like the subtle intricacies in human interactions, users form assumptions and modify their behaviors based on the auditory cues technology presents. Interestingly, these voice cues trigger identical brain regions to those stimulated in direct human-to-human exchanges [140], further attesting to the potency of voice as a social medium.

Tailoring the voice of AVs to emulate human characteristics, particularly gender, can profoundly shape users' perceptions of AVs. This notion is anchored in the similarity attraction theory, a well-recognized construct in social psychology. The theory posits that individuals tend to form affirmative connections with others who mirror them across multiple domains [97, 129, 191, 7]. Byrne (1971) succinctly expressed this idea by stating, "similarity between two individuals enhances liking, which in turn influences interactions and behaviors" [19, p. 266]. Consequently, individuals are magnetically drawn to others who reflect their demographic traits, such as age, race, education level, socio-economic status [59, 190, 60], as well as attitudes [150, 174] and beliefs [7, 90]. Grounded in this theory, individuals are likely to experience comfort and validation when surrounded by similar others, thereby fostering robust and deeper relationships [7, 19]. Furthermore, the likelihood of personal information disclosure increases among individuals who perceive similarity with others, thereby enriching the relational experience [7].

Notably, the relevance of the similarity attraction theory extends beyond human-human interactions to human-computer interactions. The CASA paradigm, when combined with the similarity attraction theory, suggests that individuals are more likely to trust and build relationships with entities or technologies that resemble them. In the realm of human-AV interactions, this implies that users might display a higher inclination to trust and engage with AVs if the AV voice is specifically tailored to reflect their characteristics, with particular focus on gender. Cognitive trust in an AV is largely determined by an individual's perception and past experiences relating to its reliability and integrity [109, 121]. It is expected that individuals will be more attracted to and seek information about AVs when the voice of the AV explanation mirrors their own. This can instigate a positive feedback loop where the relationship becomes progressively stable and stronger as trust and similarity mutually reinforce one another. As a result, the customization of an AV's voice to align with users' gender could enhance cognitive trust in AVs.

H1: *Cognitive trust in AVs can be positively influenced by matching the gender of the AV* *explanation voice to those of the user.*

Extending the influence of the similarity attraction theory to affective trust creates a comprehensive understanding of trust dynamics in human-AV interactions. Affective trust is deeply rooted in emotional bonds, feelings of empathy, and mutual care in relationships [109, 121]. When individuals perceive similarities with others - or, in this case, with technologies such as AVs - they tend to experience heightened comfort and emotional security, facilitating the growth of affective trust.The design of AV voice - in particular, its gender - can foster this kind of emotional resonance between the user and the AV. As users perceive the AV voice as similar to their own, it can spark a sense of familiarity and connection, evoking positive emotions that enhance affective trust. This emotional bond not only contributes to the development of trust but also may result in users feeling more understood and represented by the technology. This sense of validation and mutual understanding might drive users' willingness to engage more deeply with the AVs and rely on them more confidently.

H2: *Affective trust in AVs can be positively influenced by matching the gender of the AV explanation voice to those of the user.*

The relationship between gender similarity and trust can be further affected by the moderating role of gender-role congruity. According to role congruity theory, social groups tend to be favorably evaluated when their perceived characteristics align with traditional social roles [40]. Gender stereotypes often play a significant part in these assessments, focusing on the perceived fit between characteristics traditionally associated with certain genders and specific roles. When an individual's gender seems incongruous with their role, it often triggers perceptions of a "lack-of-fit," leading to negative performance evaluations and potential limitations [69, 70, 68, 116]. For instance, the disproportionate representation of males in leadership roles throughout history might result in heightened scrutiny for female leaders, as observers may perceive them as lacking leadership attributes typically associated with males [82].

In the context of AVs, we propose that gender-role congruity could influence user interactions with these vehicles and moderate the impact of gender similarity on both cognitive and affective trust. For cognitive trust, we hypothesize that incongruity between the perceived gender of the AV and the role ascribed to it could lead to less favorable evaluations of the AV, suppressing the belief in the AV's competency to fulfill its designated tasks. As such, the influence of gender similarity on cognitive trust could be less pronounced and the difference between the gender similarity and dissimilarity groups could be less significant in cases where gender and role are incongruent, compared to cases where they are congruent. Affective trust, complementing cognitive trust, could similarly be affected by congruity between role perception and voice characteristics. A mismatch between the perceived role of the AV and its voice characteristics could lead to a negative emotional response and connection, suppressing the affective trust generated by gender similarity. Therefore, we hypothesize that gender-role incongruity will suppress the increase in affective trust due to gender similarity, resulting in a less rise in affective trust compared to gender-role congruity cases.

H3: *The relationship between gender similarity and (a) cognitive trust, and (b) affective trust can be moderated by gender-role congruity. Specifically, the positive effect of gender similarity may be diminished when the gender of the AV voice is incongruent with the perceived role of the AV, resulting in a less noticeable difference in trust between gender similarity and dissimilarity groups.*

4.4 Method

This study received approval from the institutional review board at the University of Michigan in compliance with the ethical guidelines of the American Psychological Association. Informed consent was obtained from all participants.

4.4.1 Participants

Our study was carried out through an online survey involving a group of 326 U.S. drivers, recruited from the diverse participant pool at CloudResearch. The demographic makeup of our participants was well-balanced between genders, comprising 160 women and 166 men. Among the women, there were 75 younger adults whose average age was 22.51 years (with a standard deviation [SD] of 3.09 years), and 85 older adults with an average age of 62.15 years (SD = 5.37 years). In the men's group, there were 88 younger adults with an average age of 22.52 years (SD = 4.85 years) and 78 older adults, averaging 61.83 years of age (SD = 6.16 years). In the preparatory stage before launching our study, we performed a power analysis based on Cohen's (1988) criteria, to accurately determine the necessary sample size. This in-depth analysis predicted a substantial effect size of 0.8 for our research [26]. Subsequently, with the utility of the GPower3.1 statistical tool, we established our alpha at 0.05 and power at 0.8. From this, we determined that a total sample size of fewer than 96 participants would be adequate for conducting the "ANOVA: repeated measures, between factors" group comparison. This led to the realization that our actual participant count of 326 surpassed the basic requirement, thereby solidifying our potential to yield statistically significant results.

To ensure the delivery of high-quality data, we adopted two stringent measures. Primarily, we shortlisted workers who had demonstrated a high performance in prior tasks, reflected by a minimum 95% approval rating and the successful completion of at least 1,000 approved tasks. Additionally, to ward off hurried or disengaged responses, we incorporated two attention-check queries within the body of our survey. We also incorporated eligibility screening to verify participants' suitability for our study. We ensured that they held a valid driver's license, had no visual or auditory impairments that could affect the outcome, and used devices with the ability to play audio content. Upon successful completion of the survey, which generally took between 25 to 30 minutes, participants were compensated with a $5 remuneration.

4.4.2 Study Design

This study implemented a between-subjects experimental design, premised on a two-factor matrix, namely gender similarity and gender-role congruity. Both factors were classified into two categories, thus leading to a 2x2 experimental design. The research aimed to probe the effects of gender similarity between human participants and the voices utilized in the AV explanations. In addition, it investigated the potential moderating effect of gender-role congruity on two distinct types of trust: cognitive and affective.

4.4.2.1 Independent Variable

In this study, we focused on two primary independent variables: the gender similarity between humans and AV explanation voices, and the gender-role congruity. For gender similarity, we categorized participants based on whether their gender matched the gender of the AV explanation voice they heard. As such, the gender similarity variable was split into two groups - similarity and dissimilarity. The similarity group comprised participants who heard an AV explanation voice that matched their own gender (for example, a male participant hearing a male AV voice), whereas the dissimilarity group included instances where the participant's gender did not match the AV explanation voice (for instance, a male participant hearing a female AV voice). Table 4.1 illustrates the total number of participants across AV explanation voice conditions, segmented by gender.

With respect to gender-role congruity, we factored in participants' perceptions of the role

Table 4.1: Total number of participants by gender.

		Voice Gender	
		Female	Male
Participants Gender	Female	85	75
	Male	79	87

Table 4.2: Experimental design and participant distribution.

		Gender-Role Congruity		
		Congruity	Incongruity	
Gender Similarity	Similarity	79	93	**172**
	Dissimilarity	83	71	**154**
		162	**164**	

of AVs and the congruence of this role with the gender of the AV voice. We delineated the perceived roles of AVs into two categories: the ‘driving assistant‘ and the ’driving supervisor.’ Participants indicated their perceived role of an automated vehicle on a slider scale. Subsequently, we matched the AV voice gender with the participant’s self-identified perceived AV role, adhering to prevailing gender stereotypes (i.e., male for driving supervisor; female for driving assistant) [133, 62, 31]. Participants who perceived the AV as a ’driving assistant’ and heard a female voice were categorized under the gender-role congruity group. Conversely, those perceiving the AV as a ’driving assistant’ but encountered a male voice were categorized under the gender-role incongruity group. The distribution among the gender similarity groups and gender-role congruity groups can be observed in Table 4.2.

In our study, we created AV voices using two text-to-speech platforms: Murf and Uberduck. These services offer the ability to transcribe text into human-like speech, allowing for the creation of varied personas based on different demographic characteristics [192]. To counterbalance the potential interaction effect of a participant’s age and the perceived age of the AV voice, we crafted four distinct voices characterized by two levels of gender (male and female) and age (younger and older). To minimize bias, we deliberately assigned participants from different age groups to corresponding voice conditions. The Murf Text-to-Speech application facilitated the creation of ’younger’ voices, resulting in two personas: Natalie (young female voice) and Nate (young male voice). On the other hand, we utilized the Uberduck platform to generate ’older’ voices, leading to two additional personas: Charlotte (older female) and Jim (older male). Finally, to ensure an immersive experience for our participants, these voice files were synchronized with pre-recorded scenarios using the CapCut editing tool. This ensured that the spoken explanations were appropriately matched with the corresponding actions demonstrated in the videos.

Throughout the study, participants engaged with six video scenarios demonstrating AVs in action from the driver's perspective. Figure 4.1 provides a glimpse of one such scenario titled "Oversized vehicle ahead". These scenarios spanned various driving contexts, including urban, highway, and rural environments, portraying the appropriate responses of the AV in each situation. To ensure the clarity of AV actions, each scenario included a comprehensive "what+why" explanation delivered by the AV voice. These explanations articulated the actions the AV was about to undertake and provided reasons behind such decisions, serving to inform participants about the functioning and reasoning capability of the AV system. Appendix B offers a detailed outline of these "what+why" explanations for each driving scenario.

Figure 4.1: A video screenshot for "Oversized Vehicle Ahead" scenario

4.4.2.2 Dependent Variables

This study evaluated two trust-related dependent variables, namely cognitive trust and affective trust. An overview of all utilized questionnaires is provided in Table 4.3.

Cognitive Trust: To assess cognitive trust in AVs, we implemented a seven-item measure adapted from McAllister's (1995) [121] and Lee's (2022) [101] studies. To suit the context of AVs, we modified these items accordingly. Participants were asked to rate each item on a 7-point Likert scale, with "1" denoting "strongly disagree" and "7" signifying "strongly agree".

Table 4.3: Factor loading.

Variable	Items	Cronbach's Alpha	Component 1	Component 2
Cognitive Trust	This self-driving car demonstrates expertise.	0.93	0.83	
	Given the driving behavior of this self-driving car, I see no reason to doubt its competence.		0.83	
	I can rely on this self-driving car's driving capabilities.		0.83	
	I believe that other drivers who use this self-driving car will regard it as trustworthy.		0.80	
	Most people will trust and rely on this self-driving car, even those who don't know much about it.		0.64	
	This self-driving car's analysis of driving situations was accurate.		0.82	
	I can trust this self-driving car's instructions.		0.83	
Affective Trust	I can freely share my concerns with this self-driving car.	0.95		0.85
	This self-driving car will listen to me when I share my struggles in understanding its actions.			0.87
	If this self-driving car was no longer available to me, I would feel a sense of loss.			0.76
	This self-driving car would respond caringly if I shared my concerns with it about driving.			0.90
	I would make considerable emotional investments in this self-driving car.			0.82
	This self-driving car will help me with great care.			0.79
	This self-driving car will kindly help me when I need it.			0.83
	This self-driving car will take care of me with thoughtful consideration.			0.84

Affective Trust: We evaluated affective trust via a questionnaire formulated on the basis of McAllister's (1995) [121] and Lee's (2022) [101] affective trust questionnaires. The participants were instructed to rate each item on a Likert scale from 1 indicating strong disagreement to 7 implying strong agreement.

4.4.3 Study Procedure

Participants for this study were recruited through the 'Connect' crowdsourcing platform on CloudResearch. Those who agreed to participate were directed to an online survey hosted on the Qualtrics platform. Prior to starting the study, participants were provided with a clear and concise explanation of the study's objectives, and they were informed that their participation was voluntary. They were also instructed to ensure that their devices had functioning audio and visual capabilities to fully engage with the study materials.

After providing their informed consent, participants were given an overview of the capabilities of the AV used in the study. This summary emphasized that the AV was fully autonomous, adhered to traffic regulations, and could modify its routes based on alternative suggestions or available navigation data from sources like Google Maps. Before proceeding with the main study, participants were asked to provide information about their driving experience, previous encounters with AVs and automated driving systems, and their general perceptions of AVs. This background information aimed to capture participants' relevant experiences and perspectives.

The study consisted of six videos, each presenting a unique driving scenario from the viewpoint of the AV's driver's seat. After watching each video, participants were required to complete a survey designed to assess their perceptions of the AV, including measures of affective and cognitive trust. These surveys allowed participants to provide feedback on their emotional connection with the AV and their beliefs about its competence and reliability. Upon completion of all six videos and surveys, participants were asked to provide demographic information, including their age, gender, educational attainment, race/ethnicity, and income. This information helped ensure a diverse participant pool and enabled the examination of potential demographic influences on perceptions of AVs.

4.5 Results

4.5.1 Reliability and Construct Validity

An assessment of construct validity and reliability was undertaken to confirm the appropriateness and consistency of the measurement constructs employed in this study. Construct validity pertains to the degree of accuracy with which a scale encapsulates the concept it intends to measure [5]. This form of validity is further categorized into convergent and discriminant validity, both of which were evaluated using exploratory factor analysis in this study. Convergent validity is exhibited when scale items have factor loadings of 0.60 or above on their corresponding constructs, while discriminant validity is demonstrated when scale items bear loadings of 0.35 or below on unrelated constructs [52]. As delineated in Table 4.3, all scale items either met or surpassed these benchmarks, hence affirming their validity. Construct reliability signifies the internal consistency of a set of scale items [144]. A prevalent metric for assessing this is Cronbach's alpha [177, 30]. Table 4.3 showcases that the reliability of all constructs either met or exceeded the generally acceptable threshold of 0.70, thereby confirming their reliability. In addition, Table 4.4 lists the means, standard

Table 4.4: Descriptive statistics

Variable	**Mean**	**Std.Dev**	**Anxiety**	**Unsafety**
Cognitive Trust	5.165	1.033	1.000	
Affective Trust	3.541	1.372	0.475**	1.000
*Correlation is significant at the 0.01 level (2-tailed).				

deviations, and correlations.

4.5.2 Hypothesis Testing

The hypothesis was evaluated using a sample consisting of 326 participants. We utilized linear mixed models to investigate the effects of gender similarity (between the participant and the AV explanation voice) and gender-role congruity on the dependent variables, namely cognitive and affective trust. This statistical method allowed us to discern if there were significant variations between the mean values of independent groups, based on these two unique factors. To account for non-independence in all the linear mixed models, participants were considered as random effects. All statistical analyses were executed using IBM SPSS 28.0 statistical software. The significance threshold for all statistical tests was established at an alpha level of 0.05. To control for Type I error across multiple comparisons, we applied a Bonferroni correction to all post hoc analyses.

4.5.2.1 Manipulation Check

As a crucial component of the manipulation check, participants were asked to discern the perceived age and gender of the voice presented in each video after viewing it. The outcomes of this process confirmed that our manipulation of the AV explanatory voices was effective and produced the intended impact on the participants. More specifically, a significant discrepancy was observed in the perception of voice age between the younger and older voice groups ($F(1, 1956) = 1704.715$, $p < 0.001$). Participants were able to successfully differentiate between younger and older voices, validating the age-based voice manipulation. In addition, participants demonstrated a high success rate in distinguishing the intended voice gender manipulations ($F(1, 1956) = 25142.004$, $p < 0.001$). This result suggests an accurate perception of the gender of the AV explanatory voices, affirming the effectiveness of the manipulation pertaining to gender. These findings collectively confirm the reliability of our manipulations in the study.

4.5.2.2 The Effect of Gender Similarity on Cognitive Trust

To examine H1, which explores the main effect of gender similarity on cognitive trust, a one-way ANOVA was conducted. The results demonstrated a statistically significant difference in cognitive trust between the gender similarity and dissimilarity groups ($F(1, 1956) = 7.310$, $p = 0.007$). Figure 4.2 illustrates the mean scores, with participants in the gender similarity group (mean = 5.224, SD = 0.961) displaying higher levels of cognitive trust compared to those in the dissimilarity group (mean = 5.098, SD = 1.104). These findings provide support for H1, suggesting that aligning the gender of the AV voice with that of the user positively impacts cognitive trust.

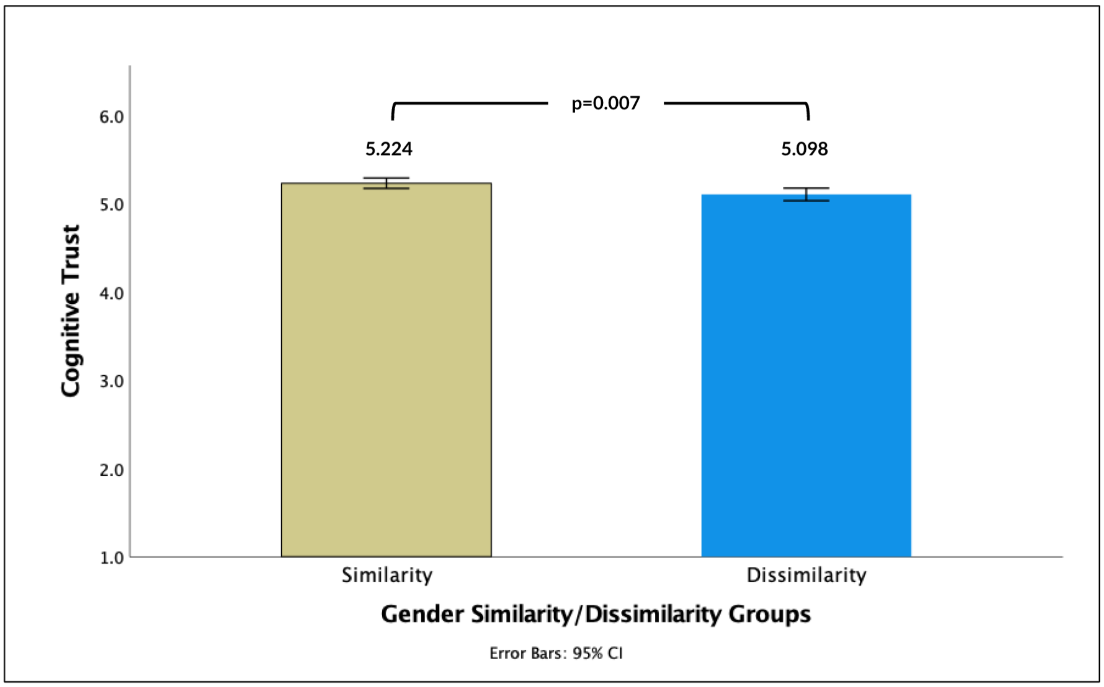

Figure 4.2: The average scores of cognitive trust between gender similarity/dissimilarity groups

4.5.2.3 The Effect of Gender Similarity on Affective Trust

The analysis of our data demonstrated a significant main effect of the gender similarity/dissimilarity group on affective trust ($F(1, 1956) = 28.017$, $p < 0.001$). Specifically, when the gender of the human and AV explanation voice matched, participants reported higher levels of affective trust (mean = 3.695, SD = 1.315) compared to the dissimilarity group (mean = 3.369, SD = 1.413). These findings provide support for H2, indicating that matching the gender of the AV voice to that of the user positively influences affective trust.

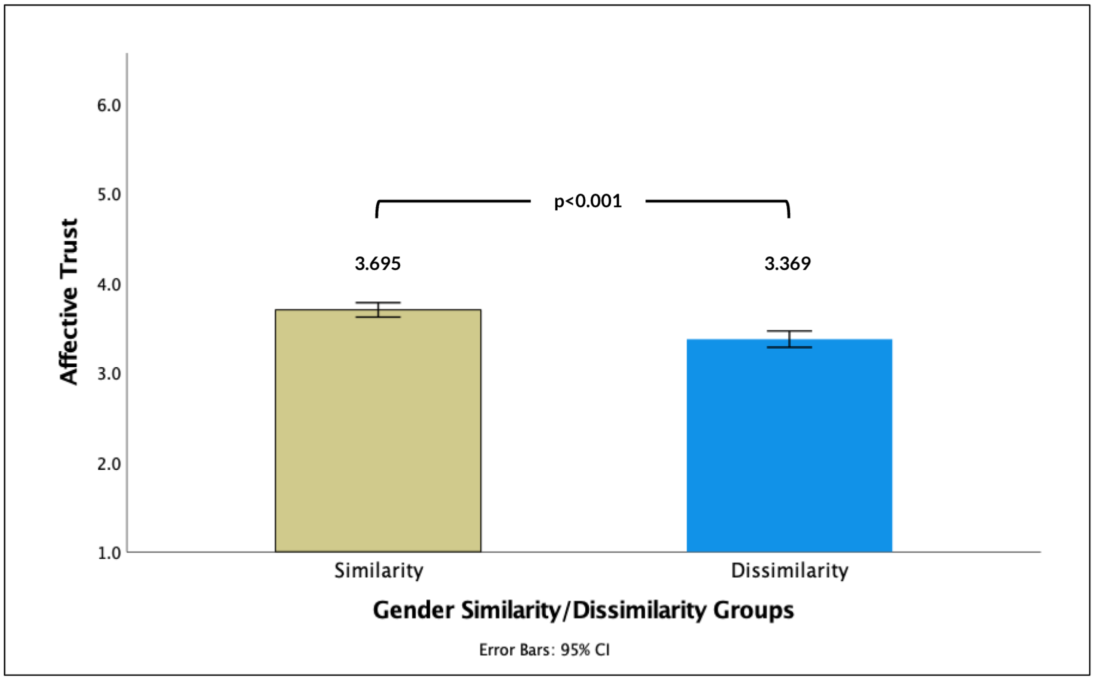

Figure 4.3: The average scores of affective trust between gender similarity/dissimilarity groups

4.5.2.4 The Interaction Effect of Gender Similarity and Gender-role Congruity on Cognitive Trust and Affective Trust

To investigate the potential moderating effect of gender-role congruity on cognitive trust (H3), a two-way ANOVA was conducted, including both gender similarity and gender-role congruity as factors. The analysis did not reveal a significant interaction effect between gender similarity and gender-role congruity on cognitive trust ($F(1, 1954) = 1.560$, $p = 0.212$), as presented in Table 4.5. However, the results did show a significant main effect of gender similarity/dissimilarity on cognitive trust ($F(1, 1954) = 7.926$, $p = 0.005$). These findings suggest that the impact of gender similarity on cognitive trust is not influenced by gender-role congruity. Therefore, H3a, which posits a moderating effect of gender-role congruity, was not supported.

For affective trust, a significant two-way interaction between gender similarity and gender-role congruity was found ($F(1, 1954) = 5.154$, $p = 0.023$), as depicted in Figure 4.4 and supported by the results in Table 4.6. The effects of gender similarity/dissimilarity were observed in both the gender-role congruity groups of participants, but the impact was more pronounced in the gender similarity group. Specifically, in the gender-role congruity group (represented by the orange line in the figure 4.4), when the gender of the AV voice and the

Table 4.5: ANOVA summary table of cognitive trust.

Source of Variation	**Sum of Squares**	**df**	**MeanSquare**	**F**	**p**
(Intercept)	51710.18	1	51710.18	48681.805	0
Gender Similarity (GS)	8.419	1	8.419	7.926	0.005**
Gender-Role Congruity (GRC)	2.282	1	2.282	2.149	0.143
GS x GRC	1.657	1	1.657	1.56	0.212
Error	2075.554	1954	1.062		
Total	54314.878	1957			
Note: "df" indicates degree of freedom; "F" indicates F statistic; "p" indicates p value.					

Table 4.6: ANOVA summary table of affective trust.

Source of Variation	**Sum of Squares**	**df**	**MeanSquare**	**F**	**p**
(Intercept)	24277.337	1	24277.337	13427.767	0
Gender Similarity (GS)	62.82	1	62.82	34.746	0.001**
Gender-Role Congruity (GRC)	84.743	1	84.743	46.871	0.001**
GS x GRC	9.318	1	9.318	5.154	0.023*
Error	3532.822	1954	1.808		
Total	28237.297	1957			
Note: "df" indicates degree of freedom; "F" indicates F statistic; "p" indicates p value.					

perceived role were congruent, the gender similarity group displayed a significantly higher level of affective trust in the AV (mean = 3.996, SD = 1.169) compared to the dissimilarity group (mean = 3.498, SD = 1.466, $p < 0.001$). However, in the gender-role incongruity group (represented by the green line in the figure 4.4), where the gender and role were incongruent, the effect of gender similarity on affective trust was weakened. The increase in affective trust when the voice gender matched the user's gender was not as substantial as in the gender-role congruity group, and the difference between the similarity (mean = 3.440, SD = 1.378) and dissimilarity group (mean = 3.218, SD = 1.334) was less statistically significant ($p = 0.011$). Moreover, posthoc comparisons among these groups showed notable differences. For both gender similarity ($p < 0.001$) and dissimilarity groups ($p = 0.015$), the affective trust was significantly higher in the gender-role congruity group than in the incongruity group. Nonetheless, our analysis revealed no significant difference between the groups characterized by gender similarity paired with gender-role incongruity and those exhibiting gender dissimilarity alongside gender-role congruity ($p = 0.483$). This intriguing outcome suggests that gender similarity is not invariably linked to enhanced affective trust. Notably, when gender similarity is coupled with gender-role incongruity, participants demonstrated a trust level comparable to the group displaying gender dissimilarity and gender-role congruity.

These findings suggest that the relationship between gender similarity/dissimilarity and affective trust is subject to the moderating effect of gender-role congruity. When there is congruence between the gender of the AV voice and the perceived role, gender similarity significantly enhances affective trust. However, when there is incongruence between the gender and role, the effect of gender similarity on affective trust is attenuated. Therefore, H3b, which proposes the moderating effect of gender-role congruity on affective trust, was supported.

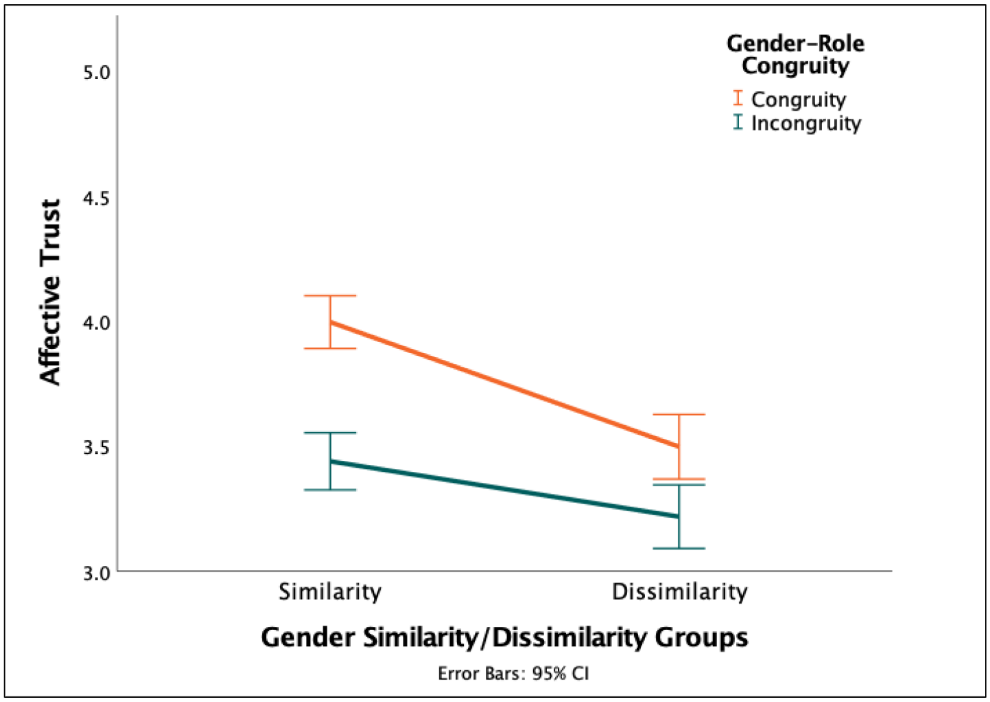

Figure 4.4: Effect of two-way interaction between gender similarity and gender-role congruity on affective trust

4.5.3 Summary of the Results

The study investigated the influence of gender similarity between human users and the gender of the AV explanation on both cognitive and affective trust. It further explored whether these relationships could be moderated by the factor of gender-role congruity. For cognitive trust, our findings demonstrate a significant impact of gender similarity. Specifically, the similarity between the user's gender and the AV explanation voice's perceived gender led to higher levels of trust compared to the dissimilarity group. Thus, H1 was supported. Affective trust followed a similar trend. Gender similarity significantly influenced people's perceptions and their affective trust in the AV. Higher levels of affective trust were observed when the gender of the user matched the perceived gender of the AV. Therefore, H2 was supported. When considering the potential moderating effect of gender-role congruity, this study found that its impact on the relationship between gender similarity and cognitive trust was not significant. The primary determinant of cognitive trust was gender similarity alone, which did not support H3a. However, the results pertaining to affective trust were different. While gender similarity continued to impact both participants in the gender-role congruity/incongruity groups, its effect was less pronounced for the incongruity group. That is, in instances where the AV explanation voice gender and perceived AV role were incongruent, the significant effect of gender similarity on affective trust was mitigated. This resulted in a less substantial difference in affective trust levels between the gender similarity and dissimilarity groups. Consequently, H3b was supported. The following section provides a more in-depth discussion of the findings and their contributions to the literature, along with the study's limitations.

4.6 Discussion

4.6.1 Gender Similarity and Cognitive and Affective Trust

Prior research in the AV field has predominantly leveraged auditory explanations to communicate information about AV actions and underlying rationale. However, a thorough exploration of diverse voice characteristics, such as gender, remains somewhat neglected despite the notable impact these characteristics can have on attitudes and behaviors towards speech-based technology. Modern technologies frequently utilize varying voices, characterized by different genders, to deliver information. However, there is no consensus on which voice type is most preferred or generates the most favorable outcomes. While some studies suggest a general preference for a female voice in voice assistants due to its comforting and warm delivery [84, 35], other research proposes the user's gender might moderate their preference for a technology's voice [42, 99]. The similarity attraction theory suggests that individuals are more likely to trust and establish relationships with entities, including technologies, that resemble themselves [42, 43, 99].

Our research endeavor seeks to extend the existing body of literature into the AV domain and reconcile these discrepancies. To this end, we scrutinize trust within the AV sphere through the lens of two distinct perspectives: cognitive trust and affective trust. Cognitive trust is primarily based on reasoning and evidence of trustworthiness. In the AV context, it involves a rational, experiential process where users discern reasons to trust AVs. Notably, our research emphasizes the importance of both the gender of the AV explanation voice and the user in shaping cognitive trust. Consistent with the similarity attraction theory, we found that users tend to trust AVs more when the voice gender is similar to their own. This preference stimulates active engagement, enhancing understanding of the AV's logic and operational environment. Positive feedback loops thus reinforce the human-AV relationship as users gain deeper insight into the AV's competencies and capabilities, providing rational and experiential grounds for placing trust in AVs. Consequently, the customization of an AV's voice to align with users' gender could enhance cognitive trust. In relation to affective trust, our findings clearly demonstrate an enhanced level of trust when the gender of the AV explanation aligns with the user's own gender. These observations underscore the validity of the similarity-attraction theory, which proposes that individuals are more inclined to connect with entities resembling themselves. In this context, users are likely to experience a greater sense of comfort and connection when interacting with AVs that mirror their own gender. This familiarity triggers positive emotional responses, thereby augmenting affective trust.

Our findings play a pivotal role in resolving the ongoing ambiguity regarding the preferred gender of the voice used in the AV domain. Historically, a female voice has been favored due to its perceived comforting presence within technological environments. However, our study indicates that this may not universally apply across all contexts and user groups. The choice of voice gender should not be a static decision but rather a dynamic one that can adapt according to the user's gender. This flexibility can promote a deeper connection, understanding, and ultimately, trust in AV systems, optimizing user interaction and engagement.

4.6.2 Gender Similarity, Gender-Role Congruity, and Cognitive and Affective Trust

In addition to the effect of gender similarity between humans and AV explanation voice, existing literature also indicates potential influences of gender-role congruity on perceptions and the adoption of technology. Prior research suggests that a technology's perceived role can significantly influence a user's preference for the gender of the voice [84, 115]. This mirrors the propensity of users to interact with technology adhering to the same social norms they employ in human-to-human interactions, which may involve gender stereotypes [74, 106, 179]. Users frequently favor male voices that project authority and confidence—characteristics stereotypically associated with men in decision-making scenarios. In contrast, when users seek assistance, they tend to prefer voices exuding comfort and cooperation, traits often linked to women. As per the role congruity theory, social groups are generally evaluated positively when their perceived attributes align with traditional social roles. However, when an individual's gender seems incongruous with their role, it frequently triggers perceptions of a "lack-of-fit", resulting in negative performance evaluations and potential limitations

[69, 68, 70, 116].

The influence of these gender-role congruity factors forms a pivotal component of our study. We seek to explore their potential moderating effects on the relationship between gender similarity and trust—both cognitive and affective—in AVs. With regard to cognitive trust, we found no evidence to support a moderating effect of gender-role congruity on the relationship between gender similarity and cognitive trust. Instead, we identified a prominent main effect: gender similarity significantly promotes people's cognitive trust in AVs. In terms of affective trust, we found the benefits of gender similarity to be less pronounced when the gender of the AV voice is incongruent with the perceived role of the AV, leading to a more gradual increase in trust. These results bring forth three crucial insights for the field.

First, in the context of affective trust, we discovered that gender similarity between humans and the AV voice plays a substantial role in enhancing trust. However, when a mismatch arises between the voice gender and the perceived role of the AV, it seems to impair people's emotional engagement. This in turn dampens the beneficial impact of gender similarity on affective trust. This observation not only resonates with prior research pointing towards the detrimental impact of gender-role incongruity on perceptions, but it also draws attention to its intricate relationship with gender similarity in molding affective trust.

Second, we noted no interaction effect between gender similarity and gender-role congruity on cognitive trust. This suggests that while gender-role incongruity could undermine people's emotional responses, it doesn't seem to impinge upon their rational assessments. In essence, individuals depend on other sources of information to formulate and validate reasons to place cognitive trust in the AV (e.g., the positive reinforcement driven by gender similarity).

Lastly, the contrasting patterns regarding the interaction effect of gender similarity and gender-role congruity between cognitive and affective trust underscore the need to distinguish between these two forms of trust. This insight carries significant implications for both theoretical advancements in trust research and practical applications in the design of AVs to maximize user trust.

4.6.3 Design Implications

The findings of this study have several implications for the design of AV voices. First, the gender of the AV voice should be deliberately designed considering the human user's gender and their perceived role of the AV. The AV should ideally use a voice that matches the user's gender to explain its actions. For example, male users should be presented with a male voice, while female users should be presented with a female voice. This strategy not only promotes cognitive trust but also helps build an emotional bond between the user and the AV, fostering affective trust and further enhancing the human-AV interaction.

Second, although gender similarity consistently leads to positive outcomes, its impact on affective trust can be compromised by gender-role incongruity. Therefore, AV design should consider not only the user's gender but also the users' perceived role of the AV to further bolster affective trust. It is generally advantageous to combine gender similarity and gender-role congruity to generate the highest level of affective trust. For instance, for male users who perceive the vehicle as a driving supervisor, a male voice explaining the AV's actions would be the optimal choice. Conversely, for female users who perceive the vehicle as a driving assistant, a female voice would be ideal. Nevertheless, scenarios where the AV voice neither matches the user's gender nor their role perception should be avoided, as these can lead to the lowest levels of affective trust. An example of such a situation would be providing a female voice for male users who perceive the vehicle as a driving supervisor.

In order to optimize the benefits of AVs and facilitate the relationship between humans and AVs, both the user's gender and perceived role of the AV should be carefully considered. This can be accomplished by setting a default voice that aligns with these factors, educating users about the potential functions and roles of the AV, and allowing drivers to personalize the voice to their preference. By addressing these elements, designers can more effectively encourage trust and user adoption of AV technology.

4.6.4 Limitations and Future Research

The limitations of this study are worth noting. First, participants were sourced from an online subject pool familiar with the format of online studies, which may not adequately represent the broader population's knowledge and experience within the AV field. Second, while the experimental setup provided strong internal validity, its external validity may be limited, necessitating future studies in field settings to bolster generalizability. Third, it's possible that participants engaged in hypothesis guessing, altering their reactions and responses based on their assumptions of the researcher's desired outcomes—although no evidence of this was found in our study. Moreover, this study exclusively addressed the gender aspect of the human-voice characteristics, leaving other traits such as personality, age, and race/ethnicity unexplored. Future research could delve into these interactions between various human and voice characteristics to determine their potential impact on human-AV interactions, providing invaluable insights for the design of future AVs. Additionally, this study did not examine other features associated with AV explanations, including the definition, generation, selection, and evaluation of alternative courses of action for individuals. These and other potential attributes related to AV explanations warrant further investigation. In conclusion, a more comprehensive understanding of the impact of AV explanations and their effective design requires additional research.

4.7 Conclusion

In this study, we delved into the effects of gender similarity and gender-role congruity on both cognitive and affective trust. To our knowledge, this is the first exploration into the impact of voice characteristics on explanation effectiveness within the context of Level 5 automation. Our findings underscore the significance of considering gender similarity between human users and the AV explanation voice, along with its interplay with gender-role congruity, when determining the effectiveness of AV explanations from the standpoints of cognitive and affective trust. Overall, our research contributes to a more comprehensive understanding of the elements influencing AV explanation effectiveness, providing invaluable insights that will prove essential for the design of future automated vehicles.

CHAPTER 5

Discussion and Conclusion

In this dissertation, we embark on an exploration of the intricate dynamics that surround the relationship between user expectations and trust formation in Automated Vehicles (AVs). This analysis is underpinned by comprehensive consideration of the cognitive and emotional mechanisms that drive trust creation. The work conducted throughout the course of this research comprises three in-depth studies, each geared towards illuminating the impact of expectations and trust on AV adoption. In doing so, it also contributes new insights to the growing body of AV literature, particularly in the context of AV design and human-AV interaction.

Chapter 2 launches our examination with a study on initial AV expectations across a diverse range of user groups. With a keen focus on identifying variances in pre-experience expectations, we aim to categorize distinct expectation levels among prospective users. This foundational understanding of users' anticipations provides an important baseline for the analysis that follows.

Chapter 3 takes a detailed look into the impact of expectations on the construction of trust in AVs. This segment scrutinizes the cognitive assessment of divergences between pre-existing expectations and the actual performance of AVs, as well as the subsequent effects on trust development. Furthermore, we explore how risk perception could potentially moderate this dynamic. Drawing upon the Expectation Disconfirmation Theory (EDT), our findings present an intricate relationship between expectations, actual performance, and the consequent disconfirmation that arises. This complex interplay substantially informs the development of cognitive trust in AVs, an insight that carries considerable implications for both theoretical understanding and practical applications in the realm of AV technology adoption.

Finally, in Chapter 4, we underscore the indispensable role of AV design in fostering trust. Through the third study, we aim to elevate human-AV interaction and perception by advocating for the inclusion of explanation voice characteristics in AV design. This design feature seeks to promote both cognitive and affective trust, guided by the principles of the similarity attraction theory and role congruity theory. By enhancing the understandability and approachability of AVs, we aim to facilitate stronger, more positive human-AV interactions, and ultimately, higher levels of AV adoption. The four main contributions of this dissertation are presented in detail below.

5.1 Contributions

5.1.1 Unraveling the Influence of Individual Differences on Expectations for Automated Vehicles

Chapter 2 of this dissertation undertakes a comprehensive exploration of the intricate ways individual differences shape expectations for autonomous vehicles (AVs), both validating and extending upon established literature in the field. Our study not only corroborates previous research suggesting that male drivers and those with higher levels of education tend to harbor greater expectations for AVs, but also broadens the understanding of this area by identifying additional influential factors.

Key among these new insights is the impact of various demographic and personality characteristics on AV expectations. We found that younger drivers and those who are unmarried tend to have more optimistic outlooks on AVs. Similarly, those who engage in frequent driving, despite having comparatively less driving experience, also manifest higher expectations of this technology. Equally noteworthy is the role of personality traits in shaping AV expectations. Our research revealed that individuals characterized by high levels of extraversion, agreeableness, emotional stability, and conscientiousness tend to be more optimistic about the capabilities and potential of AVs. These findings suggest that individual personality traits could serve as potent predictors of attitudes towards AVs.

This enriched understanding of how individual differences influence attitudes and behaviors towards AVs contributes to a comprehensive picture of the dynamics underpinning trust, satisfaction, and the intention to adopt this emerging technology. Through this research, we pave the way for more targeted and effective strategies for AV introduction and acceptance, catering to the unique expectations of diverse user groups.

5.1.2 Illuminating the Role of Expectations in Trust Formation and Automated Vehicle Adoption

Chapter 2 of this dissertation furnishes a more comprehensive understanding of the complex interplay between individual characteristics, expectations, and the adoption of AVs. Our findings affirm that older drivers, who typically maintain lower expectations of AVs, harbor more negative attitudes toward their adoption, in alignment with existing literature. Concurrently, we reinforce previous studies that suggest higher levels of AV acceptance among male drivers and those with advanced educational levels, as these groups generally foster higher expectations of AVs. Moreover, this research broadens our comprehension of the dynamics of technology adoption. By pinpointing demographic groups prone to excessively low or high AV expectations, we pave the way for the design of interventions to moderate these biases and, in turn, cultivate a more extensive acceptance of AVs.

Building upon the groundwork laid in Chapter 2, Chapter 3 delves deeper into the mechanisms and processes through which expectations impact trust. Drawing on expectation disconfirmation theory, we underscore the key role of performance-related disconfirmation in trust dynamics. Our research indicates that positive disconfirmation, which occurs when the performance of AVs surpasses expectations, engenders trust. Conversely, negative disconfirmation, resulting from performance that does not meet expectations, can erode trust. Further, our research brings to light the significant role of perceived risk—both internal and external—in modulating the impact of disconfirmation on trust. We discern that perceived risk considerably affects the velocity of trust change. Specifically, trust degrades more quickly under low-risk conditions when expectations go unmet, while it accelerates more rapidly in low-risk scenarios when performance outstrips expectations. These findings underscore that individuals' perceptions of safety, predictability, and the intricacy of the driving conditions significantly sway their trust in AVs. This revelation emphasizes the multifaceted relationship among expectations, trust formation, and the eventual adoption of AVs.

5.1.3 Suggesting Design Strategies for Optimizing Automated Vehicle Adoption

Chapter 3 of this dissertation underscores valuable insights with far-reaching implications for the design of AVs, especially in promoting user trust and catalyzing extensive adoption. The findings reveal the critical role of fulfilling, or ideally surpassing, user expectations. As such, this chapter advises the necessity for clear and open communication about the abilities and restrictions of AV systems. This level of transparency is instrumental in setting realistic user expectations concerning AV performance. Furthermore, the chapter accentuates the need for creating versatile AVs that can handle a wide array of driving conditions. The adaptability of AVs to different environments and situations is essential in meeting diverse user preferences.

Chapter 4 of this dissertation underscores the significance of careful and intentional design of AV voices, tailored according to the user's gender and their perceived roles of AVs. Such design considerations are crucial for building user trust and enhancing user engagement, drawing from the tenets of similarity-attraction theory and role congruity theory. This study serves as a pioneering effort to delineate cognitive trust from affective trust, recognizing that existing research has predominantly centered on the logical elements that contribute to trust in AVs, while often sidelining the emotional bonds users form with these systems. By differentiating cognitive from affective trust, our research offers a more in-depth and nuanced understanding of how trust is cultivated within the AV context. Our findings propose that fostering both cognitive and affective trust is facilitated when the gender of the AV voice aligns with that of the user. Moreover, the study underlines that the interplay between gender similarity and gender-role congruity significantly influences affective trust. Specifically, when there is an incongruity between the voice's gender and the perceived role, it could potentially soften the effect of gender similarity on affective trust.

This chapter, in presenting the interaction between gender similarity and gender-role congruity, provides tangible insights for AV voice design aimed at optimizing user interaction and engagement. Our research accentuates the need for a holistic approach in the design of AVs, one that incorporates both logical and emotional aspects. This could catalyze a heightened level of user trust, potentially driving higher acceptance and adoption rates of AVs.

5.2 Limitation and Future Work

Our research offers valuable insights into the relationship between individual traits, expectations, performance perception, and trust in AVs. However, it's important to note its limitations.

First of all, the dissertation, composed of three individual studies, heavily relies on survey methods and video simulations to gather data. While these methodologies have their benefits, such as providing controlled environments and easy data collection, they come with inherent drawbacks that may limit the scope and applicability of the findings. First, surveys, although widely used due to their efficiency in collecting large amounts of data, could potentially introduce method bias. This can distort the data, rendering results that are not entirely reflective of real-world scenarios. Second, the use of video simulations, while useful in presenting standardized scenarios to all participants, may not perfectly emulate real-world conditions. The virtual environment can't fully reproduce the intricacies and unpredictability of actual road conditions, thus limiting the extent to which we can generalize the findings. Given these limitations, future research should consider adopting more realistic, field-based methodologies to supplement survey and simulation data. Real-world experiments, like driving tests in actual AVs or interactive experiments in controlled traffic environments, could provide richer, more accurate data about drivers' behaviors and perceptions. In addition, using mixed methods research - blending qualitative and quantitative approaches - can yield more nuanced insights. For example, in-depth interviews or focus group discussions could reveal underlying reasons or motivations for drivers' attitudes towards AVs that might not be captured by survey or simulation methods.

Second, while the investigation of demographic and driving-related factors provides a solid starting point, it is imperative to broaden the scope of research in order to fully comprehend the intricacies of perception and trust in AVs. To achieve a comprehensive understanding of trust dynamics in AVs and their adoption, it is essential for future research to explore the main or moderating effects of additional factors. One such variable to explore is an individual's social system, which encompasses the network of interpersonal influences, media exposure, cultural norms, and societal pressures. The cultural context and societal norms prevalent in a given society can significantly shape the level of trust placed in technological innovations such as AVs. A society that embraces and welcomes technological advancements may foster individuals who are more inclined to trust and utilize AVs. By examining the impact of social systems on trust in AVs, researchers can unravel the complex interplay between societal factors and individual perceptions. Furthermore, the propensity to be an early adopter of new technologies warrants consideration. Drawing on the Diffusion of Innovations theory, individuals who demonstrate a predisposition towards early adoption are likely to be more receptive to embracing AVs. Their openness to change and willingness to adapt to new technologies can influence their acceptance and adoption of AVs. Understanding the characteristics of these early adopters, such as their motivations and attitudes, can inform strategic approaches to the introduction and implementation of AV technology. In conclusion, by delving into the influence of factors such as the social system and early adopter propensity, future research can enhance our understanding of the variables that shape perceptions and trust in AVs. Expanding the scope of investigation will provide a more nuanced perspective and contribute to the development of a comprehensive model of AV acceptance.

APPENDIX A

Questionnaires

A.1 Individual Difference and Expectation

A.1.1 Expectation Questionnaire

Please answer the following questions. (Low=1; High=7)

- How would you rate your overall expectations regarding the driving of a self-driving car?
- How would you rate your expectations regarding the effectiveness of a self-driving car?
- How would you rate your expectations regarding the safety of a self-driving car?

A.1.2 Demographics and Personality Questionnaires

Age: What is your age?
Gender: How would you describe your gender?

- Male
- Female
- Prefer not to answer
- Prefer to self describe

Region: Which region do you live in?

- Northeast (Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont)

- Midwest (Illinois, Indiana, Iowa, Kansas, Michigan, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin)
- South (Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, District of Columbia, and West Virginia)
- West (Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming)

Ethnicity: Please specify your ethnicity.

- White non-Hispanic
- Black or African American, non-Hispanic
- Hispanic
- Other

Education: What is the highest level of education you have received or are pursuing?

- High school graduate (high school diploma or equivalent including GED)
- Some college
- College graduate

Marital Status: What is your marital status?

- Never married
- Married
- Living with partner
- Widowed
- Divorced/Separated

Income: What was your total household income before taxes during the past 12 months?

- Less than $25,000
- $25,000 to $34,999

- $35,000 to $49,999
- $50,000 to $74,999
- $75,000 to $99,999
- $100,000 to $149,999
- $150,000 or more

Driving Frequency: How often do you drive?

- Drives almost every day
- Drives sometimes or rarely
- Never drives

Driving Experience: By year

Personality: Here are a number of personality traits that may or may not apply to you. Please select a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other. (Strongly disagree=1; Strongly agree=7)

I see myself as:

- Extraverted, enthusiastic
- Critical, quarrelsome
- Dependable, self-disciplined
- Anxious, easily upset
- Open to new experiences, complex
- Reserved, quiet
- Sympathetic, warm
- Disorganized, careless
- Calm, emotionally stable
- Conventional, uncreative

A.2 Expectations and Trust in Automated Vehicles

A.2.1 Expectation Questionnaire

Please answer the following questions. (Low=1; High=7)

- How would you rate your overall expectations regarding the driving of a self-driving car?
- How would you rate your expectations regarding the effectiveness of a self-driving car?
- How would you rate your expectations regarding the safety of a self-driving car?

A.2.2 Perceived Performance Questionnaire

Please answer the following questions. (Low=1; High=7)

- How would you rate the driving of the just shown self-driving car?
- How would you rate the effectiveness of the just shown self-driving car?
- How would you rate the safety of the just shown self-driving car?

A.2.3 Trust in AVs Questionnaire

Please select the number for each question which best describes your feeling or your impression. (Strongly disagree=1; Strongly agree=7)

- The just shown self-driving car was deceptive.
- The just shown self-driving car behaved in an underhanded manner.
- I was suspicious of the just shown self-driving car's intent, action, or output.
- I worried that the just shown self-driving car's action will have a harmful or injurious outcome.
- I was confident in the just shown self-driving car.
- The just shown self-driving car provided security.
- The just shown self-driving car had integrity.
- The just shown self-driving car was dependable.

- The just shown self-driving car was reliable.
- I can trust the just shown self-driving car.
- I am familiar with the just shown self-driving car.

A.2.4 Risk Conditions Questionnaires

Weather Condition: Do you agree or disagree that each of the following sentences describes the just shown video? Please select one answer for each statement. (Strongly disagree=1; Strongly agree=7)

- The weather made the driving situation risky.
- Due to the weather conditions the likelihood of a collision was high.
- There was a high chance of an accident occurring because of the weather.
- Due to the weather conditions the driving situation was unpredictable.

AV Driving Behavior: Do you agree or disagree that each of following word or phases describe the just shown self-driving car? Please select one answer for each statement. (Strongly disagree=1; Strongly agree=5)

- Dangerous
- Safe
- Hazardous
- Risky
- Could get hurt easily
- Unsafe
- Chance of death
- Scary

A.3 AV Voice and Cognitive and Affective Trusts

A.3.1 Cognitive Trust Questionnaire

Do you agree or disagree with each of the following sentences describes this self-driving car? Please select one answer for each statement. (Strongly disagree=1; Strongly agree=7)

- This self-driving car demonstrates expertise.
- Given the driving behavior of this self-driving car, I see no reason to doubt its competence.
- I can rely on this self-driving car's driving capabilities.
- I believe that other drivers who use this self-driving car will regard it as trustworthy.
- Most people will trust and rely on this self-driving car, even those who don't know much about it.
- This self-driving car's analysis of driving situations was accurate.
- I can trust this self-driving car's instructions.

A.3.2 Affective Trust Questionnaire

Do you agree or disagree with each of the following sentences describes this self-driving car? Please select one answer for each statement. (Strongly disagree=1; Strongly agree=7)

- I can freely share my concerns with this self-driving car.
- This self-driving car will listen to me when I share my struggles in understanding its actions.
- If this self-driving car was no longer available to me, I would feel a sense of loss.
- This self-driving car would respond caringly if I shared my concerns with it about driving.
- I would make considerable emotional investments in this self-driving car.
- This self-driving car will help me with great care.
- This self-driving car will kindly help me when I need it.
- This self-driving car will take care of me with thoughtful consideration.

A.3.3 Perceived AV Role Questionnaire

Please adjust the slider to represent your perceived role of a self-driving car.

- My driving assistant
- My driving supervisor

APPENDIX B

Scenario Description - Chapter 4

B.1 Scenario: Severe Accident Ahead

- Description: After identifying the severe road hazard ahead, the automated vehicle rerouted.
- Explanation: "Severe accident ahead, Rerouting."
- Video Example: https://www.youtube.com /watch?v=-_A67F4p6a8&t=1s

B.2 Scenario: Abrupt Stop Ahead

- Description: The automated vehicle changed lanes due to a roadway obstruction.
- Explanation: "Roadway obstruction, Changing lanes."
- Video Example: https://www.youtube.com /watch?v=OzdIeb-Stng

B.3 Scenario: Oversized Vehicle Ahead

- Description: As an oversized vehicle blocked the roadway ahead, the automated vehicle slowed down until the vehicle turned at the intersection.
- Explanation: "Oversized vehicle blocking the roadway, Slowing down."
- Video Example: https://www.youtube.com /watch?v=66lcgGRpz3s

B.4 Scenario: Emergency Vehicle Approaching from the Rear

- Description: The automated vehicle yielded when an emergency vehicle approaching from behind activated its siren.
- Explanation: ”Emergency vehicle approaching from the rear, Yielding.”
- Video Example: https://www.youtube.com /watch?v=YQLA1k9uj1M

B.5 Scenario: Emergency Vehicle Approaching from the Front

- Description: The automated vehicle stopped when an emergency vehicle approaching from the front activated its siren.
- Explanation: ”Emergency vehicle approaching from the front, Stopping.”
- Video Example: https://www.youtube.com /watch?v=CvqBO66WowY

B.6 Scenario: Road Obstruction

- Description: The automated vehicle rerouted in view of road construction ahead.
- Explanation: ”Identified road obstruction, Rerouting.”
- Video Example: https://www.youtube.com /watch?v=tRYr_xvUYNA&t=1s

BIBLIOGRAPHY

[1] Hillary Abraham, Chaiwoo Lee, Samantha Brady, Craig Fitzgerald, Bruce Mehler, Bryan Reimer, and Joseph F Coughlin. Autonomous vehicles, trust, and driving alternatives: A survey of consumer preferences. *Massachusetts Inst. Technol, AgeLab, Cambridge*, 1(16):2018–12, 2016.

[2] Ritu Agarwal and Jayesh Prasad. Are individual differences germane to the acceptance of new information technologies? *Decision sciences*, 30(2):361–391, 1999. Publisher: Wiley Online Library.

[3] Gordon W Allport and Henry S Odbert. Trait-names: A psycho-lexical study. *Psychological monographs*, 47(1):i, 1936. Publisher: Psychological Review Company.

[4] Hebert Azevedo-Sa, Huajing Zhao, Connor Esterwood, X Jessie Yang, Dawn M Tilbury, and Lionel P Robert Jr. How internal and external risks affect the relationships between trust and driver behavior in automated driving systems. *Transportation research part C: emerging technologies*, 123:102973, 2021. Publisher: Elsevier.

[5] Richard P Bagozzi, Youjae Yi, and Lynn W Phillips. Assessing construct validity in organizational research. *Administrative science quarterly*, pages 421–458, 1991. Publisher: JSTOR.

[6] Bernard Barber. The logic and limits of trust. 1983.

[7] JE Barbuto and Gregory T Gifford. Motivation and leader-member exchange: Evidence counter to similarity attraction theory. *International Journal of Leadership Studies*, 7(1):18–28, 2012.

[8] Sven Beiker. Legal Aspects of Autonomous Driving. *Santa Clara Law Review*, 52(4):1145, December 2012.

[9] Anol Bhattacherjee. Understanding information systems continuance: An expectation-confirmation model. *MIS quarterly*, pages 351–370, 2001. Publisher: JSTOR.

[10] Anol Bhattacherjee and G Premkumar. Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. *MIS quarterly*, pages 229–254, 2004. Publisher: JSTOR.

[11] Taylor C Boas, Dino P Christenson, and David M Glick. Recruiting large online samples in the United States and India: Facebook, mechanical turk, and qualtrics.

Political Science Research and Methods, 8(2):232–250, 2020. Publisher: Cambridge University Press.

[12] Mads Borup, Nik Brown, Kornelia Konrad, and Harro Van Lente. The sociology of expectations in science and technology. *Technology analysis & strategic management*, 18(3-4):285–298, 2006. Publisher: Taylor & Francis.

[13] George EP Box and Norman R Draper. *Empirical model-building and response surfaces.* John Wiley & Sons, 1987.

[14] Cheryl Campanella Bracken and Matthew Lombard. Social Presence and Children: Praise, Intrinsic Motivation, and Learning with Computers. *Journal of Communication*, 54(1):22–37, March 2004.

[15] Inge K Broverman, Susan Raymond Vogel, Donald M Broverman, Frank E Clarkson, and Paul S Rosenkrantz. Sex-Role Stereotypes: A Current Appraisal 1. *Journal of Social issues*, 28(2):59–78, 1972. Publisher: Wiley Online Library.

[16] Austin Brown, Jeffrey Gonder, and Brittany Repac. An analysis of possible energy impacts of automated vehicles. *Road vehicle automation*, pages 137–153, 2014. Publisher: Springer.

[17] Susan A Brown, Viswanath Venkatesh, and Sandeep Goyal. Expectation confirmation in technology use. *Information Systems Research*, 23(2):474–487, 2012. Publisher: INFORMS.

[18] Susan A Brown, Viswanath Venkatesh, and Sandeep Goyal. Expectation confirmation in information systems research. *MIS quarterly*, 38(3):729–A9, 2014. Publisher: JSTOR.

[19] Donn Erwin Byrne. *The attraction paradigm*, volume 462. Academic press, 1971.

[20] Ernest R Cadotte, Robert B Woodruff, and Roger L Jenkins. Expectations and norms in models of consumer satisfaction. *Journal of marketing Research*, 24(3):305–314, 1987. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[21] Sanya Carley, Rachel M Krause, Bradley W Lane, and John D Graham. Intent to purchase a plug-in electric vehicle: A survey of early impressions in large US cites. *Transportation Research Part D: Transport and Environment*, 18:39–45, 2013. Publisher: Elsevier.

[22] Hsin Hsin Chang and Su Wen Chen. The impact of online store environment cues on purchase intention: Trust and perceived risk as a mediator. *Online information review*, 32(6):818–841, 2008. Publisher: Emerald Group Publishing Limited.

[23] Ching Seng Yap and John Wee Huu Hii. Factors Affecting the Adoption of Mobile Commerce in Malaysia. *IUP Journal of Information Technology*, 5(3):24–37, September 2009.

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

[25] Ben Clark, Graham Parkhurst, and Miriam Ricci. Understanding the socioeconomic adoption scenarios for autonomous vehicles: A literature review. 2016.

[26] Sheldon Cohen. Perceived stress in a probability sample of the United States. In *The social psychology of health*, The Claremont Symposium on Applied Social Psychology, pages 31–67. Sage Publications, Inc, Thousand Oaks, CA, US, 1988.

[27] SAE On-Road Automated Vehicle Standards Committee. J3016C: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International, 2021.

[28] C Cooper. *Individual differences*, volume 2 of *null*. 2002.

[29] Brian J Corbitt, Theerasak Thanasankit, and Han Yi. Trust and e-commerce: a study of consumer perceptions. *Electronic commerce research and applications*, 2(3):203–215, 2003. Publisher: Elsevier.

[30] Lee J Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334, 1951. Publisher: Springer.

[31] Andreea Danielescu. Eschewing gender stereotypes in voice assistants to promote inclusion. pages 1–3, 2020.

[32] Fred D Davis, Richard P Bagozzi, and Paul R Warshaw. User acceptance of computer technology: A comparison of two theoretical models. *Management science*, 35(8):982–1003, 1989. Publisher: INFORMS.

[33] John M Digman. Higher-order factors of the Big Five. *Journal of personality and social psychology*, 73(6):1246, 1997. Publisher: American Psychological Association.

[34] Tamara Dinev and Paul Hart. An extended privacy calculus model for e-commerce transactions. *Information systems research*, 17(1):61–80, 2006. Publisher: Informs.

[35] Jiayuan Dong, Emily Lawson, Jack Olsen, and Myounghoon Jeon. Female voice agents in fully autonomous vehicles are not only more likeable and comfortable, but also more competent. volume 64, pages 1033–1037. SAGE Publications Sage CA: Los Angeles, CA, 2020. Issue: 1.

[36] Her-Sen Doong and Hsiangchu Lai. Exploring usage continuance of e-negotiation systems: expectation and disconfirmation approach. *Group Decision and Negotiation*, 17(2):111–126, 2008. Publisher: Springer.

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

[38] Na Du, Feng Zhou, Dawn Tilbury, Lionel Peter Robert, and X. Jessie Yang. Designing Alert Systems in Takeover Transitions: The Effects of Display Information and Modality. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 173–180, Leeds United Kingdom, September 2021. ACM.

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

[40] Alice H Eagly and Steven J Karau. Role congruity theory of prejudice toward female leaders. *Psychological review*, 109(3):573, 2002. Publisher: American Psychological Association.

[41] Alice H Eagly and Valerie J Steffen. Gender stereotypes stem from the distribution of women and men into social roles. *Journal of personality and social psychology*, 46(4):735, 1984. Publisher: American Psychological Association.

[42] Chad Edwards, Autumn Edwards, Brett Stoll, Xialing Lin, and Noelle Massey. Evaluations of an artificial intelligence instructor's voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior*, 90:357–362, 2019. Publisher: Elsevier.

[43] Chad Edwards and Jake Harwood. Social Identity in the Classroom: An Examination of Age Identification Between Students and Instructors. *Communication Education*, 52(1):60–65, January 2003. Publisher: Routledge _eprint: https://doi.org/10.1080/03634520302463.

[44] Jeffrey R Edwards. *Person-job fit: A conceptual integration, literature review, and methodological critique.* John Wiley & Sons, 1991.

[45] Jeffrey R Edwards and Mark E Parry. On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management journal*, 36(6):1577–1613, 1993. Publisher: Academy of Management Briarcliff Manor, NY 10510.

[46] Jeffrey R Edwards and R Van Harrison. Job demands and worker health: three-dimensional reexamination of the relationship between person-environment fit and strain. *Journal of Applied Psychology*, 78(4):628, 1993. Publisher: American Psychological Association.

[47] Julia B Edwards. Speed adjustment of motorway commuter traffic to inclement weather. *Transportation research part F: traffic psychology and behaviour*, 2(1):1–14, 1999. Publisher: Elsevier.

[48] José Manuel Ortega Egea and María Victoria Román González. Explaining physicians' acceptance of EHCR systems: An extension of TAM with trust and risk factors. *Computers in Human Behavior*, 27(1):319–332, 2011. Publisher: Elsevier.

[49] Connor Esterwood, X Jessie Yang, and Lionel P Robert. Barriers to AV bus acceptance: A national survey and research agenda. *International Journal of Human–Computer Interaction*, 37(15):1391–1403, 2021. Publisher: Taylor & Francis.

[50] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, 132:138–161, 2019. Publisher: Elsevier.

[51] Ewan Ferlie, John Gabbay, Louise Fitzgerald, Louise Locock, and Sue Dopson. Evidence-based medicine and organisational change: an overview of some recent qualitative research. 2001. Publisher: Palgrave Macmillan.

[52] Claes Fornell and David F Larcker. Structural equation models with unobservable variables and measurement error: Algebra and statistics. 1981. Publisher: Sage Publications Sage CA: Los Angeles, CA.

[53] Yannick Forster, Frederik Naujoks, and Alexandra Neukum. Increasing anthropomorphism and trust in automated driving functions by adding speech output. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 365–372, June 2017.

[54] Jesse Fox and Andrew Gambino. Relationship development with humanoid social robots: Applying interpersonal theories to human–robot interaction. *Cyberpsychology, Behavior, and Social Networking*, 24(5):294–299, 2021. Publisher: Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New . . . .

[55] Anna-Katharina Frison, Laura Aigner, Philipp Wintersberger, and Andreas Riener. Who is generation A? Investigating the experience of automated driving for different age groups. pages 94–104, 2018.

[56] Andrew Gambino, Jesse Fox, and Rabindra A Ratan. Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, 1:71–85, 2020. Publisher: Communication and Social Robotics Labs Kalamazoo, Michigan.

[57] Stefan K Gehrig and Fridtjof J Stein. Dead reckoning and cartography using stereo vision for an autonomous car. volume 3, pages 1507–1512. IEEE, 1999.

[58] Christian Gold, Moritz Körber, Christoph Hohenberger, David Lechner, and Klaus Bengler. Trust in automation–before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing*, 3:3025–3032, 2015. Publisher: Elsevier.

[59] Caren B Goldberg. Relational demography and similarity-attraction in interview assessments and subsequent offer decisions: Are we missing something? *Group & Organization Management*, 30(6):597–624, 2005. Publisher: Sage Publications Sage CA: Thousand Oaks, CA.

[60] Peter R. Grant. Reactions to intergroup similarity: Examination of the similarity-differentiation and the similarity-attraction hypotheses. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 25:28–44, 1993. Place: Canada Publisher: Canadian Psychological Association.

[61] Taehyun Ha, Sangyeon Kim, Donghak Seo, and Sangwon Lee. Effects of explanation types and perceived risk on trust in autonomous vehicles. *Transportation research part F: traffic psychology and behaviour*, 73:271–280, 2020. Publisher: Elsevier.

[62] Florian Habler, Valentin Schwind, and Niels Henze. Effects of Smart Virtual Assistants' Gender and Language. In *Proceedings of Mensch und Computer 2019*, pages 469–473. 2019.

[63] Allison W Harrison and R Kelly Rainer Jr. The influence of individual differences on skill in end-user computing. *Journal of Management Information Systems*, 9(1):93–111, 1992. Publisher: Taylor & Francis.

[64] Nathan Andrew Hatfield. The Effects of Automation Transparency and Ethical Outcomes on User Trust and Blame Towards Fully Autonomous Vehicles. 2018.

[65] Nathalie Hauk, Joachim Hüffmeier, and Stefan Krumm. Ready to be a Silver Surfer? A Meta-analysis on the Relationship Between Chronological Age and Technology Acceptance. *Computers in Human Behavior*, 84:304–319, July 2018.

[66] Bob E Hayes, Jill Perander, Tara Smecko, and Jennifer Trask. Measuring perceptions of workplace safety: Development and validation of the work safety scale. *Journal of Safety research*, 29(3):145–161, 1998. Publisher: Elsevier.

[67] Fan He and Catherine M. Burns. A Battle of Voices: A Study of the Relationship between Driving Experience, Driving Style, and In-Vehicle Voice Assistant Character. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 236–242, Seoul Republic of Korea, September 2022. ACM.

[68] Madeline E Heilman. Sex bias in work settings: The lack of fit model. *Research in organizational behavior*, 1983. Publisher: JAI Press, Inc.

[69] Madeline E Heilman. Gender stereotypes and workplace bias. *Research in organizational Behavior*, 32:113–135, 2012. Publisher: Elsevier.

[70] Madeline E Heilman, Caryn J Block, and Richard F Martell. Sex stereotypes: Do they influence perceptions of managers? *Journal of Social behavior and Personality*, 10(4):237, 1995. Publisher: Select Press.

[71] Kevin Anthony Hoff and Masooda Bashir. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3):407–434, 2015. Publisher: Sage Publications Sage CA: Los Angeles, CA.

[72] Mohammad Alamgir Hossain and Mohammed Quaddus. Expectation–confirmation theory in information system research: A review and analysis. *Information systems theory*, pages 441–469, 2012. Publisher: Springer.

[73] Daniel Howard and Danielle Dai. Public Perceptions of Self-Driving Cars: The Case of Berkeley, California. 2014. Number: 14-4502.

[74] Mary Lee Hummert, Teri A. Garstka, Jaye L. Shaner, and Sharon Strahm. Stereotypes of the Elderly Held by Young, Middle-Aged, and Elderly Adults. *Journal of Gerontology*, 49(5):P240–P249, September 1994.

[75] Jon ; Carter Jasperson, Pamela E. A Comprehensive Conceptualization of Post-Adoptive Behaviors Associated with Information Technology Enabled Work Systems. *MIS quarterly*, 29(3). Publisher: Management Information Systems Research Center, University of Minnesota.

[76] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics*, 4(1):53–71, 2000. Publisher: Taylor & Francis.

[77] James J Jiang and Gary Klein. Expectation-confirmation theory: Capitalizing on descriptive power. In *Handbook of research on contemporary theoretical models in information systems*, pages 384–401. IGI Global, 2009.

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

[79] William J Kettinger and Choong C Lee. Zones of tolerance: Alternative scales for measuring information systems service quality. *MIS quarterly*, pages 607–623, 2005. Publisher: JSTOR.

[80] Markku ; Summala Kilpeläinen, Heikki. Effects of weather and weather forecasts on driver behaviour. *Transportation research. Part F, Traffic psychology and behaviour*, 10(4). Publisher: Elsevier India Pvt Ltd.

[81] Dan J Kim, Donald L Ferrin, and H Raghav Rao. A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems*, 44(2):544–564, 2008. Publisher: Elsevier.

[82] Joseph K Kim, Crystal M Harold, and Brian C Holtz. Evaluations of abusive supervisors: The moderating role of the abuser's gender. *Journal of Organizational Behavior*, 43(3):465–482, 2022. Publisher: Wiley Online Library.

[83] Amanda J Koch, Susan D D'Mello, and Paul R Sackett. A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of applied psychology*, 100(1):128, 2015. Publisher: American Psychological Association.

[84] Anne M Koenig. Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in psychology*, 9:1086, 2018. Publisher: Frontiers Media SA.

[85] Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 9(4):269–275, November 2015.

[86] Jeamin Koo, Dongjun Shin, Martin Steinert, and Larry Leifer. Understanding driver responses to voice alerts of autonomous car operations. *International Journal of Vehicle Design*, 70(4):377–392, January 2016. Publisher: Inderscience Publishers.

[87] Praveen K Kopalle and Donald R Lehmann. Strategic management of expectations: The role of disconfirmation sensitivity and perfectionism. *Journal of Marketing Research*, 38(3):386–394, 2001. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[88] Sahil Koul and Ali Eydgahi. Utilizing technology acceptance model (TAM) for driverless car technology adoption. *Journal of technology management & innovation*, 13(4):37–46, 2018. Publisher: SciELO Chile.

[89] Carol T Kulik, Greg R Oldham, and J Richard Hackman. Work design as an approach to person-environment fit. *Journal of vocational behavior*, 31(3):278–296, 1987. Publisher: Elsevier.

[90] Ko Kuwabara, Jiyin Cao, Soomin Sophie Cho, and Paul Ingram. Lay Theories of Instrumental Relations: Explaining Individual Differences in Dispositional Similarity-Attraction. *Academy of Management Journal*, February 2022. Publisher: Academy of Management.

[91] Miltos Kyriakidis, Riender Happee, and Joost CF de Winter. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation research part F: traffic psychology and behaviour*, 32:127–140, 2015. Publisher: Elsevier.

[92] Moritz Körber, Lorenz Prasch, and Klaus Bengler. Why Do I Have to Drive Now? Post Hoc Explanations of Takeover Requests. *Human Factors*, 60(3):305–323, May 2018. Publisher: SAGE Publications Inc.

[93] Karl Reiner Lang and Sirkka Jarvenpaa. Managing the paradoxes of mobile technology. *Information systems management*, 22(4):7–23, 2005.

[94] Nancy Lankton, D Harrison McKnight, and Jason Bennett Thatcher. Incorporating trust-in-technology into Expectation Disconfirmation Theory. *The Journal of Strategic Information Systems*, 23(2):128–145, 2014. Publisher: Elsevier.

[95] Nancy K Lankton and Harrison D McKnight. Examining two expectation disconfirmation theory models: Assimilation and asymmetry effects. *Journal of the Association for Information Systems*, 13(2):1, 2012.

[96] S. Laumer, A Eckhardt, Y. Dwivedi, M. Wade, and S. Schneberger. *Information systems theory*, volume null of *null*. 2012.

[97] Bruce D Layton and Chester A Insko. Anticipated interaction and the similarity-attraction effect. *Sociometry*, pages 149–162, 1974. Publisher: JSTOR.

[98] Chaiwoo Lee, Bruce Mehler, Bryan Reimer, and Joseph F Coughlin. User perceptions toward in-vehicle technologies: Relationships to age, health, preconceptions, and hands-on experience. *International Journal of Human-Computer Interaction*, 31(10):667–681, 2015. Publisher: Taylor & Francis.

[99] Eun Ju Lee, Clifford Nass, and Scott Brave. Can computer-generated speech have gender? an experimental test of gender stereotype. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '00, pages 289–290, New York, NY, USA, April 2000. Association for Computing Machinery.

[100] Hyun-Joo Lee, Hyeon Jeong Cho, Wenwen Xu, and Ann Fairhurst. The influence of consumer traits and demographics on intention to use retail self-service checkouts. *Marketing Intelligence & Planning*, 28(1):46–58, 2010. Publisher: Emerald Group Publishing Limited.

[101] Jae-gil Lee and Kwan Min Lee. Polite speech strategies and their impact on drivers' trust in autonomous vehicles. *Computers in Human Behavior*, 127:107015, February 2022.

[102] John Lee and Neville Moray. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10):1243–1270, 1992. Publisher: Taylor & Francis.

[103] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1):50–80, 2004. Publisher: SAGE Publications Sage UK: London, England.

[104] Jong-Eun Roselyn Lee and Clifford I Nass. Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication. In *Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives*, pages 1–15. IGI Global, 2010.

[105] Kwan Min Lee, Katharine Liao, and Seoungho Ryu. Children's Responses to Computer-Synthesized Speech in Educational Media: Gender Consistency and Gender Similarity Effects. *Human Communication Research*, 33(3):310–329, July 2007.

[106] Sanguk Lee, Rabindra Ratan, and Taiwoo Park. The voice makes the car: Enhancing autonomous vehicle perceptions and adoption intention through voice agent gender and style. *Multimodal Technologies and Interaction*, 3(1):20, 2019. Publisher: MDPI.

[107] Roy J Lewicki and Barbara B Bunker. Developing and maintaining trust in work relationships. *Trust in organizations: Frontiers of theory and research*, 114:139, 1996.

[108] Roy J. Lewicki, Edward C. Tomlinson, and Nicole Gillespie. Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions. *Journal of Management*, 32:991–1022, 2006. Place: US Publisher: Sage Publications.

[109] J. David Lewis and Andrew Weigert. Trust as a Social Reality. *Social Forces*, 63(4):967–985, June 1985.

[110] Mengyao Li, Brittany E Holthausen, Rachel E Stuck, and Bruce N Walker. No risk no trust: Investigating perceived risk in highly automated driving. pages 177–185, 2019.

[111] Paul Liernert and Maria Caspani. Americans still don't trust self-driving cars, Reuters/Ipsos poll finds | Reuters, April 2019.

[112] Peng Liu, Run Yang, and Zhigang Xu. Public acceptance of fully automated driving: Effects of social trust and risk/benefit perceptions. *Risk Analysis*, 39(2):326–341, 2019. Publisher: Wiley Online Library.

[113] Nora Ni Loideain and Rachel Adams. From Alexa to Siri and the GDPR: the gendering of virtual personal assistants and the role of data protection impact assessments. *Computer Law & Security Review*, 36:105366, 2020. Publisher: Elsevier.

[114] Matthew Lombard and Kun Xu. Social responses to media technologies in the 21st century: The media are social actors paradigm. *Human-Machine Communication*, 2:29–55, January 2021. Publisher: Communication and Social Robotics Labs.

[115] Sydney Lynch and Marianne Campbell. Adolescents Voice Preference in Auditory Advertisements: A Study in Gender Stereotypes and Multi-Media Marketing. *Journal of Student Research*, 10(1), 2021.

[116] Karen S Lyness and Madeline E Heilman. When fit is fundamental: performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91(4):777, 2006. Publisher: American Psychological Association.

[117] Qi Ma, Alan HS Chan, and Pei-Lee Teh. Insights into older adults' technology acceptance through meta-analysis. *International Journal of Human–Computer Interaction*, 37(11):1049–1062, 2021. Publisher: Taylor & Francis.

[118] Ruth Madigan, Tyron Louw, Marc Wilbrink, Anna Schieben, and Natasha Merat. What influences the decision to use automated public transport? Using UTAUT to understand public acceptance of automated road transport systems. *Transportation research part F: traffic psychology and behaviour*, 50:55–64, 2017. Publisher: Elsevier.

[119] Likoebe M Maruping, Hillol Bala, Viswanath Venkatesh, and Susan A Brown. Going beyond intention: Integrating behavioral expectation into the unified theory of acceptance and use of technology. *Journal of the Association for Information Science and Technology*, 68(3):623–637, 2017. Publisher: Wiley Online Library.

[120] Roger C Mayer, James H Davis, and F David Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995. Publisher: Academy of Management Briarcliff Manor, NY 10510.

[121] Daniel J. McAllister. Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal*, 38(1):24–59, February 1995. Publisher: Academy of Management.

[122] R. R. McCrae, P. T. Costa, R. Hogan, J. Johnson, and S. Briggs. *Handbook of personality psychology*, volume null of *null*. 1997.

[123] Barbara A. Mellers, Alan Schwartz, Katty Ho, and Ilana Ritov. Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options. *Psychological Science*, 8(6):423–429, 1997. Publisher: [Association for Psychological Science, Sage Publications, Inc.].

[124] Matthew L Meuter, Mary Jo Bitner, Amy L Ostrom, and Stephen W Brown. Choosing among alternative service delivery modes: An investigation of customer trial of self-service technologies. *Journal of marketing*, 69(2):61–83, 2005. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[125] Jonas Meyer, Henrik Becker, Patrick M Bösch, and Kay W Axhausen. Autonomous vehicles: The next jump in accessibilities? *Research in transportation economics*, 62:80–91, 2017. Publisher: Elsevier.

[126] Jaroslav Michalco, Jakob Grue Simonsen, and Kasper Hornbæk. An exploration of the relation between expectations and user experience. *International Journal of Human-Computer Interaction*, 31(9):603–617, 2015. Publisher: Taylor & Francis.

[127] Donald E. ; Johnson Miles, Gregory L. Aggressive driving behaviors: are there psychological and attitudinal predictors? *Transportation research. Part F, Traffic psychology and behaviour*, 6(2). Publisher: Elsevier India Pvt Ltd.

[128] Vincent-Wayne Mitchell. Consumer perceived risk: conceptualisations and models. *European Journal of marketing*, 1999. Publisher: MCB UP Ltd.

[129] R Matthew Montoya and Robert S Horton. A meta-analytic investigation of the processes underlying the similarity-attraction effect. *Journal of Social and Personal Relationships*, 30(1):64–94, 2013. Publisher: Sage Publications Sage UK: London, England.

[130] Michael G. Morris and Viswanath Venkatesh. Age Differences in Technology Adoption Decisions: Implications for a Changing Work Force. *Personnel Psychology*, 53(2):375–403, 2000. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1744-6570.2000.tb00206.x.

[131] Bonnie M Muir. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11):1905–1922, 1994. Publisher: Taylor & Francis.

[132] Guido Möllering. The nature of trust: From Georg Simmel to a theory of expectation, interpretation and suspension. *Sociology*, 35(2):403–420, 2001. Publisher: Cambridge University Press.

[133] Procheta Nag and Özge Nilay Yalçın. Gender stereotypes in virtual agents. pages 1–8, 2020.

[134] Clifford Nass and Li Gong. Speech interfaces from an evolutionary perspective. *Communications of the ACM*, 43(9):36–43, 2000. Publisher: ACM New York, NY, USA.

[135] Clifford Nass, Ing-Marie Jonsson, Helen Harris, Ben Reaves, Jack Endo, Scott Brave, and Leila Takayama. Improving automotive safety by pairing driver emotion and car voice emotion. pages 1973–1976, 2005.

[136] Clifford Nass, Youngme Moon, and Paul Carney. Are People Polite to Computers? Responses to Computer-Based Interviewing Systems1. *Journal of Applied Social Psychology*, 29(5):1093–1109, 1999. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1559-1816.1999.tb00142.x.

[137] Clifford Nass, Youngme Moon, and Nancy Green. Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. *Journal of Applied Social Psychology*, 27(10):864–876, 1997. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1559-1816.1997.tb00275.x.

[138] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. Computers are social actors. pages 72–78, 1994.

[139] Clifford I Nass, Youngme Moon, and John Morkes. Computers are social actors: A review of current. *Human values and the design of computer technology*, 72:137, 1997.

[140] Clifford Ivar Nass and Scott Brave. *Wired for speech: How voice activates and advances the human-computer relationship.* MIT press Cambridge, 2005.

[141] Ilja Nastjuk, Bernd Herrenkind, Mauricio Marrone, Alfred Benedikt Brendel, and Lutz M. Kolbe. What drives the acceptance of autonomous driving? An investigation of acceptance factors from an end-user's perspective. *Technological Forecasting and Social Change*, 161:120319, December 2020.

[142] Frederik Naujoks, Yannick Forster, Katharina Wiedemann, and Alexandra Neukum. A Human-Machine Interface for Cooperative Highly Automated Driving. In Neville A. Stanton, Steven Landry, Giuseppe Di Bucchianico, and Andrea Vallicelli, editors, *Advances in Human Aspects of Transportation*, pages 585–595, Cham, 2017. Springer International Publishing.

[143] Frederik Naujoks, Yannick Forster, Katharina Wiedemann, and Alexandra Neukum. Improving usefulness of automated driving by lowering primary task interference through HMI design. *Journal of Advanced Transportation*, 2017, 2017. Publisher: Hindawi.

[144] RG Netemeyer, WO Bearden, and S Sharma. Scaling Procedures: Issues and ApplicationsSage Publications. *Thousands Oak, CA*, 2003.

[145] Sina Nordhoff, Joost De Winter, Miltos Kyriakidis, Bart Van Arem, and Riender Happee. Acceptance of driverless vehicles: Results from a large cross-national questionnaire study. *Journal of Advanced Transportation*, 2018, 2018. Publisher: Hindawi.

[146] John Han Numan. Knowledge-based systems as companions: Trust, human computer interaction and complex systems. 1998.

[147] Luis Oliveira, Christopher Burns, Jacob Luton, Sumeet Iyer, and Stewart Birrell. The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour*, 72:280–296, 2020. Publisher: Elsevier.

[148] Richard L Oliver. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, 17(4):460–469, 1980. Publisher: Sage Publications Sage CA: Los Angeles, CA.

[149] Thomas A O'Neill and Natalie J Allen. Personality and the prediction of team performance. *European Journal of Personality*, 25(1):31–42, 2011. Publisher: SAGE Publications Sage UK: London, England.

[150] Christopher Orpen. Attitude Similarity, Attraction, and Decision-Making in the Employment Interview. *The Journal of Psychology*, 117(1):111–120, May 1984. Publisher: Routledge _eprint: https://doi.org/10.1080/00223980.1984.9923666.

[151] Paul A. Pavlou. Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model, 2003.

[152] Miranda AG Peeters, Harrie FJM Van Tuijl, Christel G Rutte, and Isabelle MMJ Reymen. Personality and team performance: a meta-analysis. *European journal of personality*, 20(5):377–396, 2006. Publisher: SAGE Publications Sage UK: London, England.

[153] Hazel Pettifor, Charlesette Wilson, John Axsen, Wokje Abrahamse, and J Anable. Social influence in the global diffusion of alternative fuel vehicles–A meta-analysis. *Journal of Transport Geography*, 62:247–261, 2017. Publisher: Elsevier.

[154] Diane M Phillips and Hans Baumgartner. The role of consumption emotions in the satisfaction response. *Journal of Consumer psychology*, 12(3):243–252, 2002. Publisher: Elsevier.

[155] Lingyun Qiu and Izak Benbasat. A study of demographic embodiments of product recommendation agents in electronic commerce. *International Journal of Human-Computer Studies*, 68(10):669–688, 2010. Publisher: Elsevier.

[156] John K Rempel, John G Holmes, and Mark P Zanna. Trust in close relationships. *Journal of personality and social psychology*, 49(1):95, 1985. Publisher: American Psychological Association.

[157] Yuna Ro and Youngwook Ha. A factor analysis of consumer expectations for autonomous cars. *Journal of Computer Information Systems*, 59(1):52–60, 2019. Publisher: Taylor & Francis.

[158] L. P Robert. *AI & SOCIETY*, 34(3):687, 2019.

[159] Lionel P Robert. Are automated vehicles safer than manually driven cars? *AI & SOCIETY*, 34(3):687–688, 2019. Publisher: Springer.

[160] Lionel P. Robert and Tracy Ann Sykes. Extending the Concept of Control Beliefs: Integrating the Role of Advice Networks. *Information Systems Research*, 28(1):84–96, March 2017. Publisher: INFORMS.

[161] Lionel P Robert Jr, Rasha Alahmad, Connor Esterwood, Sangmi Kim, Sangseok You, and Qiaoning Zhang. A review of personality in human–robot interactions. *Foundations and Trends® in Information Systems*, 4(2):107–212, 2020. Publisher: Now Publishers, Inc.

[162] Sandra L Robinson. Trust and breach of the psychological contract. *Administrative science quarterly*, pages 574–599, 1996. Publisher: JSTOR.

[163] Everett M Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.

[164] Denise M Rousseau, Sim B Sitkin, Ronald S Burt, and Colin Camerer. Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3):393–404, 1998. Publisher: Academy of Management Briarcliff Manor, NY 10510.

[165] Peter A. M. Ruijten, Jacques M. B. Terken, and Sanjeev N. Chandramouli. Enhancing Trust in Autonomous Vehicles through Intelligent User Interfaces That Mimic Human Behavior. *Multimodal Technologies and Interaction*, 2(4):62, December 2018. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[166] Christina Rödel, Susanne Stadler, Alexander Meschtscherjakov, and Manfred Tscheligi. Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. pages 1–8, 2014.

[167] Subhro Sarkar and Arpita Khare. Influence of expectation confirmation, network externalities, and flow on use of mobile shopping apps. *International Journal of Human–Computer Interaction*, 35(16):1449–1460, 2019. Publisher: Taylor & Francis.

[168] Brandon Schoettle and Michael Sivak. A survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia. Technical report, University of Michigan, Ann Arbor, Transportation Research Institute, 2014.

[169] Sergej S Shadrin and Anastasiia A Ivanova. ANALYTICAL REVIEW OF STANDARD SAE J3016 TAXONOMY AND DEFINITIONS FOR TERMS RELATED TO DRIVING AUTOMATION SYSTEMS FOR ON-ROAD MOTOR VEHICLES WITH LATEST UPDATES. *Avtomobil'. Doroga. Infrastruktura.*, (3 (21)):10, 2019.

[170] Ravi Shanker, Adam Jonas, Scott Devitt, Katy Huberty, Simon Flannery, William Greene, Benjamin Swinburne, Gregory Locraft, Adam Wood, and Keith Weiss. Autonomous cars: Self-driving the new auto industry paradigm. *Morgan Stanley blue paper*, pages 1–109, 2013. Publisher: Morgan Stanley & Co. LLC.

[171] Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. Psychological roadblocks to the adoption of self-driving vehicles. *Nature Human Behaviour*, 1(10):694–696, 2017. Publisher: Nature Publishing Group UK London.

[172] Yuan Shen, Shanduojiao Jiang, Yanlin Chen, Eileen Yang, Xilun Jin, Yuliang Fan, and Katie Driggs Campbell. To explain or not to explain: A study on the necessity of explanations for autonomous vehicles. *arXiv preprint arXiv:2006.11684*, 2020.

[173] TB Sheridan. Trustworthiness of command and control systems. In *Analysis, Design and Evaluation of Man–Machine Systems 1988*, pages 427–431. Elsevier, 1989.

[174] Ramadhar Singh, Sherie E-Lin Yeo, Patrick K. F. Lin, and Lydia Tan. Multiple Mediators of the Attitude Similarity-Attraction Relationship: Dominance of Inferred Attraction and Subtlety of Affect. *Basic and Applied Social Psychology*, 29(1):61–74, April 2007. Publisher: Routledge _eprint: https://doi.org/10.1080/01973530701331007.

[175] Benjamin K Sovacool, Johannes Kester, Lance Noel, and Gerardo Zarazua de Rubens. Income, political affiliation, urbanism and geography in stated preferences for electric vehicles (EVs) and vehicle-to-grid (V2G) technologies in Northern Europe. *Journal of Transport Geography*, 78:214–229, 2019. Publisher: Elsevier.

[176] D. Sandy Staples, Ian Wong, and Peter B. Seddon. Having expectations of information systems benefits that match received benefits: does it really matter? *Information & Management*, 40(2):115–131, December 2002.

[177] David L Streiner. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, 80(1):99–103, 2003. Publisher: Taylor & Francis.

[178] Araz Taeihagh and Hazel Si Min Lim. Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport reviews*, 39(1):103–128, 2019. Publisher: Taylor & Francis.

[179] Benedict Tay, Younbo Jung, and Taezoon Park. When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, 38:75–84, 2014. Publisher: Elsevier.

[180] James YL Thong, Se-Joon Hong, and Kar Yan Tam. The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. *International Journal of human-computer studies*, 64(9):799–810, 2006. Publisher: Elsevier.

[181] Antonella Toffetti, Ellen S Wilschut, Marieke H Martens, Anna Schieben, Amon Rambaldini, Natasha Merat, and Frank Flemisch. CityMobil: Human factor issues regarding highly automated vehicles on eLane. *Transportation research record*, 2110(1):1–8, 2009. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[182] T. Triplett, R. Santos, S. Rosenbloom, and B Tefft. *American driving survey: 2014–2015*, volume null of *null*. 2016.

[183] David K Tse and Peter C Wilton. Models of consumer satisfaction formation: An extension. *Journal of marketing research*, 25(2):204–212, 1988. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[184] I. P. Tussyadiah, F. J. Zach, and J Wang. *Information and communication technologies in tourism 2017*, volume null of *null*. 2017.

[185] Gregg G Van Ryzin. Expectations, performance, and citizen satisfaction with urban services. *Journal of policy analysis and management*, 23(3):433–448, 2004. Publisher: Wiley Online Library.

[186] Viswanath Venkatesh and Fred D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2):186–204, 2000. ISBN: 0025-1909 Publisher: INFORMS.

[187] Viswanath Venkatesh and Sandeep Goyal. Expectation disconfirmation and technology adoption: polynomial modeling and response surface analysis. *MIS quarterly*, pages 281–303, 2010. Publisher: JSTOR.

[188] Frank M. F. Verberne, Jaap Ham, and Cees J. H. Midden. Trust in Smart Systems: Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(5):799–810, October 2012.

[189] P. F Waller. *Human Factors*, 33(5):499, 1991.

[190] Janelle E Wells and Thomas J Aicher. Follow the leader: A relational demography, similarity attraction, and social identity theory of leadership approach of a team's performance. *Gender Issues*, 30:1–14, 2013. Publisher: Springer.

[191] Christopher G Wetzel and Chester A Insko. The similarity-attraction relationship: Is there an ideal one? *Journal of Experimental Social Psychology*, 18(3):253–276, 1982. Publisher: Elsevier.

[192] Lucas Whittaker, Jan Kietzmann, Kate Letheren, Rory Mulcahy, and Rebekah Russell-Bennett. Brace yourself! Why managers should adopt a synthetic media incident response playbook in an age of falsity and synthetic media. *Business Horizons*, August 2022.

[193] Gesa Wiegand, Matthias Schmidmaier, Thomas Weber, Yuanting Liu, and Heinrich Hussmann. I Drive - You Trust: Explaining Driving Behavior Of Autonomous Cars. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, pages 1–6, New York, NY, USA, May 2019. Association for Computing Machinery.

[194] Gavin Walter Wille. The Stepwise Procedure for the Empirical Assessment of Latent Variable Models. 1996. Publisher: University of Port Elizabeth.

[195] Andrew F Wood and Matthew J Smith. *Online communication: Linking technology, identity, & culture*. Routledge, 2004.

[196] R. S. Woodworth and D. G Marquis. *Psychology (psychology revivals): A study of mental life*, volume null of *null*. 2014.

[197] Kun Xu, Xiaobei Chen, and Luling Huang. Deep mind in social responses to technologies: A new approach to explaining the Computers are Social Actors phenomena. *Computers in Human Behavior*, 134:107321, September 2022.

[198] Qing Yang, Chuan Pang, Liu Liu, David C Yen, and J Michael Tarn. Exploring consumer perceived risk and trust for online payments: An empirical study in China's younger generation. *Computers in human behavior*, 50:9–24, 2015. Publisher: Elsevier.

[199] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. Understanding the effect of accuracy on trust in machine learning models. pages 1–12, 2019.

[200] Akbar Zaheer, Bill McEvily, and Vincenzo Perrone. Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization science*, 9(2):141–159, 1998. Publisher: INFORMS.

[201] Valarie A Zeithaml, Ananthanarayanan Parasuraman, Leonard L Berry, and Leonard L Berry. *Delivering quality service: Balancing customer perceptions and expectations*. Simon and Schuster, 1990.

[202] Qiaoning Zhang, X Jessie Yang, and Lionel P Robert. What and When to Explain? A Survey of the Impact of Explanation on Attitudes Toward Adopting Automated Vehicles. *IEEE Access*, 9:159533–159540, 2021. Publisher: IEEE.

[203] Qiaoning Zhang, Xi Jessie Yang, and Lionel P Robert. Drivers' age and automated vehicle explanations. *Sustainability*, 13(4):1948, 2021. Publisher: Multidisciplinary Digital Publishing Institute.

[204] Tingru Zhang, Da Tao, Xingda Qu, Xiaoyan Zhang, Rui Lin, and Wei Zhang. The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation research part C: emerging technologies*, 98:207–220, 2019. Publisher: Elsevier.

[205] Tingru Zhang, Da Tao, Xingda Qu, Xiaoyan Zhang, Jihong Zeng, Haoyu Zhu, and Han Zhu. Automated vehicle acceptance in China: Social influence and initial trust are key determinants. *Transportation research part C: emerging technologies*, 112:220–233, 2020. Publisher: Elsevier.

[206] Johanna Zmud, Ipek N. Sener, Jason Wagner, and Texas A&M Transportation Institute. Consumer acceptance and travel behavior : impacts of automated vehicles : final report. Technical Report PRC 15-49 F, January 2016.

[207] Lynne G Zucker. Production of trust: Institutional sources of economic structure, 1840–1920. *Research in organizational behavior*, 1986. Publisher: JAI Press, Inc.