World Electric Vehicle Journal

MDPI

*Article*

# Analysis of Cross-Cultural Trust and Vehicle Operation Metrics for Self-Driving Cars †

**Steven Tolbert * and Mehrdad Nojoumian**

Department of Electrical Engineering and Computer Science, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA; mnojoumian@fau.edu

* Correspondence: stolbert2014@fau.edu

† This paper is an extended version of our paper published in the Hawaii International Conference on Human Factors in Design, Engineering, and Computing (AHFE), Honolulu, HI, USA, 20–22 July 2025; Volume 199, pp. 137–146.

**Abstract**

This paper presents an exploratory cross-cultural analysis of autonomous vehicle expectations through a 57-question survey distributed in the United States ($n$ = 50), Germany ($n$ = 66), and Panama ($n$ = 41). Five scales are presented and validated: Driving Behavior Aggressiveness (DBA), Self-Driving Car Aggressiveness (SDCA), Artificial Intelligence (AI) Trust (AIT), AI Driving Mechanics Trust (AIDMT), and Driver Safety Score (DSS). Each scale is validated via confirmatory factor analysis and multi-group measurement invariance testing. Results show that drivers prefer a self-driving car driving style *more conservative than their own*; however, participants who are more trustful of AI show *DBA–SDCA equivalence*, consistent with acceptance of a driving style comparable to their own. Significant cross-cultural differences emerge, with Panama diverging from the United States and Germany on DBA, SDCA, AIDMT, and DSS; these country effects largely persist after controlling for demographics. These findings suggest that self-driving car behaviors should be tailored to regional expectations and passenger trust profiles to improve adoption.

**Keywords:** self-driving cars; driving behavior; cross-cultural expectations; user experience; trust in autonomy

## 1. Introduction

Over the past decade, the technologies powering self-driving cars have advanced rapidly toward commercial deployment. The current standard for levels of driving automation is defined by the Society of Automotive Engineers (SAE) International's J3016 Levels of Driving Automation [1]. The first three levels, 0, 1, and 2, vary from no driving automation to partial driving automation where the human driver monitors the road. Beginning at Level 3, the automated driving system monitors the road, and as the levels progress to 4 and 5, the abilities of the driving system increase, such that at Level 5, the vehicle is fully capable of automation in all scenarios. Mercedes-Benz's "Drive Pilot" system represents a Level 3 automation system that fully handles the driving task within well-defined operational domains at slow to moderate speeds [2]. Level 4 systems have progressed to full commercial deployment, with Waymo operating a driverless robotaxi service across multiple US cities that has demonstrated substantially lower crash rates than human drivers [3]. Level 5 systems, however, remain unrealized due to the number of edge cases that such a system would need to handle. The questions for the immediate future of self-driving car technology then revolve around the adaptability of the technology rather



Academic Editor: Grzegorz Sierpiński

Received: 6 February 2026
Revised: 19 March 2026
Accepted: 19 March 2026
Published: 22 March 2026

than the technology itself. While self-driving car technologies continue to mature, the large-scale adoption of self-driving cars remains uncertain. Current research accordingly points to limited acceptance of the technology among the general public.

In 2021, Morning Consult surveyed 2200 adults in the United States and found that 47% of respondents believed autonomous vehicles were less safe than their human-driven counterparts and that only 22% believed autonomous vehicles were safer than a human driver [4]. This skepticism has persisted: the AAA's 2025 survey found that 6 in 10 US drivers remain afraid of riding in a self-driving car, with only 13% expressing trust in the technology [5]. With a sustained majority of the public expressing skepticism toward the technology even as these systems enter commercial deployment, research on improving consumer trust in self-driving cars is needed [6,7]. Research conducted by Lee et al. suggests that there are several factors affecting the adoption of self-driving cars. These factors include the perceived usefulness, self-efficacy, perceived risk, and psychological ownership of the vehicle [8]. Similarly, Choi and Ji found that trust and perceived usefulness are the strongest determinants of intention to adopt autonomous vehicles [9]. Underlying most of these factors is the inherent trust a user has in the self-driving technology. Trust is a key factor in a person's willingness to accept the technology and is largely shaped by prior experiences [7,10]. Lee and See [6] provide a foundational framework for understanding trust in automation, defining it as the attitude that an agent has that will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability. Parasuraman and Riley [11] further distinguish between appropriate use, misuse (over-reliance), and disuse (under-reliance) of automation, each driven by the calibration of trust relative to actual system capability.

Survey studies have revealed a social dilemma in autonomous vehicle ethics: while respondents generally approve of utilitarian autonomous vehicles (AVs) that sacrifice their passenger to save a greater number of lives, they themselves would prefer to ride in a self-protective vehicle [12]. This tension between collective endorsement and individual self-interest complicates trust formation, as passengers must accept a system whose ethical programming may not prioritize their personal safety. Shariff et al. [13] propose that the discussion of risk needs to be posed in terms of "absolute risk" rather than relative risk: by driving a self-driving car, one diminishes one's total risk of injury, and therefore one should not focus on edge cases where one's safety may not be prioritized. When considering risk from an absolute perspective, users may be more likely to adopt a self-driving car, as their chances of survival on any given drive are overall maximized by doing so.

One method of cultivating trust for self-driving cars is to improve human–machine interaction. By designing self-driving cars that communicate with passengers and give them a more active role in the experience, developers can deliver a more trustworthy system, e.g., through adaptive mood control [14] and adaptive driving mode [15]. This is supported by research conducted by Hartwich et al., whose evidence suggests that even given an SAE Level 4–5 system where no human interaction is required, the introduction of monitoring tools significantly improves passenger trust [16]. Further research conducted by Hartwich et al. shows that the first experience with self-driving cars greatly impacts the trust one associates with the technology [17].

In addition to the first experience significantly affecting a user's perception of the technology, research from Shahrdar et al. shows how trust is greatly affected by the driving style used and that defensive driving builds more trust than aggressive driving in virtual reality simulated tests [18]. These tests also showed that while initial experiences were important, trust in the system can be rebuilt following faulty behavior given enough time experiencing safer and more defensive driving from the self-driving car. Cegarra et al. found that initial trust affects how conventional vehicle drivers behave in traffic with AVs,

with higher trust leading to more risk-taking behavior around autonomous vehicles [19]. Furthermore, the amount of control a user has appears to be a significant factor in a user's ability to trust a given system. Research has shown that in the classical scenario of a person being chauffeured, there is an increased level of discomfort with being a passenger compared with being an active driver [20], and it appears that this analog extends to the self-driving car scenario; yet, there is still a decreased amount of trust in robotic drivers versus a human driver given equivalent driving behaviors as shown by Mühl et al. [21]. This suggests that self-driving cars not only have to perform as well as a human driver but must also perform even better to gain equivalent trust from their passengers.

Studies by Kolekar et al. suggest self-driving behavior can be made more human-like with the introduction of "driver risk field" modeling where the car's behavior is tuned to a given driver's perceived risk when executing driving maneuvers [22]. This generates autonomous behavior that is more in line with human driving than the current status quo, which only tries to minimize the real risk posed by a driving scenario. Beyond increasing the interactions passengers can experience with a self-driving car, the driving style can also be modified to increase trust in the system. Research conducted by Basu et al. showed that a more defensive driving style led to higher comfort and preference in autonomous driving scenarios [23]. When participants were surveyed on their driving preferences, they responded that they would want an experience similar to their own driving style for a self-driving car. Yet, when passengers were placed in a simulation, participants preferred a driving style that they thought was their own but was in fact much more conservative than their actual driving style [23].

This coincides with research conducted by Hajiseyedjavadi et al., who showed that in a simulator, drivers preferred their own driving styles over a faster one but still provided negative feedback when replaying their own driving style on urban roads versus rural roads, suggesting that environmental conditions play a significant part in preferred driving style [24]. These results are partly supported by Dettmann et al., who showed that younger drivers preferred autonomous driving styles similar to their own, while older drivers showed greater tolerance for driving styles more aggressive than their own [25]. Dettmann et al. also concluded by stating that the main factors in driving style preference depend on "speed, acceleration and deceleration behavior as well as distance control" [25]. In a related study, Schlüter et al. found that technological affinity and skepticism toward the technology should also be taken into account when designing adaptive driving style autonomous vehicle technologies [26].

Further research conducted by Bellem et al. showed how participants in a simulation study preferred driving styles that minimized acceleration and jerk when performing driving actions such as lane changing [27]. Bellem et al. also reported that personality traits associated with participants did not have any significant observable effects on autonomous driving preferences [27]. Consistent with the finding that drivers prefer conservative self-driving car (SDC) behavior, Craig et al. reported that surveyed participants showed that they expect a self-driving car to behave in a slightly less aggressive manner than their own driving style [28]. Methods proposed by Park et al. suggest adapting the driving behavior based on electroencephalography feedback to establish and maintain trust in the system [29,30].

Beyond any technical limitations, there are also numerous legal issues that arise with highly autonomous systems. From a legal standpoint, liability frameworks for autonomous vehicles remain fragmented; while some manufacturers now accept responsibility when their automated driving systems are engaged [2], no uniform international standard has emerged. In the case of semi-autonomous systems, liability attribution becomes less clear. Research from Awad et al. suggests that drivers are blamed more than the automated

systems in these semi-autonomous situations even when both make errors [31]. These findings suggest that more clearly defining liability attribution in autonomous systems could help improve public trust in the technology. With these questions in mind, we must further consider how users will respond to these technologies outside of the demographics in which research is collected. An important question remains as to how research participants are biased by the infrastructure and cultural norms of the country in which research takes place. There are some surveys that provide an international view, such as research conducted by Deloitte in 2020, which provided responses by country (South Korea, Japan, United States, Germany, India, and China) detailing the percentage of consumers who believe SDCs will not be safe. The results provided by Deloitte for most countries follow the United States' sentiments (∼50% believe they will not be safe) with some outliers, such as China, whose survey data suggests a more trusting sentiment, and India, whose survey data suggests a less trusting sentiment [32]. Large-scale cross-national studies further illustrate cross-cultural variation: Kyriakidis et al. [33] surveyed over 5000 respondents from 109 countries and found that respondents from more developed countries expressed greater concern about automated driving, while Muzammel et al. [34] demonstrated that Hofstede's cultural dimensions significantly predict national-level variation in AV acceptance. More recent multi-country validation studies [35,36] confirm that cultural values moderate AV acceptance, underscoring the need for cross-cultural validation of trust instruments.

With prominent Level 4 deployments such as Waymo's driverless service operating primarily in US cities [3], it becomes challenging to understand the global needs of this technology when so much of this research is based on the experiences of US drivers on American roads. Throughout this paper, we establish new scales through surveys that allow for measurement of various driving behaviors and expected behaviors from self-driving cars. Because trust calibration theory predicts that higher trust in automation reduces the perceived need for conservative safety margins [6], we hypothesize that individuals with greater AI trust will show a smaller gap between their self-reported driving aggressiveness and the aggressiveness they prefer from an SDC. Using these scales, we address the following research questions:

**RQ1.** Do drivers prefer SDC driving behavior that is more conservative than their own, and does this preference vary across countries?

**RQ2.** Is Artificial Intelligence (AI) trust level associated with the gap between self-reported driving aggressiveness and preferred SDC aggressiveness?

**RQ3.** Do cross-cultural differences exist in driving behavior, SDC preferences, AI trust, and driver safety behaviors across the US, Germany, and Panama?

This journal version extends the conference paper [37] with multi-group measurement invariance testing, discriminant validity analysis of AI Trust (AIT) versus AI Driving Mechanics Trust (AIDMT), AIT threshold sensitivity analysis, multivariate regression controlling for demographics, Bonferroni correction for multiple comparisons, and post-hoc power analysis.

## 2. Materials and Methods

All statistical analyses were implemented in Python (version 3.12.3) and R (version 4.3.3) by the authors; Anthropic Claude (Claude Opus 4.6, 2026) was used as a coding assistant and to verify consistency of reported statistics across the manuscript.

### *2.1. Participants and Sampling*

This study contributes cross-cultural data on driving behavior and SDC preferences through several newly defined metrics based on question groupings that further characterize user expectations of self-driving cars. We surveyed 157 people across the United

States, Germany, and Panama who were recruited through local networking and through PollPool.com. Panamanian data were primarily collected through the distribution of the survey among local contacts in the area who distributed the survey further among their peers. Data from Germany and the United States were primarily collected through PollPool. These regions were chosen as the US and Germany represent two global hubs of automotive manufacturing and are at the forefront of self-driving technologies at scale. In contrast, Panama was chosen as an emerging market due to its high number of consumer sales across Central America, where in 2019 it had the highest number of vehicles registered or sold in the region [38]. Panama also provides an important contrast as a Latin American market where road safety remains a significant public health challenge [39] and where cross-cultural communication norms may shape technology acceptance differently than in Western industrialized nations [40], making AV deployment both promising and sensitive to cultural context. Research on AV perceptions in Latin America remains scarce, with Marroquin et al. [41] providing one of the few regional studies, finding that interpersonal trust is the strongest predictor of AV acceptance across 18 Latin American countries.

Survey data were machine-translated from English into German and Spanish and subsequently reviewed by a native speaker of each language to verify the accuracy and cultural appropriateness of each item. Formal back-translation was not conducted; however, the measurement invariance analyses reported in Section 3.9 provide partial empirical evidence that the scales function similarly across the three language versions, though measurement invariance alone cannot establish semantic equivalence (see Section 4.5). Data were collected over the course of 2022. The respondents totaled 50 from the United States, 66 from Germany, and 41 from Panama. Surveys distributed in the United States and Germany were administered via Google Forms, with PollPool used as a recruitment platform to direct participants to the survey, introducing potential self-selection bias toward tech-savvy respondents. Panamanian data were collected through snowball sampling among local contacts, which may introduce network-based sampling bias. Both represent convenience samples and are not nationally representative; however, multivariate analyses controlling for demographic covariates partially address potential demographic skew across groups (see Section 2.9). This analysis contributes cross-cultural driving behavior and SDC preference data from an underrepresented Latin American market (Panama) alongside two Western automotive hubs (US and Germany), pairing self-reported driving behavior metrics with SDC driving style preference metrics across multiple countries.

### 2.2. Survey Procedure and Instruments

Participants were asked to complete a survey. The survey asked 57 questions relating to demographics, personal driving behaviors, and trust in AI and self-driving cars. The survey was structured into seven distinct parts as follows:

- **Part-1** provides demographic information including data about what country the driver currently resides in and what country they have driven the most in as well as age, gender, ethnicity, education, employment status, income range, etc.
- **Part-2** provides information regarding how a driver behaves on non-highway roads, and it is used to define an aggressiveness score for the driver.
- **Part-3** provides information regarding how a driver behaves on highway roads, and it is also used to define an aggressiveness score for the driver.
- **Part-4** provides information regarding general driving behaviors such as parking, turning, and driving under difficult weather conditions.
- **Part-5** provides information regarding how much drivers currently trust Artificial Intelligence (AI) and its applications to self-driving cars on a 5-point Likert-type scale: distrust, somewhat distrust, neutral, somewhat trust, and trust.

- **Part-6** provides data regarding how drivers expect AI/autonomy to perform on non-highway roads.
- **Part-7** provides data regarding how drivers expect AI/autonomy to perform on highway roads.

### 2.3. Quantitative Measurement

Each question asked can be related to a quantitative value to define *Driving Behavior Aggressiveness* (DBA), *Self-Driving Car Aggressiveness* (SDCA), *AI Driving Mechanics Trust* (AIDMT), general *AI Trust* (AIT), and *Driver Safety Score* (DSS) metrics.

Responses from highway-based and non-highway-based questions were averaged together to provide a more general scoring of the driver's aggressiveness in all situations. The same averaging method was applied to the SDCA items across highway and non-highway contexts. For DBA scores, a score of 0 represents a conservative driver and a score of 1 represents an aggressive driver. For SDCA scores, a score of 0 represents a conservative SDC and a score of 1 represents an aggressive SDC. These scores can then be used to contrast expectations of an SDC to a participant's own driving behaviors. For the trust scales, AIT and AIDMT scores of 0 represent full distrust and scores of 1 represent full trust. For DSS, a score of 0 represents the safest (most cautious) driving actions and 1 represents the least safe.

### 2.4. Unidimensionality of Scales

As a prerequisite to most consistency and reliability tests as they relate to new scales, the assumption of the unidimensionality of each scale must be evaluated. In this survey, we introduce five new scales to measure various driving behaviors based on survey responses. These scales are constructed from subsets of questions hypothesized to measure a single underlying factor. To evaluate whether these questions are unidimensional (i.e., they measure a single larger factor), we conducted an iterative confirmatory factor analysis (CFA) of these questions, in which theoretically motivated models were refined through inspection of modification indices and item diagnostics.

For the Driving Behavior Aggressiveness (DBA) scores, we constructed a scale from seven items relating to highway and non-highway driving. Correlated residuals (error covariances) were specified between item pairs guided primarily by empirical modification indices and, where applicable, substantive rationale such as parallel item content across road types (e.g., the same driving mechanics measured in highway versus non-highway contexts). The same correlated residual specifications were applied consistently to both the DBA and SDCA scales. Similarly, for the Self-Driving Car Aggressiveness metric, we constructed a scale consisting of 7 items that measure how aggressively a given user would want their self-driving car to behave. The 7 items used in this scale are the same as the 7 items asked in the DBA scale but now in the context of the self-driving car performing the action. The same correlated residual structure applied to the DBA scale was applied to the SDCA scale to maintain consistency. The next scale constructed was the AI Driving Mechanics Trust (AIDMT) scale, which employed 6 items that gauge on a 5-point Likert-type scale how much trust a user has in an AI car performing a variety of highway driving mechanics such as controlling speed and lane changing. Subsequently, the AI Trust scale (AIT) was constructed from 6 items that determine how much a user would trust AI performance in various complex scenarios. Finally, the Driver Safety Score (DSS) scale was created from seven items that determine how safe the sampled user is based on how they respond to various driving conditions.

During CFA model development, items with standardized factor loadings below 0.40 or whose removal improved model fit were candidates for exclusion. The AIT scale was

reduced from 7 to 6 items (one item dropped due to local dependency with an adjacent item), the DBA and SDCA scales from 10 to 7 items each (braking items from both road types and the non-highway speed item were removed), and the DSS scale from 10 to 7 items. The AIDMT scale retained 6 of its original 16 items; notably, all retained items concerned highway driving tasks, so the scale as validated reflects highway-specific driving mechanics trust rather than general driving mechanics trust. Full original item pools are available in the archived survey instruments [42].

The CFA considered how each item in a scale related to a single latent variable using polychoric correlations to account for the categorical nature of the ordinal response scales [43]. All CFA models were estimated using the Weighted Least Squares Mean and Variance adjusted (WLSMV) estimator, which is recommended for ordinal categorical data [43]. No missing data were present at the item level. The DSS model included one correlated residual between items involving lateral vehicle movement (turning behavior and night driving lane changes), guided by modification indices and specified on the same empirical basis as the DBA and SDCA residuals. Goodness of fit for each model was evaluated using the mean-adjusted (scaled) fit indices produced by WLSMV, which correct for non-normality due to the ordinal data. The primary fit indices reported are the $\chi^2$ test statistic and its $p$-value, the Comparative Fit Index (CFI), and the Tucker–Lewis Index (TLI). Root Mean Square Error of Approximation (RMSEA) is also provided but should be interpreted cautiously for this study due to known issues with overestimating misfit when the degrees of freedom and sample size are small [44]. Table 1 presents the fit parameters for each scale. Standardized factor loadings range from 0.45 to 0.92 for DBA, 0.74 to 0.89 for SDCA, 0.75 to 0.85 for AIT, 0.74 to 0.87 for AIDMT, and 0.49 to 0.72 for DSS. The $\chi^2$ $p$-value for each scale is greater than 0.05 and thus non-significant. CFI and TLI values exceed 0.95 for all five scales, indicating good model fit [45]. RMSEA values range from 0.025 to 0.074, indicating close to reasonable fit, though as noted above, RMSEA should be interpreted cautiously given the small sample size and low degrees of freedom; the CFI/TLI values provide stronger evidence of model fit.

**Table 1.** Goodness of fit metrics for each created scale.

| | **Goodness of Fit Metrics** | | | | | |
|---|---|---|---|---|---|---|
| | $\chi^2$ | **df** | $p(\chi^2)$ | **CFI** | **TLI** | **RMSEA** |
| DBA | 16.540 | 11 | 0.122 | 0.989 | 0.979 | 0.057 |
| SDCA | 18.356 | 11 | 0.074 | 0.995 | 0.991 | 0.065 |
| AIDMT | 14.656 | 9 | 0.101 | 0.997 | 0.995 | 0.063 |
| AIT | 16.609 | 9 | 0.055 | 0.995 | 0.991 | 0.074 |
| DSS | 14.294 | 13 | 0.353 | 0.994 | 0.990 | 0.025 |

### *2.5. Consistency and Reliability in Measurement*

To establish the reliability of our analysis, the consistency of each scale must be demonstrated. While Cronbach's alpha is often used for this purpose in the literature, its assumptions are frequently violated in practice, as it assumes tau-equivalence of the items within each set, i.e., each item contributes equally to the general factor being measured, as well as assumes that the general factor being measured is unidimensional [46,47].

Furthermore, if either of these requirements is not met, the reported Cronbach's alpha value may over- or underestimate the reliability of the test, depending on the pattern of loadings and error correlations [46,48]. To address these issues, we consider a model-based measure of reliability known as McDonald's Omega [48,49], which takes a factor analysis approach to deriving correlations between items. McDonald's Omega maintains the same

range and threshold of accepted consistency as Cronbach's alpha, where Omega values greater than 0.7 are considered acceptable, while not requiring tau-equivalence among the items. Similar to Cronbach's alpha, the data are required to be unidimensional, which we have established for our data using the CFA in the previous section. Table 2 presents the resulting Omega values. Our results for McDonald's Omega exceed 0.7, confirming good reliability for all five scales.

**Table 2.** McDonald's Omega for each created scale.

| McDonald's Omega | |
|---|---|
| **Scale** | $\omega_u$ |
| DBA | 0.788 |
| SDCA | 0.863 |
| AIDMT | 0.908 |
| AIT | 0.896 |
| DSS | 0.705 |

### 2.6. Measurement Invariance

This study compares scale scores across three culturally and linguistically distinct groups; as such, it is essential to establish that the scales function equivalently across groups. To do so, we established evidence for measurement invariance [50,51]. Without this evidence, observed cross-cultural differences may reflect measurement artifacts rather than true group differences. Prior work by Chien et al. [52] validated a trust-in-automation scale across US, German, Taiwanese, and Turkish samples, demonstrating the importance of such validation for cross-cultural trust research. Multi-group CFA was conducted for each of the five scales using the WLSMV estimator with theta parameterization, which is preferred for ordinal response data [53]. Three progressively restrictive models were tested: (1) *configural invariance* (same factor structure across groups), (2) *metric invariance* (equal factor loadings), and (3) *scalar invariance* (equal loadings and thresholds). Model comparisons used the change in CFI (ΔCFI), where $|\Delta\text{CFI}| < 0.01$ indicates that invariance holds [50]. We note that Chen [50] suggested a stricter threshold of $|\Delta\text{CFI}| < 0.005$ for small, unequal samples; however, the 0.01 criterion remains the most widely applied standard in applied measurement invariance research [51] and is adopted here. For scales where some response categories were empty within specific country groups (DSS, SDCA, and DBA), sparse categories were collapsed to enable model convergence. Results are presented in Section 3.9.

### 2.7. Discriminant Validity: AIT Versus AIDMT

The moderate observed Pearson correlation between the AIT and AIDMT scales ($r = 0.636$) raises the question of whether these scales measure genuinely distinct constructs. AIT captures *general* trust in AI and autonomous technologies, whereas AIDMT measures trust in *specific driving mechanics* performed by an autonomous vehicle. This distinction mirrors the established generalized versus task-specific trust framework in organizational psychology [54]. To empirically validate this distinction, three complementary analyses were conducted: (1) a comparison of two-factor versus one-factor CFA models using chi-square difference testing, (2) the Fornell–Larcker criterion, requiring each scale's Average Variance Extracted (AVE) to exceed the squared inter-factor correlation, and (3) the Heterotrait–Monotrait (HTMT) ratio, where HTMT $< 0.85$ supports discriminant validity [55]. Results are presented in Section 3.10.

### 2.8. AIT Threshold Justification

This study classifies participants as "trustful" (AIT $> 0.5$) or "distrustful" (AIT $\leq 0.5$) when examining the relationship between AI trust and SDC aggressiveness preferences (Section 3). The 0.5 threshold represents the neutral midpoint of the 0–1 scale: an average item response above 0.5 indicates a net positive disposition toward trust, while a score at or below 0.5 indicates a net negative disposition, yielding $n_{trust} = 67$ and $n_{distrust} = 90$. To verify that findings are not sensitive to this particular threshold, a separate sensitivity analysis was conducted across seven thresholds (0.35, 0.40, 0.45, 0.50, 0.55, 0.60, and 0.65); in the sensitivity analysis, participants scoring exactly at each threshold were excluded to create clean group separation (e.g., at 0.50, 16 participants with AIT $= 0.5$ were excluded, yielding $n_{trust} = 67$ and $n_{distrust} = 74$). Additionally, continuous analyses using Spearman correlations and Ordinary Least Squares (OLS) regression were performed to complement the dichotomous approach. To move beyond a null-result interpretation for the trustful group, a Two One-Sided Tests (TOST) equivalence procedure [56], implemented using two one-sided Wilcoxon signed-rank tests on the paired DBA–SDCA differences, was applied to determine whether the DBA–SDCA difference in the trustful group was statistically equivalent to zero within a meaningful margin. The equivalence margin was set to $\pm 0.125$, equal to the distrustful group's observed pseudo-median, representing the smallest empirically meaningful gap. TOST results are presented alongside the paired DBA–SDCA comparison in Section 3; threshold sensitivity results are presented in Section 3.11.

### 2.9. Multivariate Analysis

To assess whether observed cross-cultural differences persisted after accounting for demographic heterogeneity, OLS regression was conducted for each scale with country as the primary predictor and age, gender, and education level as covariates. Demographics were harmonized across the three survey languages by standardizing response categories. Education was coded ordinally (0 = none, 1 = high school or equivalent, 2 = bachelor's, 3 = master's, and 4 = doctoral/professional), with country-specific degree names mapped to the nearest equivalent level. Income was excluded due to incompatible currency units across countries. Driving experience (years of licensure) was not collected in the survey instrument and therefore could not be included as a covariate. Seven participants were excluded from the regression analyses due to missing demographic data (4 from Panama and 3 from Germany), yielding $N = 150$. OLS regression on averaged ordinal scale scores is standard practice in psychometric research when the dependent variable represents a multi-item composite [57,58]; averaging across 6–7 items produces a quasi-continuous distribution, and with $N = 150$, the Central Limit Theorem supports the assumption that the sampling distributions of the OLS coefficient estimates are approximately normal. Results are presented in Section 3.12.

## 3. Results

This section compares the scores generated from each of the defined quantitative metrics (DBA, SDCA, AIDMT, AIT, and DSS) against the collected demographic data to test whether there are statistically observable differences across demographics within a nation's population as well as from an international perspective. Distributions were compared using two-sided Mann–Whitney U tests for independent groups and two-sided Wilcoxon signed-rank tests for paired within-subject comparisons (e.g., DBA versus SDCA and AIT versus AIDMT) and reported if the resulting $p$-value was less than 0.05. Direction of effects was conveyed through signed effect size measures (Cliff's $\delta$, Hodges–Lehmann estimates, pseudo-medians, and rank-biserial correlations) rather than test sidedness. In comparisons involving multiple countries, a per-construct Bonferroni correction [59] was

applied within each scale to control the family-wise error rate. Because each scale measures a conceptually distinct construct, the correction was applied within each family of three country-pair comparisons ($\alpha = 0.05/3 = 0.0167$) rather than across all 15 tests globally. This approach scopes the correction to tests that share the same construct-specific null hypothesis, avoiding the inflation of type II error that results from penalizing each construct for comparisons on unrelated scales. Under a global Bonferroni correction across all 15 tests ($\alpha = 0.05/15 = 0.0033$), four of eight significant results would no longer reach significance (DBA US–Panama, SDCA US–Panama, SDCA Panama–Germany, and DSS Panama–Germany). Within-country paired comparisons (DBA vs. SDCA and AIT vs. AIDMT) are reported at $\alpha = 0.05$ without correction, as each paired test addresses a distinct within-subject hypothesis.

As the Mann–Whitney U test provides only a significance value without quantifying the magnitude of difference, additional statistical tools were employed for measures found significant. The Hodges–Lehmann (HL) estimator [60], bootstrapped with 10,000 iterations, was used to estimate the median location shift between distributions along with 95% confidence intervals; positive HL values indicate the first group scored higher than the second. This analysis also reports Cliff's $\delta$ [61] as a standardized effect size measure describing the tendency for scores in one group to be higher than the other. Cliff's $\delta$ is measured on a scale of $-1$ to $+1$, where $+1$ represents the case when all values in group A are greater than all values in group B, 0 represents groups with perfect overlap, and $-1$ represents all values in group A being less than all values in group B. For paired within-subject comparisons, the pseudo-median of within-subject differences and its 95% confidence interval are reported as the location-shift measure, along with the matched-pair rank-biserial correlation $r_{rb}$ as the effect size. Positive pseudo-median values indicate the first scale scored higher; $r_{rb}$ ranges from $-1$ to $+1$ with the same directional interpretation.

### *3.1. Summary Statistics*

Table 3 summarizes the demographic characteristics of participants by country. Panamanian respondents were older on average ($M = 33.9$; $SD = 12.8$) than US ($M = 27.7$; $SD = 8.3$) and German ($M = 26.7$; $SD = 5.2$) respondents. The US sample was predominantly male (68.0%), whereas the German sample was predominantly female (63.6%). Education levels varied, with Germany having the highest proportion of bachelor's (55.6%) and master's (23.8%) degree holders, while Panama had the highest proportion of high school or some college education (60.0%). Seven participants had incomplete demographic data (four from Panama and three from Germany) and were excluded from regression analyses, yielding $N = 150$ for those models.

To provide further evidence that our metrics measure different aspects of a respondent's driving profile, we considered the correlations among their responses. The Pearson correlation coefficient was evaluated between each measured metric across all demographics. The basic correlation analysis shows that most inter-scale correlations are weak to moderate, with the exception of AIT and AIDMT ($r = 0.636$), whose distinctness is empirically verified through discriminant validity analysis (Section 3.10). The remaining correlations support the conclusion that each scale is measuring a different aspect of driving behaviors and preferences. These results are shown in Table 4. The summary statistics for each surveyed demographic were also generated and are shown in Table 5. These tables provide a high-level overview of how each demographic tended to respond in each measured scale. A more detailed examination of how these distributions compare with one another is shown in the following sections; however, some differences in the distribution of responses across demographics are already apparent. Detailed paired and between-

group distribution comparisons for all scales are presented in Table 6, and all cross-country Mann–Whitney U tests with Bonferroni correction are summarized in Table 7.

**Table 3.** Participant demographics by country. Education percentages are computed from respondents with non-missing education data ($n_{US} = 50$, $n_{Panama} = 40$, and $n_{Germany} = 63$; $N = 153$).

| | **US ($n = 50$)** | **Panama ($n = 41$)** | **Germany ($n = 66$)** | **Overall ($N = 157$)** |
|---|---|---|---|---|
| Age | | | | |
| *M* (*SD*) | 27.7 (8.3) | 33.9 (12.8) | 26.7 (5.2) | 28.8 (9.1) |
| Range | 18–51 | 18–64 | 20–50 | 18–64 |
| Gender (%) | | | | |
| Male | 68.0 | 56.1 | 36.4 | 51.6 |
| Female | 32.0 | 43.9 | 63.6 | 48.4 |
| Education (%) | | | | |
| High School/Some College | 36.0 | 60.0 | 14.3 | 33.3 |
| Bachelor's | 48.0 | 35.0 | 55.6 | 47.7 |
| Master's | 16.0 | 5.0 | 23.8 | 16.3 |
| Professional/Doctoral | 0.0 | 0.0 | 6.3 | 2.6 |

*Note.* Seven participants had incomplete demographic data (four from Panama and three from Germany) and were excluded from regression analyses.

**Table 4.** Pearson correlation of scales across all demographics surveyed. Scale scores are computed using CFA-validated items.

| **Pearson Correlation of Average Responses for Each Scale** | | | | | |
|---|---|---|---|---|---|
| | **DBA** | **SDCA** | **AIDMT** | **AIT** | **DSS** |
| DBA | 1 | 0.365 | 0.102 | −0.081 | 0.523 |
| SDCA | 0.365 | 1 | 0.266 | 0.300 | 0.494 |
| AIDMT | 0.102 | 0.266 | 1 | 0.636 | −0.040 |
| AIT | −0.081 | 0.300 | 0.636 | 1 | 0.037 |
| DSS | 0.523 | 0.494 | −0.040 | 0.037 | 1 |

**Table 5.** Summary statistics across each sampled demographic: Std. Dev. stands for standard deviation.

| | **International Summary Statistics** | | | | |
|---|---|---|---|---|---|
| | **DBA** | **SDCA** | **AIT** | **AIDMT** | **DSS** |
| Mean | 0.400 | 0.326 | 0.481 | 0.546 | 0.222 |
| Std. Dev. | 0.183 | 0.209 | 0.247 | 0.268 | 0.164 |
| Median | 0.429 | 0.357 | 0.500 | 0.542 | 0.214 |
| Min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Max | 0.786 | 1.000 | 1.000 | 1.000 | 0.750 |
| | **United States Summary Statistics** | | | | |
| | **DBA** | **SDCA** | **AIT** | **AIDMT** | **DSS** |
| Mean | 0.404 | 0.374 | 0.535 | 0.619 | 0.269 |
| Std. Dev. | 0.172 | 0.227 | 0.228 | 0.247 | 0.170 |
| Median | 0.429 | 0.357 | 0.542 | 0.604 | 0.250 |
| Min | 0.036 | 0.000 | 0.083 | 0.000 | 0.000 |
| Max | 0.714 | 1.000 | 1.000 | 1.000 | 0.750 |

**Table 5.** *Cont.*

| | Germany Summary Statistics | | | | |
|---|---|---|---|---|---|
| | **DBA** | **SDCA** | **AIT** | **AIDMT** | **DSS** |
| Mean | 0.454 | 0.346 | 0.468 | 0.576 | 0.235 |
| Std. Dev. | 0.176 | 0.181 | 0.243 | 0.241 | 0.168 |
| Median | 0.429 | 0.393 | 0.458 | 0.542 | 0.179 |
| Min | 0.000 | 0.000 | 0.000 | 0.083 | 0.000 |
| Max | 0.786 | 0.786 | 1.000 | 1.000 | 0.750 |
| | **Panama Summary Statistics** | | | | |
| | **DBA** | **SDCA** | **AIT** | **AIDMT** | **DSS** |
| Mean | 0.309 | 0.233 | 0.435 | 0.410 | 0.145 |
| Std. Dev. | 0.175 | 0.204 | 0.269 | 0.288 | 0.117 |
| Median | 0.321 | 0.214 | 0.500 | 0.417 | 0.143 |
| Min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Max | 0.714 | 0.643 | 1.000 | 1.000 | 0.429 |

**Table 6.** Distribution comparisons across multiple countries. All $p$-values are two-sided. All scale scores are computed using CFA-validated items (see Tables 8–12). Panels (a)–(c) report paired Wilcoxon signed-rank tests (W statistic, pseudo-median of within-subject differences, and matched-pair rank-biserial $r_{rb}$) for within-subject comparisons; direction is indicated by the sign of the pseudo-median and $r_{rb}$. Panels (d)–(e) report Mann–Whitney U tests (HL estimate and Cliff's $\delta$) for between-group comparisons; positive HL and $\delta$ indicate the first-named group scored higher. For Bonferroni-corrected significance across all 15 country-pair comparisons, see Table 7. Individual trust-question $p$-values in panels (f)–(h) are exploratory and uncorrected for multiple comparisons.

**(a) Paired Wilcoxon Signed-Rank Comparison of Each Participant's DBA Versus SDCA, Split by AIT Trust Threshold (>0.5 = Trustful; ≤0.5 = Distrustful).**

| AIT | W | $p$-Value | Pseudo-med. | 95% CI | $r_{rb}$ |
|---|---|---|---|---|---|
| ≤0.5 | 2739.5 | <0.001 | 0.125 | (0.089, 0.179) | 0.65 |
| >0.5 | 891.5 | 0.863 | 0.000 | (−0.054, 0.054) | −0.03 |

**(b) Paired Wilcoxon Signed-Rank Comparison of DBA Versus SDCA Within Each Country and Combined Internationally (Two-Sided). Direction DBA > SDCA Based on Positive Pseudo-Medians and $r_{rb}$.**

| Country | W | $p$-Value | Pseudo-med. | 95% CI | $r_{rb}$ |
|---|---|---|---|---|---|
| US | 584.5 | 0.449 | 0.036 | (−0.036, 0.107) | 0.13 |
| Germany | 1366.5 | <0.001 | 0.107 | (0.054, 0.143) | 0.54 |
| Panama | 501.5 | 0.024 | 0.071 | (0.018, 0.143) | 0.43 |
| International | 6892.5 | <0.001 | 0.071 | (0.036, 0.107) | 0.38 |

**(c) Paired Wilcoxon Signed-Rank Comparison of AIT Versus AIDMT Within Each Country and Combined Internationally (Two-Sided). Direction AIDMT > AIT Based on Positive Pseudo-Medians and $r_{rb}$.**

| Country | W | $p$-Value | Pseudo-med. | 95% CI | $r_{rb}$ |
|---|---|---|---|---|---|
| US | 781.5 | 0.008 | 0.083 | (0.021, 0.146) | 0.45 |
| Germany | 1407.5 | <0.001 | 0.104 | (0.062, 0.167) | 0.59 |
| Panama | 224.5 | 0.460 | −0.021 | (−0.083, 0.042) | −0.15 |
| International | 6536.0 | <0.001 | 0.063 | (0.042, 0.104) | 0.38 |

**Table 6.** *Cont.*

| **(d) America vs. Panama—Distribution Comparison. Positive HL and $\delta$ Indicate US Scored Higher.** | | | | | |
|---|---|---|---|---|---|
| **Scale** | **U Statistic** | ***p*-Value** | **HL Estimate** | **95% CI** | **Cliff's $\delta$** |
| DBA | 1342.0 | 0.011 | 0.107 | (0.036, 0.179) | 0.31 |
| SDCA | 1383.0 | 0.004 | 0.143 | (0.071, 0.214) | 0.35 |
| AIDMT | 1477.0 | <0.001 | 0.250 | (0.125, 0.375) | 0.44 |
| DSS | 1484.0 | <0.001 | 0.107 | (0.071, 0.179) | 0.45 |
| **(e) Germany vs. Panama—Distribution Comparison. Positive HL and $\delta$ Indicate Germany Scored Higher.** | | | | | |
| **Scale** | **U Statistic** | ***p*-Value** | **HL Estimate** | **95% CI** | **Cliff's $\delta$** |
| DBA | 1959.5 | <0.001 | 0.143 | (0.071, 0.214) | 0.45 |
| SDCA | 1770.0 | 0.007 | 0.143 | (0.000, 0.214) | 0.31 |
| AIDMT | 1823.0 | 0.003 | 0.208 | (0.083, 0.292) | 0.35 |
| DSS | 1752.5 | 0.010 | 0.071 | (0.000, 0.143) | 0.30 |
| **(f) Germany vs. Panama—Trust Questions. Positive HL and $\delta$ Indicate Germany Scored Higher.** | | | | | |
| **AIT Question** | **U Statistic** | ***p*-Value** | **HL Estimate** | **95% CI** | **Cliff's $\delta$** |
| AIT4 | 1053.0 | 0.047 | −0.250 | (−0.250, 0.000) | −0.22 |
| AIT5 | 1739.5 | 0.011 | 0.250 | (0.000, 0.250) | 0.29 |
| **(g) America vs. Germany—Trust Questions. Positive HL and $\delta$ Indicate US Scored Higher.** | | | | | |
| **AIT Question** | **U Statistic** | ***p*-Value** | **HL Estimate** | **95% CI** | **Cliff's $\delta$** |
| AIT4 | 2105.5 | 0.009 | 0.250 | (0.000, 0.250) | 0.28 |
| **(h) America vs. Panama—Trust Questions. Positive HL and $\delta$ Indicate US Scored Higher.** | | | | | |
| **AIT Question** | **U Statistic** | ***p*-Value** | **HL Estimate** | **95% CI** | **Cliff's $\delta$** |
| AIT2 | 1291.5 | 0.028 | 0.250 | (0.000, 0.250) | 0.26 |
| AIT6 | 1393.5 | 0.003 | 0.250 | (0.000, 0.250) | 0.36 |

**Table 7.** All cross-country Mann–Whitney U tests with per-construct Bonferroni correction and Cliff's $\delta$ effect sizes. Scale scores are computed using CFA-validated items (see Tables 8–12). Each scale's three country-pair comparisons are corrected as an independent family ($\alpha = 0.05/3 = 0.0167$). Effect size interpretation: N = negligible ($|\delta| < 0.147$), S = small (<0.33), M = medium (<0.474), and L = large (≥0.474). Positive $\delta$ indicates the first group scored higher. All *p*-values are two-sided. * $p < 0.0167$ (Bonferroni-corrected).

| **Scale** | **Comparison** | **U** | ***p*** | **Cliff's $\delta$** | **Sig.** |
|---|---|---|---|---|---|
| DBA | US vs. Panama | 1342.0 | 0.011 | 0.309 (S) | * |
| DBA | US vs. Germany | 1408.5 | 0.177 | −0.146 (N) | |
| DBA | Panama vs. Germany | 746.5 | <0.001 | −0.448 (M) | * |
| SDCA | US vs. Panama | 1383.0 | 0.004 | 0.349 (M) | * |
| SDCA | US vs. Germany | 1756.0 | 0.552 | 0.064 (N) | |
| SDCA | Panama vs. Germany | 936.0 | 0.007 | −0.308 (S) | * |

**Table 7.** *Cont.*

| Scale | Comparison | U | $p$ | Cliff's $\delta$ | Sig. |
|---|---|---|---|---|---|
| AIT | US vs. Panama | 1266.5 | 0.054 | 0.236 (S) | |
| AIT | US vs. Germany | 1914.5 | 0.140 | 0.160 (S) | |
| AIT | Panama vs. Germany | 1236.5 | 0.457 | −0.086 (N) | |
| AIDMT | US vs. Panama | 1477.0 | <0.001 | 0.441 (M) | * |
| AIDMT | US vs. Germany | 1850.5 | 0.264 | 0.122 (N) | |
| AIDMT | Panama vs. Germany | 883.0 | 0.003 | −0.347 (M) | * |
| DSS | US vs. Panama | 1484.0 | <0.001 | 0.448 (M) | * |
| DSS | US vs. Germany | 1895.0 | 0.172 | 0.148 (S) | |
| DSS | Panama vs. Germany | 953.5 | 0.010 | −0.295 (S) | * |

**Table 8.** DBA scale items.

| | DBA Questions Used for Scale |
|---|---|
| DBA1 | Which best describes your driving behavior most of the time in terms of speed while driving on: THE HIGHWAY |
| DBA2 | Which best describes your lane changing behavior when driving on: THE HIGHWAY |
| DBA3 | How would you describe the way you accelerate and decelerate while driving on: THE HIGHWAY |
| DBA4 | How often do you pass others when driving on: THE HIGHWAY |
| DBA5 | What best describes your lane changing behavior when driving on: NON-HIGHWAY ROADS |
| DBA6 | How would you describe the way you accelerate and decelerate while driving on: NON-HIGHWAY ROADS |
| DBA7 | How often do you pass other vehicles when driving on: NON-HIGHWAY ROADS |

**Table 9.** SDCA scale items.

| | SDCA Questions Used for Scale |
|---|---|
| SDCA1 | If you are traveling in a self-driving car, and the car is in control of the speed, what range speed would you feel most comfortable with when driving on: HIGHWAY ROADS |
| SDCA2 | If you are traveling in a self-driving car, on HIGHWAY ROADS, you expect the car to change lanes: |
| SDCA3 | If you are traveling in a self-driving car, on HIGHWAY ROADS, your preference for the way it accelerates and decelerates would be: |
| SDCA4 | If you are traveling in a self-driving car, how often would you expect the car to pass other vehicles when driving on: HIGHWAY ROADS |
| SDCA5 | If you are traveling in a self-driving car, on NON-HIGHWAY ROADS, you expect the car to change lanes: |
| SDCA6 | If you are traveling in a self-driving car, on NON-HIGHWAY ROADS, your preference for the way it accelerates and decelerates would be: |
| SDCA7 | If you are traveling in a self-driving car, how often would you expect the car to pass other vehicles when driving on: NON-HIGHWAY ROADS |

**Table 10.** AIT scale items.

| | AIT Questions Used for Scale |
|---|---|
| AIT1 | What is your trust level to utilize Artificial Intelligence or Fully Autonomous Technologies? |
| AIT2 | What is your trust level that self-driving cars will keep your own safety as its primary objective? |
| AIT3 | What is your trust level that self-driving cars will be able to navigate in construction zones that include temporary detours that would ordinarily go against the flow of traffic? |
| AIT4 | What is your trust level that self-driving cars will be able to navigate in crowded pedestrian areas? |
| AIT5 | What is your trust level that self-driving cars will successfully get you to the EXACT destination you requested? |
| AIT6 | What is your trust level in the ability of self-driving cars to navigate safely with no person in the vehicle? |

**Table 11.** AIDMT scale items.

| | AIDMT Questions Used for Scale |
|---|---|
| AIDMT1 | If you were in a self-driving car, what tasks do you feel comfortable handing over to the car to perform autonomously? Please rank your trust level when it comes to the following driving tasks being performed by a self-driving car when driving on: HIGHWAY ROADS [The speed of the car] |
| AIDMT2 | If you were in a self-driving car, what tasks do you feel comfortable handing over to the car to perform autonomously? Please rank your trust level when it comes to the following driving tasks being performed by a self-driving car when driving on: HIGHWAY ROADS [Changing lanes] |
| AIDMT3 | If you were in a self-driving car, what tasks do you feel comfortable handing over to the car to perform autonomously? Please rank your trust level when it comes to the following driving tasks being performed by a self-driving car when driving on: HIGHWAY ROADS [Signaling] |
| AIDMT4 | If you were in a self-driving car, what tasks do you feel comfortable handing over to the car to perform autonomously? Please rank your trust level when it comes to the following driving tasks being performed by a self-driving car when driving on: HIGHWAY ROADS [Driving in hazardous weather conditions] |
| AIDMT5 | If you were in a self-driving car, what tasks do you feel comfortable handing over to the car to perform autonomously? Please rank your trust level when it comes to the following driving tasks being performed by a self-driving car when driving on: HIGHWAY ROADS [Braking] |
| AIDMT6 | If you were in a self-driving car, what tasks do you feel comfortable handing over to the car to perform autonomously? Please rank your trust level when it comes to the following driving tasks being performed by a self-driving car when driving on: HIGHWAY ROADS [Maintaining a certain distance from the cars around you] |

**Table 12.** DSS scale items.

| | **DSS Questions Used for Scale** |
|---|---|
| DSS1 | What best describes your driving behavior when making a turn? |
| DSS2 | What best describes your driving behavior when driving on a winding road? |
| DSS3 | When driving in that less than perfect weather condition, what happens to your speed? |
| DSS4 | How would you describe your lane changing behavior when driving in that less than perfect weather condition? |
| DSS5 | Which best describes your driving behavior most of the time in terms of speed while driving at night? |
| DSS6 | Which best describes your behavior most of the time when it comes to signaling lane changes while driving at night: |
| DSS7 | Which best describes your lane changing behavior when driving at night? |

### *3.2. International DBA and SDCA Metrics Across All Demographics*

The following DBA and SDCA scores shown in Figure 1a,b were generated and compared at an international level. These distributions show that across the combined international sample, most drivers behave in a conservative to moderate fashion and most drivers prefer a more conservative self-driving car compared with their own driving behaviors. This finding supports previous research conducted in [28], which provided similar distributions across these metrics (see Table 6).

Using a two-sided Wilcoxon signed-rank test (paired), we found evidence that these distributions are statistically different, with the underlying distribution of the DBA score being statistically greater than the underlying distribution of the SDCA score (positive pseudo-median and $r_{rb}$; see Table 6) with a $p$-value $< 0.001$. Also relevant is the relationship between people's trust in AI and how they would like their SDC to perform. For this analysis, we considered a driver with an AIT score of less than or equal to 0.5 to be generally distrustful of AI technology and a driver with a trust score greater than 0.5 to be generally trustful of AI technology. Under these parameters, the following SDCA score distributions are considered and shown in Figure 1. Using a two-sided Wilcoxon signed-rank test, we compared each participant's own DBA and SDCA scores within each trust group. The comparison of the SDCA of drivers who are distrustful of AI technology with their own driving behaviors shows a $p$-value of $<0.001$, indicating that the underlying distributions are not equal and agreeing with the general result that people want an SDC that is *more conservative than their own driving behavior*. However, this result changes when we consider the SDCA of drivers who are trustful of AI technology compared with their own driving behavior. In this case, we failed to reject the null hypothesis with a $p$-value of 0.863, finding no evidence that the trustful SDCA distribution differs from the DBA distribution. A TOST equivalence test confirmed that the trustful group's DBA–SDCA difference is statistically equivalent to zero within the equivalence margin of $\pm 0.125$ (the distrustful group's pseudo-median): TOST $p < 0.001$; the 90% confidence interval of the trustful pseudo-median $[-0.036, 0.036]$ falls entirely within the equivalence bounds. Thus, the trustful group's DBA–SDCA gap is not merely non-significant but *demonstrably negligible*, supporting the interpretation that users who are more trustful of AI technology show *DBA–SDCA equivalence, consistent with acceptance of a driving style comparable to their own.*

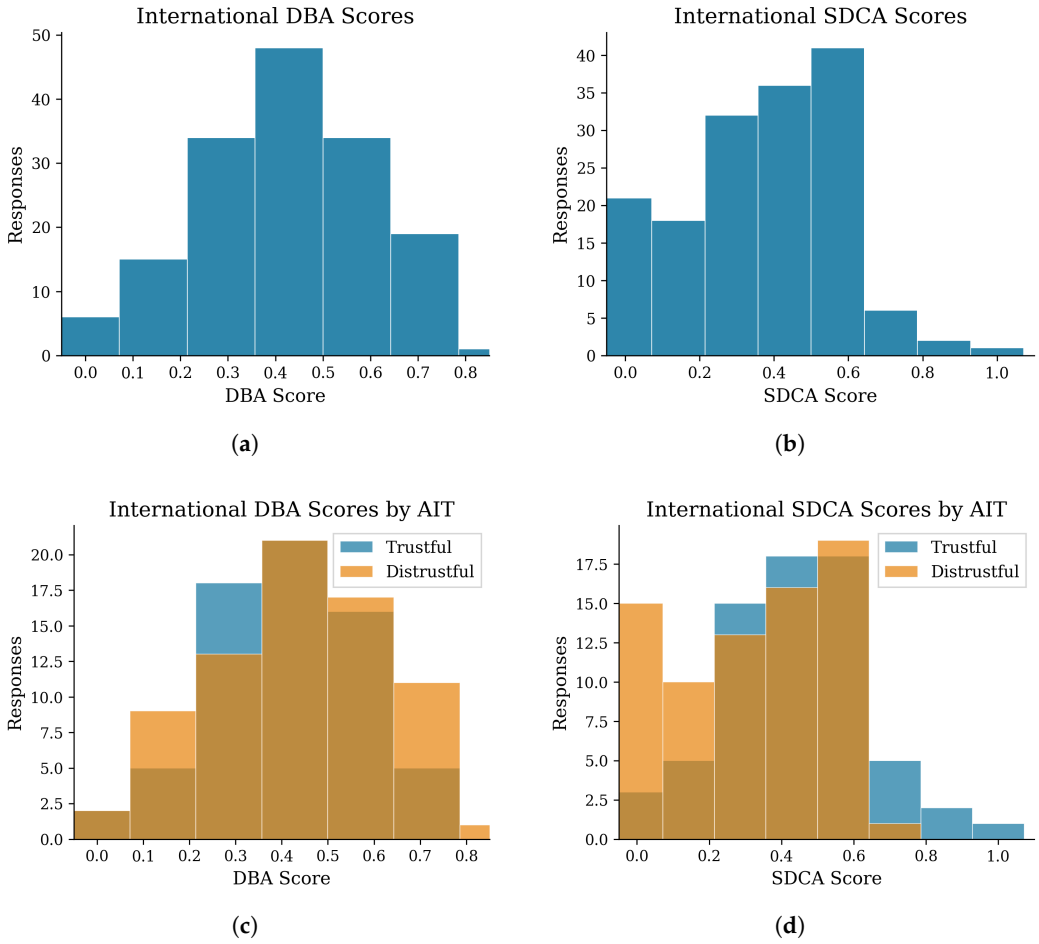



**Figure 1.** International DBA and SDCA scores as well as distributions split by AIT threshold. (**a**) DBA score measured across all demographics; (**b**) SDCA score measured across all demographics; (**c**) international DBA with AIT $>$ 0.5 represented as trustful; (**d**) international SDCA with AIT $>$ 0.5 represented as trustful.

### *3.3. American Respondent Analysis*

This analysis examines whether the same difference in preference holds when comparing DBA and SDCA scores for American respondents. A two-sided Wilcoxon signed-rank test (paired) failed to reject the null hypothesis when comparing these two distributions ($p = 0.449$; see Table 6). No evidence was found that American respondents prefer a different SDC aggressiveness level than their own driving behavior; however, this null result should not be interpreted as evidence of equivalence given the modest sample size ($n = 50$). The two trust metrics AIT and AIDMT were computed and showed that *American respondents had higher trust in AI's ability to perform the mechanics of driving than in the technology as a whole* with a corresponding $p$-value of 0.008 from a two-sided Wilcoxon signed-rank test (paired; see Table 6), providing evidence that the AIT and AIDMT distributions are stochastically different (AIDMT $>$ AIT based on positive pseudo-median).

### *3.4. German Respondent Analysis*

The German responses show a clear difference between the DBA and SDCA distributions. Using a two-sided Wilcoxon signed-rank test (paired), we observed a $p$-value of $<$0.001, providing statistical evidence that the German DBA and SDCA distributions are stochastically different (DBA $>$ SDCA based on positive pseudo-median; see Table 6). This measurement supports the idea that *German drivers prefer a more conservative SDC*

*than their own driving behaviors*. Notably, *Germans had the highest mean DBA score across all participants* for this research, though this difference was not statistically significant versus US respondents after Bonferroni correction (see Table 7). When considering AI Trust metrics, German respondents showed higher trust in AI's ability to perform the mechanics of driving (AIDMT) than in the technology as a whole (AIT). This is supported by a $p$-value of $<0.001$ from a two-sided Wilcoxon signed-rank test (paired; see Table 6) (AIDMT $>$ AIT based on positive pseudo-median).

### 3.5. Panamanian Respondent Analysis

The Panamanian responses showed a statistical difference between the DBA and SDCA scores with a $p$-value of 0.024 using a two-sided Wilcoxon signed-rank test (paired), providing evidence that the DBA and SDCA distributions are stochastically different (DBA $>$ SDCA based on positive pseudo-median; see Table 6). This result provides statistical evidence that *Panamanian drivers prefer a more conservative car behavior than their own driving behavior*. The two trust metrics AIT and AIDMT were also evaluated. Notably, *Panamanian respondents had significantly lower AIDMT scores compared with their American and German counterparts* (see Table 7), while AIT differences did not reach significance after Bonferroni correction. Furthermore, Panamanians had nearly equal trust in AI's ability to perform driving mechanics and in the technology as a whole; a two-sided Wilcoxon signed-rank test (paired) failed to reject the null hypothesis ($p = 0.460$; see Table 6), failing to provide evidence that the two distributions are stochastically different. This is in sharp contrast with both Americans and Germans, who had more trust in AI's ability to manage driving mechanics versus the technology as a whole.

### 3.6. Americans Versus Germans

When we compared Americans versus Germans, we found no significant differences in the generated distributions for any of the five metrics, DBA, SDCA, AIT, AIDMT, and DSS, after per-construct Bonferroni correction (all $p > 0.0167$; see Table 7). Note that post hoc power was below 0.26 for all US–Germany comparisons (see Section 4.5), so these null results should not be interpreted as evidence of equivalence. This paper also considers which individual AI Trust questions yielded statistically different distributions across demographics. These questions give insight into the needs of each demographic in terms of general AI trust. The resulting distributions show that Americans reported higher trust in the AI's ability to navigate crowded pedestrian areas compared with Germans, with a $p$-value of 0.009. Note that the individual trust–question comparisons reported in the cross-country subsections below are exploratory and are not corrected for multiple comparisons; they should be interpreted as hypothesis-generating rather than confirmatory.

### 3.7. Americans Versus Panamanians

When we compared Americans versus Panamanians, we found statistically significant results in four of the five metrics, DBA, SDCA, AIDMT, and DSS, with $p$-values of 0.011, 0.004, $<0.001$, and $<0.001$, respectively, in two-sided Mann–Whitney U tests, all surviving per-construct Bonferroni correction ($p < 0.0167$; see Table 7), with positive Cliff's $\delta$ values confirming that the distributions for these metrics were statistically higher for American respondents when compared with Panamanian respondents. The results illustrate that *Panamanian drivers prefer a more conservative SDC experience relative to American drivers*. Additionally, the results show that *Panamanians reported lower trust in AI's ability to perform driving mechanics*. The lower DSS scores further indicate that *Panamanian drivers tend to take more cautious driving actions compared with American drivers*. When considering the statistically significant AI trust questions, the results show that Americans believe AI will keep their safety as the highest priority compared with Panamanians, with a $p$-value of

0.028. Additionally, Americans reported higher trust in AI's ability to perform with no person in the vehicle compared with Panamanians, with a $p$-value of 0.003 in two-sided Mann–Whitney U tests.

### 3.8. Panamanians Versus Germans

When we compared Panamanians with Germans, we found statistically significant differences in DBA, SDCA, AIDMT, and DSS scores ($p < 0.001$, $p = 0.007$, $p = 0.003$, and $p = 0.010$, respectively; all $p < 0.0167$; see Table 7). These distributions indicate that *Panamanian drivers prefer a more conservative driving style compared with German drivers*. Additionally, *Panamanian drivers expect a more conservative self-driving car compared with German drivers*.

A statistically significant difference also emerged in how Panamanians trust the SDC to be able to perform driving mechanics. The AIDMT distributions show *Panamanians are less trustful in AI's ability to perform driving mechanics compared with Germans*. The lower DSS score distribution indicates that Panamanian drivers tend to take more cautious driving actions compared with German drivers. When considering the statistically significant AI trust questions, the resulting distributions show that Panamanians reported higher trust in AI's ability to navigate a crowded pedestrian area than Germans ($p = 0.047$). Additionally, Germans reported higher trust in AI's ability to navigate to an exact destination compared with Panamanians ($p = 0.011$).

### 3.9. Measurement Invariance Results

Multi-group CFA results for measurement invariance are summarized in Table 13 and Figure 2. Four of five scales (AIT, AIDMT, SDCA, and DBA) achieved full scalar invariance, with all $|\Delta CFI|$ values below the 0.01 threshold recommended by Chen [50]. The DSS scale achieved configural invariance but failed the configural-to-metric transition ($\Delta CFI = 0.023$), indicating that factor loadings may not be fully equivalent across groups for this scale. To investigate further, each item loading was systematically freed one at a time to identify the source of non-invariance. Freeing the loading for item DSS6 ("Which best describes your behavior most of the time when it comes to signaling lane changes while driving at night?") was sufficient to achieve partial metric invariance ($\Delta CFI = 0.000$; $CFI = 1.000$). Item DSS5 also independently achieved partial invariance when freed ($\Delta CFI = 0.001$). Following Byrne et al. [62], cross-cultural comparisons on the DSS remain interpretable under partial metric invariance, as the majority of factor loadings (six of seven) are invariant across groups. Notably, the two primary scales of interest for cross-cultural comparisons (SDCA and DBA) both demonstrate full scalar invariance, supporting the validity of mean-level comparisons across groups.

**Table 13.** Measurement invariance results across US ($n = 50$), Panama ($n = 41$), and Germany ($n = 66$). * Passes Chen [50] criterion ($|\Delta CFI| < 0.01$). † Partial metric invariance with item DSS6 loading freed across groups. Estimator: WLSMV with theta parameterization.

| Scale | Config. CFI | Metric CFI | Scalar CFI | $\Delta CFI_{C \to M}$ | $\Delta CFI_{M \to S}$ |
|---|---|---|---|---|---|
| AIT | 1.000 | 0.999 | 1.000 | 0.001 * | −0.001 * |
| AIDMT | 1.000 | 0.999 | 1.000 | 0.001 * | −0.001 * |
| SDCA | 1.000 | 1.000 | 1.000 | 0.000 * | 0.000 * |
| DBA | 1.000 | 1.000 | 0.994 | 0.000 * | 0.006 * |
| DSS | 1.000 | 0.977 | 0.986 | 0.023 | −0.009 * |
| DSS † | 1.000 | 1.000 | — | 0.000 * | — |

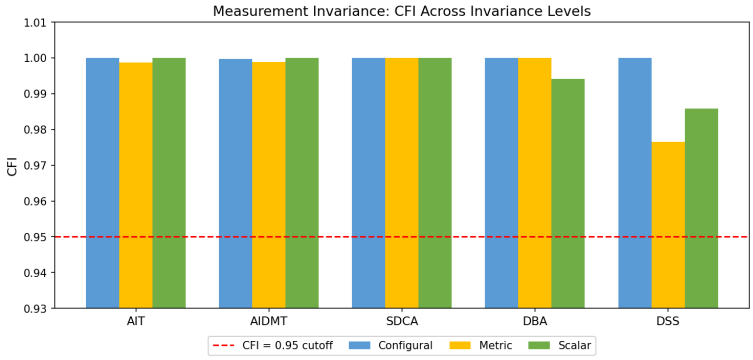



**Figure 2.** CFI values across configural, metric, and scalar invariance levels for each scale. All scales except DSS achieve full scalar invariance (all CFI values above 0.95 cutoff).

### *3.10. Discriminant Validity Results*

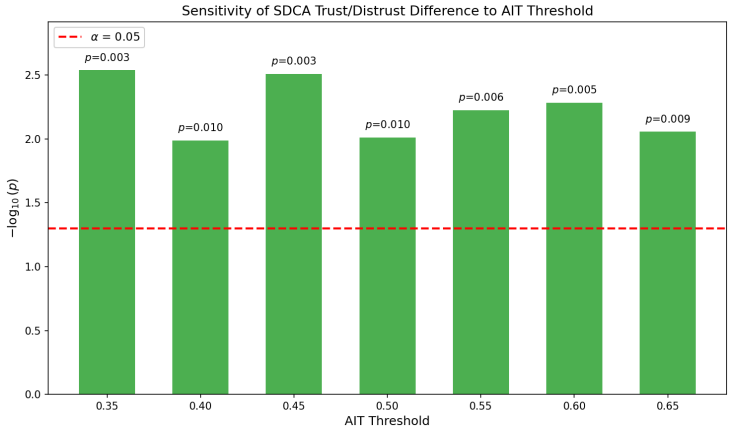
Three complementary analyses confirmed discriminant validity between AIT and AIDMT. First, the two-factor CFA model (CFI = 0.995) fit significantly better than a one-factor model (CFI = 0.975; $\chi^2$ difference $p < 0.001$), confirming that AIT and AIDMT measure distinct latent constructs. Second, the Fornell–Larcker criterion was satisfied: AVE(AIT) = 0.649 and AVE(AIDMT) = 0.689, both exceeding the squared inter-factor correlation of 0.545. Third, the HTMT ratio was 0.711, well below the 0.85 threshold recommended by Henseler et al. [55]. These results show that while AIT and AIDMT are related (latent factor $r = 0.738$; observed Pearson $r = 0.636$; see Table 4), they measure different aspects of trust.

### *3.11. AIT Threshold Sensitivity Results*

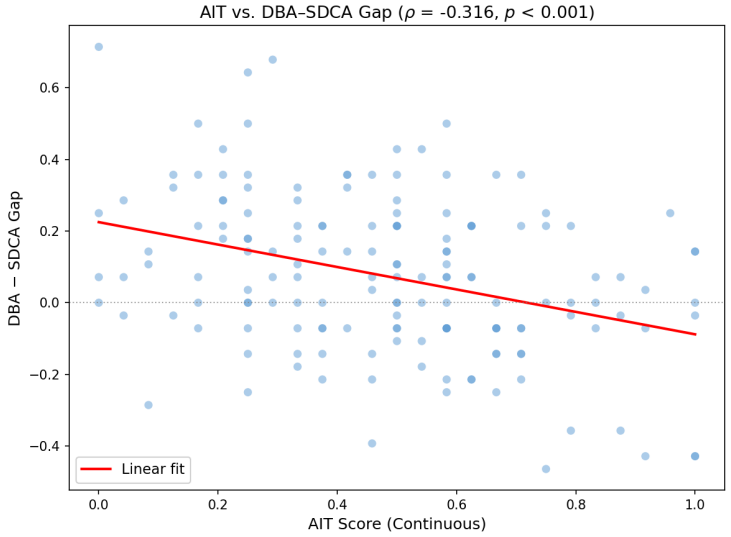
Table 14 and Figure 3 present the sensitivity analysis results. The SDCA distributions of trustful and distrustful groups differ significantly at all seven tested thresholds (0.35–0.65), with trustful individuals consistently reporting higher SDCA scores. Combined with the paired DBA–SDCA tests in Table 6a, this supports the finding that AI-trustful individuals show a smaller DBA–SDCA gap than distrustful individuals regardless of threshold choice. Continuous analyses also support this finding: the Spearman correlation between AIT scores and the DBA–SDCA gap was $\rho = -0.316$ ($p < 0.001$), indicating that higher AI trust is associated with a smaller gap between one's own driving aggressiveness and preferred SDC aggressiveness. OLS regression on the full sample ($N = 157$) confirmed that AIT significantly predicts SDCA ($b = 0.229$; $p < 0.001$) after controlling for country, with an overall model $R^2 = 0.144$.

**Table 14.** Sensitivity analysis: Mann–Whitney U tests comparing SDCA distributions of AI-Trustful versus AI-distrustful groups across different AIT thresholds. Participants scoring exactly at each threshold were excluded to create clean group separation; total $n$ varies accordingly. Effect size: S = small ($|\delta| < 0.33$).

| Threshold | $n_{trust}$ | $n_{distrust}$ | U | $p$-Value | Cliff's $\delta$ | Significant? |
|---|---|---|---|---|---|---|
| 0.35 | 104 | 53 | 3551.5 | 0.003 | 0.289 (S) | Yes |
| 0.40 | 96 | 61 | 3634.5 | 0.010 | 0.241 (S) | Yes |
| 0.45 | 91 | 66 | 3827.5 | 0.003 | 0.275 (S) | Yes |
| 0.50 | 67 | 74 | 3099.5 | 0.010 | 0.250 (S) | Yes |
| 0.55 | 62 | 95 | 3704.0 | 0.006 | 0.258 (S) | Yes |
| 0.60 | 46 | 111 | 3271.5 | 0.005 | 0.281 (S) | Yes |
| 0.65 | 39 | 118 | 2940.5 | 0.009 | 0.278 (S) | Yes |




(**a**)




(**b**)

**Figure 3.** AIT threshold sensitivity and continuous trust analysis. (**a**) Sensitivity of SDCA trust/distrust difference across AIT thresholds. Green bars indicate significant results ($p < 0.05$). (**b**) Continuous AIT score versus DBA–SDCA gap. Spearman's $\rho = -0.316$; $p < 0.001$.

### *3.12. Multivariate Regression Results*

Table 15 presents the results of OLS regression models controlling for age, gender, and education. After demographic adjustment, the Panama country effect remained significant for DBA ($b = -0.129$; $p = 0.004$), AIDMT ($b = -0.177$; $p = 0.004$), and DSS ($b = -0.110$; $p = 0.006$), indicating that the observed driving aggressiveness, trust, and safety differences between Panamanian respondents and the reference group (Germany) persist beyond demographic composition. Gender emerged as a significant predictor for AIT ($b = 0.157$; $p < 0.001$) and AIDMT ($b = 0.136$; $p = 0.002$), with male respondents reporting higher trust scores. The SDCA country effect for Panama was attenuated after demographic controls ($p = 0.079$), suggesting that some of the observed cross-cultural differences in SDC aggressiveness preferences may be partially attributable to demographic heterogeneity across samples. The persistence of DBA, AIDMT, and DSS country effects after controlling for demographics provides evidence that cross-cultural differences in driving behavior, AI driving trust, and safety behaviors reflect cultural variation rather than demographic differences alone. These results suggest that SDC aggressiveness preferences (SDCA) may be influenced by both cultural and demographic factors, while driving behavior aggressiveness (DBA), trust perceptions (AIDMT), and safety behaviors (DSS) are more

strongly tied to cultural context. The modest $R^2$ values (0.079–0.170) indicate that country and demographics explain only a small proportion of variance in scale scores; unmeasured factors such as driving experience, prior AV exposure, and personality traits likely account for additional variance.

**Table 15.** OLS regression results: Scale ~ Country + Age + Gender + Education. Reference group: Germany. $b$ = unstandardized coefficient. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. Education covariate included but omitted from table (non-significant for all scales). $N = 150$ (7 participants excluded due to missing demographic data). Model fit: DBA $R^2 = 0.093$, SDCA $R^2 = 0.079$, AIT $R^2 = 0.126$, AIDMT $R^2 = 0.170$, and DSS $R^2 = 0.103$. All scale scores use CFA-validated items.

| Scale | Predictor | *b* | SE | *t* | *p* | Sig. |
|---|---|---|---|---|---|---|
| DBA | Panama | −0.129 | 0.044 | −2.94 | 0.004 | ** |
| | US | −0.047 | 0.036 | −1.30 | 0.195 | |
| | Gender (M) | 0.028 | 0.030 | 0.92 | 0.359 | |
| | Age | −0.000 | 0.002 | −0.06 | 0.954 | |
| SDCA | Panama | −0.090 | 0.051 | −1.77 | 0.079 | |
| | US | 0.023 | 0.041 | 0.55 | 0.582 | |
| | Gender (M) | 0.033 | 0.035 | 0.92 | 0.357 | |
| | Age | −0.003 | 0.002 | −1.40 | 0.165 | |
| AIT | Panama | −0.081 | 0.058 | −1.39 | 0.168 | |
| | US | 0.020 | 0.048 | 0.41 | 0.683 | |
| | Gender (M) | 0.157 | 0.041 | 3.87 | <0.001 | *** |
| | Age | −0.001 | 0.002 | −0.35 | 0.728 | |
| AIDMT | Panama | −0.177 | 0.061 | −2.91 | 0.004 | ** |
| | US | 0.022 | 0.050 | 0.45 | 0.657 | |
| | Gender (M) | 0.136 | 0.042 | 3.22 | 0.002 | ** |
| | Age | −0.000 | 0.002 | −0.16 | 0.875 | |
| DSS | Panama | −0.110 | 0.039 | −2.80 | 0.006 | ** |
| | US | 0.018 | 0.032 | 0.57 | 0.567 | |
| | Gender (M) | 0.023 | 0.027 | 0.83 | 0.407 | |
| | Age | −0.001 | 0.002 | −0.40 | 0.692 | |

### 3.13. Bonferroni-Corrected Results

Table 7 presents all 15 cross-country Mann–Whitney U tests with per-construct Bonferroni correction [59]. Because each of the five scales measures a conceptually distinct construct, the correction is applied within each family of three country-pair comparisons ($\alpha = 0.05/3 = 0.0167$), rather than globally across all 15 tests. All eight tests significant at the uncorrected $\alpha = 0.05$ level also survive the per-construct Bonferroni correction. Four of five scales show significant differences between both the US and Panama and between Germany and Panama (DBA, SDCA, AIDMT, and DSS); only the AIT scale does not reach significance for any country pair. No significant differences were found between the US and Germany on any scale.

### 3.14. Post Hoc Power Analysis

Post hoc power analysis (computed at the Bonferroni-corrected $\alpha = 0.0167$) revealed that 5 of 15 pairwise comparisons achieved adequate statistical power ($\geq$0.80), with a median power of 0.711 across all tests. Tests involving the largest effect sizes (AIDMT comparisons with Panama and DSS US versus Panama) achieved a power of $\geq$0.85. However, comparisons involving small effect sizes between the US and Germany were underpowered

(power <0.26 in all cases), which should be considered when interpreting non-significant US–Germany differences.

## 4. Discussion

### *4.1. Statistically Significant Measures*

The approach taken when examining the survey data was an exploratory approach in which we compared relationships between metrics across demographics. When comparing multiple countries for each distribution, we applied per-construct Bonferroni correction [59] ($\alpha = 0.05/3 = 0.0167$ within each scale) to control the family-wise type I error rate. Distribution analysis done within each representative country used the standard 0.05 threshold of significance. Identifying statistical differences between distributions enables a better understanding of how to serve future passengers in self-driving cars. The greatest differences were found when comparing both US and German responses to those from Panama. Regarding RQ1, drivers internationally do prefer more conservative SDC behavior than their own, but this preference varies by country (significant for Germany and Panama but non-significant for the US). Regarding RQ2, AI trust level is associated with variation in the DBA–SDCA gap: trustful individuals show no significant gap, while distrustful individuals prefer significantly more conservative SDC behavior. Regarding RQ3, significant cross-cultural differences exist in DBA, SDCA, AIDMT, and DSS between both the US and Panama and between Germany and Panama but not between the US and Germany. These results are summarized in the following list and described in statistical detail in Table 6.

- **International DBA and SDCA Metrics**
  International distribution of DBA compared with SDCA showed drivers preferred a self-driving car that was more conservative than their own driving style ($p < 0.001$). International drivers who had higher AIT scores showed DBA–SDCA equivalence (TOST $p < 0.001$), consistent with acceptance of a driving style comparable to their own. In comparison, international drivers who had a low AIT score followed the trend of wanting a self-driving car that was more conservative than their own driving style ($p < 0.001$).
- **United States Metrics**
  American respondents showed no significant DBA–SDCA gap ($p = 0.449$), consistent with US drivers being more accepting of an SDC driving style comparable to their own. This non-significance is consistent with the trust-associated pattern observed internationally, as US respondents had moderate AIT scores, though the modest sample size ($n = 50$) limits this inference. American respondents had higher trust in AI's ability to perform the mechanics of driving than in the technology as a whole ($p = 0.008$).
- **Germany Metrics**
  German drivers prefer a more conservative SDC than their own driving behaviors ($p < 0.001$).
  German respondents had higher trust in AI's ability to perform the mechanics of driving than in the technology as a whole ($p < 0.001$).
- **Panama Metrics**
  Panamanian drivers prefer a more conservative SDC than their own driving behaviors ($p = 0.024$). Panamanian respondents uniquely showed no significant AIT–AIDMT gap ($p = 0.460$), suggesting that general AI trust and driving-specific trust are at comparable levels for Panamanian respondents, without the AIDMT premium observed in US and German respondents. This pattern may be related to lower regional digital adoption [63], though this study did not measure individual-level technology exposure.

- **United States vs. Germany Metrics**
  No significant differences between US and German respondents were found for any of the five scale-level metrics after per-construct Bonferroni correction (all $p > 0.0167$; see Table 7). At the individual question level (exploratory, uncorrected), US respondents had a higher level of trust that the self-driving car would be able to navigate a crowded pedestrian area compared with German respondents ($p = 0.009$).
- **United States vs. Panama Metrics**
  US respondents had higher metrics for DBA, SDCA, AIDMT, and DSS scores ($p = 0.011$, $p = 0.004$, $p < 0.001$, and $p < 0.001$, respectively; all $p < 0.0167$; see Table 7). The overall AIT scale did not differ significantly between US and Panama after correction ($p = 0.054 > 0.0167$).
  At the individual question level (exploratory, uncorrected), US respondents had a higher trust that the self-driving car would keep their safety as a priority compared with Panamanian respondents ($p = 0.028$), and US respondents had a higher level of trust that the self-driving car would be able to navigate safely with no person in the vehicle compared with Panamanian respondents ($p = 0.002$).
- **Germany vs. Panama Metrics**
  German respondents had higher metrics for DBA, SDCA, AIDMT, and DSS scores but scored similarly for the AIT metric ($p < 0.001$, $p = 0.007$, $p = 0.003$, and $p = 0.010$, respectively; all $p < 0.0167$; AIT $p = 0.457$; see Table 7).
  At the individual question level (exploratory, uncorrected), German respondents had a higher level of trust that the self-driving car would be able to navigate to an exact destination compared with Panamanian respondents ($p = 0.011$).

Several factors may underlie the observed differences. One factor could be the difference in road quality. According to the road quality indicator provided by the World Economic Forum [64], US and German roads score substantially higher than Central American roads, and US and German roads score similarly (5.5 and 5.3, respectively). Additionally, the digital adoption index (DAI) provided by the World Bank shows the US and Germany having a higher DAI than most Central American countries [63]. The low adoption rate of digital technologies could be one factor associated with the significantly lower AIDMT scores observed among Panamanian respondents, though general AI trust (AIT) did not differ significantly across countries. Overall, understanding the proper context driving these differences will be essential for delivering autonomous vehicles and AI technology internationally.

### *4.2. Psychological Interpretation*

AI-trustful individuals showed DBA–SDCA equivalence (TOST $p < 0.001$; 90% CI of pseudo-median within $\pm 0.125$), providing positive evidence that their preferred SDC aggressiveness is comparable to their own driving behavior. Distrustful individuals, by contrast, preferred a substantially more conservative SDC style, with their stated DBA consistently exceeding their preferred SDCA. Within the trust-in-automation framework [11], this pattern is consistent with distrustful users displaying conservative trust calibration, while trustful users exhibit calibration that permits the vehicle to operate closer to human-normative behavior. We note that the original framework was developed for contexts involving direct experience with automation, whereas the present data reflect stated preferences from participants without firsthand AV experience; accordingly, these calibration inferences should be treated as preliminary. The between-group SDCA difference persisted across all seven sensitivity thresholds, and the continuous analysis confirmed a direct association between AIT scores and the DBA–SDCA gap ($\rho = -0.316$; $p < 0.001$), indicating that the relationship is robust rather than an artifact of dichotomization.

The distrustful group's pattern of self-reporting one driving style while expressing a preference for a more conservative autonomous equivalent is thematically consistent with Basu et al. [23], who found in a simulator that participants' revealed preferences were more conservative than their self-reported preferences. While that study captured a stated-versus-revealed gap, the present study documents a comparable directional pattern across two stated measures (self-reported driving style and stated SDC preference). This pattern is also consistent with findings that driver–AV style congruence increases trust and reduces takeover interventions [65] and that initial trust level moderates how users calibrate to different AV driving styles [66]. The present cross-sectional design cannot distinguish whether high trust causes acceptance of driving styles similar to one's own or whether familiarity with one's own driving style in an autonomous context builds trust; disentangling this bidirectional relationship would require longitudinal or experimental designs.

These results extend the evidence on the trust–aggressiveness preference relationship specifically, which has primarily been examined within single-culture samples, to a multi-country sample spanning the US, Germany, and Panama, despite significant baseline differences in driving aggressiveness and driving mechanics trust between Panama and the two Western countries. The multivariate regression results showing that gender significantly predicts AI trust (with males reporting higher trust) are consistent with findings that emotional responses mediate gender differences in willingness to use automated cars [67] and that males express less concern about autonomous vehicles [68]. Age, however, was not a significant predictor of any scale in the present regression models, which may reflect the relatively young and narrow age range of our sample (overall $M = 28.8$; $SD = 9.1$).

### *4.3. Engineering Implications*

These findings suggest potential directions for the design of adaptive autonomous driving systems, though we note that all inferences are based on stated survey preferences rather than observed in-vehicle behavior and would require experimental validation before implementation. First, SDC manufacturers deploying vehicles across international markets should consider calibrating default driving profiles to regional driving mechanics trust (AIDMT) baselines. For markets with lower mean AIDMT scores (such as Panama in this study), more conservative default driving parameters (lower speeds, larger following distances, and less frequent lane changes) may improve initial acceptance and trust formation. Second, the observed association between AIT scores and SDCA preferences suggests that onboard trust assessment could serve as an input for adaptive driving mode systems [14,15]: passengers with higher measured trust could be offered driving profiles closer to human-normative behavior, while those with lower trust would experience more conservative defaults. Third, for US and German respondents, AIDMT scores were significantly higher than AIT scores, suggesting that these passengers may trust an SDC more with specific highway driving mechanics (e.g., speed control and lane changing) than with the complex scenarios captured by general AI trust (e.g., navigating construction zones and operating without a human present). This pattern did not hold for Panama, where no significant AIT–AIDMT gap was observed, suggesting similar levels of general and task-specific trust within this group. For markets exhibiting this pattern, trust-building efforts may need to address both dimensions simultaneously rather than relying on task-specific competence demonstrations as an entry point for broader trust formation.

In practice, the scales developed in this study could inform a parameterized driving controller by mapping regional SDCA baselines to vehicle behavior parameters such as target following headway, maximum longitudinal acceleration, and lane-change initiation thresholds. A market where respondents preferred conservative SDC behavior (low SDCA, as observed in Panama) would receive defaults with larger following headways, lower

acceleration limits, and more conservative gap acceptance criteria. At the individual level, a short onboard trust questionnaire derived from the AIT items could classify a passenger's trust profile and select from a library of pre-validated driving profiles, similar to existing comfort, normal, and sport driving modes but informed by empirically measured trust–aggressiveness relationships rather than arbitrary presets.

### *4.4. Contextual Factors*

Notably, the AIT scale showed no significant between-country differences for any pair (all $p > 0.05$; see Table 7), while all four other scales differed significantly between Panama and both Western countries. This suggests that general AI trust may be more consistent across these cultures than driving-specific trust and behavior and that, at least among the three populations studied here, cross-cultural variation in AV acceptance may be driven more by task-specific expectations and driving norms than by baseline attitudes toward AI. The absence of significant differences between the US and Germany on all five scales is consistent with their similar road quality indices, comparable digital adoption levels, and shared regulatory maturity for automated vehicles. However, these US–Germany comparisons were underpowered (all powers $<0.26$), so the null results should not be interpreted as evidence of true equivalence; adequately powered replication studies are needed to determine whether the apparent US–Germany similarity is genuine.

Beyond these factors, there are several other contextual factors that may contribute to the observed differences. Legal liability frameworks for autonomous vehicles differ substantially across countries; for example, Germany is among the early markets where Level 3 systems have received regulatory approval [2], whereas many emerging markets have yet to establish dedicated regulatory frameworks for autonomous vehicles. Such regulatory differences may influence public perception of safety guarantees. Edelmann et al. [69] found that cultural context significantly shapes acceptance of automated vehicle decisions across Germany, the US, Japan, and China, which further supports the idea that regulatory and cultural environments jointly affect AV trust. Additionally, differences in urbanization patterns, traffic density, and road infrastructure design (e.g., highway interchange complexity and pedestrian zone prevalence) may shape expectations about what autonomous vehicles should prioritize. The significantly lower DSS scores among Panamanian respondents (where lower scores indicate safer, more cautious driving actions) indicate that Panamanian respondents selected more cautious options on the self-reported safety scale, though whether this reflects actual driving practices, adaptation to road conditions [39], cultural norms, or social desirability bias in self-reports cannot be determined from the present data.

### *4.5. Limitations*

Several limitations of this study should be acknowledged. First, the per-group sample sizes (US: $n = 50$; Panama: $n = 41$; Germany: $n = 66$) are adequate for single-group CFA with six to seven indicators under WLSMV estimation [70] but are below the ideal minimum for multi-group CFA. The Panama subsample ($n = 41$) is particularly small, and results involving this group should be interpreted with caution. Post hoc power analysis revealed that several comparisons, particularly those between the US and Germany, were underpowered, suggesting that non-significant differences between these groups should not be interpreted as evidence of equivalence. Second, both recruitment methods represent convenience sampling and are not nationally representative. PollPool respondents may skew toward younger academics, while snowball sampling introduces network-based homogeneity.

Third, the DSS scale achieved only partial metric invariance (with item DSS6 loading freed across groups), indicating that cross-cultural comparisons on this specific scale should

be interpreted with this caveat in mind, though the majority of DSS loadings (six out of seven) were invariant. Fourth, the study relies on self-reported driving behaviors and stated preferences rather than observed behavior, which may introduce social desirability bias; moreover, participants presumably had no firsthand experience riding in an autonomous vehicle, so all SDC preferences are hypothetical. Fifth, the survey was administered in English (US), German (Germany), and Spanish (Panama); translations were machine-translated and reviewed by native speakers, but formal back-translation was not conducted. Although the measurement invariance results (Section 3.9) provide empirical support for equivalent statistical functioning across language versions, measurement invariance cannot fully substitute for back-translation in establishing semantic equivalence, and subtle differences in item interpretation across languages cannot be fully ruled out. Sixth, driving experience (years of licensure and annual mileage) was not collected and could not be included as a covariate. Driving experience may independently influence self-reported driving behavior; for example, a less experienced driver may exhibit different DBA patterns than a more seasoned driver regardless of cultural background. This omission represents a potential confound that is only partially mitigated by the inclusion of age as a covariate in the regression models. Seventh, the TOST equivalence margin ($\pm 0.125$) was derived from the distrustful group's observed pseudo-median in the same dataset; while the resulting 90% CI ($[-0.036, 0.036]$) is well within the bounds, replication with a pre-specified, theory-based margin would strengthen the equivalence conclusion. Eighth, sparse response categories were collapsed for some scales (DSS, SDCA, and DBA) to enable convergence in multi-group CFA; this standard practical step may affect threshold comparability across groups.

Finally, the cross-sectional design precludes causal inferences about the relationship between trust and driving style preferences. The CFA models used to validate scale unidimensionality were fitted on the same sample used for subsequent hypothesis testing; an independent validation sample was not available given the total $N = 157$. This is standard practice in small-sample cross-cultural survey research but means that the scale structure and hypothesis tests are not independently validated. As an exploratory study, the statistical tests reported herein should be treated as hypothesis-generating; the Bonferroni corrections were applied conservatively to reduce false positives but do not substitute for pre-registered confirmatory testing. The three countries sampled (US, Germany, and Panama) represent a narrow slice of global cultural diversity; findings should not be generalized beyond these specific populations without further replication. Findings warrant replication using larger, probability-based samples across additional cultural contexts.

## 5. Conclusions

Regarding RQ1, drivers internationally prefer SDC behavior that is more conservative than their own, though this preference reached significance only in Germany and Panama, not in the US. Regarding RQ2, AI trust level is consistently associated with the DBA–SDCA gap across all tested thresholds and in continuous analysis, with trustful individuals showing DBA–SDCA equivalence (TOST $p < 0.001$). Regarding RQ3, significant cross-cultural differences exist on four out of five scales (DBA, SDCA, AIDMT, and DSS) between Panama and both Western countries but not between the US and Germany; notably, general AI trust (AIT) did not differ significantly for any country pair.

At an international level, comparing the three countries' survey results combined, this analysis found that drivers with higher trust in AI technologies (based on their AIT scores) showed DBA–SDCA equivalence (TOST $p < 0.001$), consistent with acceptance of a driving style comparable to their own, while drivers who were distrustful of AI technologies preferred a self-driving car that was more conservative than their own driving style.

When comparing individual countries, further observations emerge. Notably, Panamanian respondents had the lowest average SDCA score, indicating a preference for a more conservative SDC experience compared with US respondents ($p = 0.004 < 0.0167$), with a significant difference also observed relative to German respondents ($p = 0.007 < 0.0167$). When comparing Panamanian respondents to German respondents, statistical differences were found in four of the five quantitative measurements, suggesting that technology design or deployment strategy may need to be adapted for populations with trust and driving behavior profiles similar to those observed in Panama to improve social acceptability of SDCs. However, the SDCA country effect was attenuated after controlling for demographics ($p = 0.079$), suggesting that SDC aggressiveness preferences may be influenced by both cultural and demographic factors.

Future work could include expanding the sample size and increasing the number of nations surveyed to better understand the global needs of SDC technology. As autonomous vehicles continue to be deployed [71] and mixed traffic scenarios become more common [72], public trust is likely to evolve, warranting periodic reassessment of these metrics across diverse populations.

Based on the findings of this exploratory study, we suggest the following directions for SDC manufacturers and policymakers:

1. Default driving profiles could be calibrated to regional driving mechanics trust (AIDMT) baselines and aggressiveness preferences, with more conservative defaults for markets with lower measured AIDMT scores.
2. Adaptive driving systems may benefit from incorporating real-time trust assessment to personalize the driving experience.
3. Since US and German respondents trusted specific driving mechanics (AIDMT) more than general AI scenarios (AIT), trust-building efforts for these markets may benefit from addressing the specific low-trust AIT scenarios (e.g., construction zone navigation and autonomous operation without a human present) rather than relying solely on demonstrations of driving mechanics competence as an entry point for broader trust formation.
4. Given the significant cross-cultural differences observed in four out of five scales despite similar general AI trust levels, cross-cultural validation of trust metrics should be considered as a component of international SDC deployment.
5. For populations exhibiting significantly lower driving mechanics trust (AIDMT) and aggressiveness scores (such as Panamanian respondents in this study), longitudinal research should investigate how trust evolves with exposure to autonomous vehicle technology and whether incremental introduction of autonomous features can build trust over time.

**Author Contributions:** Conceptualization, M.N.; methodology, S.T. and M.N.; software, S.T.; validation, S.T.; formal analysis, S.T.; investigation, S.T. and M.N.; resources, M.N.; data curation, S.T.; writing—original draft preparation, S.T. and M.N.; writing—review and editing, S.T. and M.N.; visualization, S.T.; supervision, M.N.; project administration, M.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the Florida Atlantic University Social (protocol code: IRBNet ID 1845471-1 and date of approval: 6 December 2021).

**Informed Consent Statement:** Informed consent was obtained from all individual participants included in the study.

**Data Availability Statement:** All source data used to generate the results of this analysis, the survey templates containing the 57 questions asked in each respective language, and all analysis code are freely available on Zenodo [42].

**Acknowledgments:** We thank the anonymous reviewers for their inspiring and constructive feedback. During the preparation of this manuscript, the author(s) used Anthropic Claude (Claude Opus 4.6, 2026) for coding assistance, drafting, editorial refinement, and verification of statistical reporting consistency. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AI | Artificial Intelligence |
| AIDMT | Artificial Intelligence Driving Mechanics Trust |
| AIT | Artificial Intelligence Trust |
| AV | Autonomous Vehicles |
| AVE | Average Variance Extracted |
| CFA | Confirmatory Factor Analysis |
| CFI | Comparative Fit Index |
| CI | Confidence Interval |
| DAI | Digital Adoption Index |
| DBA | Driving Behavior Aggressiveness |
| DSS | Driver Safety Score |
| HL | Hodges–Lehmann |
| HTMT | Heterotrait–Monotrait Ratio |
| IRB | Institutional Review Board |
| OLS | Ordinary Least Squares |
| RMSEA | Root Mean Square Error of Approximation |
| SAE | Society of Automotive Engineers |
| SD | Standard Deviation |
| SDC | Self-Driving Car |
| SDCA | Self-Driving Car Aggressiveness |
| SE | Standard Error |
| TLI | Tucker–Lewis Index |
| TOST | Two One-Sided Test |
| WLSMV | Weighted Least Squares Mean and Variance Adjusted |

## References

1. SAE International. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*; Technical Report J3016_202104; SAE International: Warrendale, PA, USA, 2021. [CrossRef]
2. Nature Machine Intelligence. Safe Driving Cars. *Nat. Mach. Intell.* **2022**, *4*, 95–96. [CrossRef]
3. Kusano, K.D.; Scanlon, J.M.; Chen, Y.H.; McMurry, T.L.; Gode, T.; Victor, T. Comparison of Waymo Rider-Only Crash Rates by Crash Type to Human Benchmarks at 56.7 Million Miles. *Traffic Inj. Prev.* **2025**, *26*, S8–S20. [CrossRef] [PubMed]
4. Teale, C. Amid Tesla's Autopilot Probe, Nearly Half the Public Thinks Autonomous Vehicles Are Less Safe Than Normal Cars. 2021. Available online: https://web.archive.org/web/2023/https://morningconsult.com/2021/09/02/autonomous-vehicles-safety-consumer-interest-polling/ (accessed on 15 January 2022).
5. American Automobile Association. AAA: Fear in Self-Driving Vehicles Persists. 2025. Available online: https://newsroom.aaa.com/2025/02/aaa-fear-in-self-driving-vehicles-persists/ (accessed on 1 March 2026).
6. Lee, J.D.; See, K.A. Trust in automation: Designing for appropriate reliance. *Hum. Factors* **2004**, *46*, 50–80. [CrossRef]
7. Hoff, K.A.; Bashir, M. Trust in automation: Integrating empirical evidence on factors that influence trust. *Hum. Factors* **2015**, *57*, 407–434. [CrossRef]

8. Lee, J.; Lee, D.; Park, Y.; Lee, S.; Ha, T. Autonomous vehicles can be shared, but a feeling of ownership is important: Examination of the influential factors for intention to use autonomous vehicles. *Transp. Res. Part C Emerg. Technol.* **2019**, *107*, 411–422. [CrossRef]
9. Choi, J.K.; Ji, Y.G. Investigating the importance of trust on adopting an autonomous vehicle. *Int. J. Hum.-Comput. Interact.* **2015**, *31*, 692–702. [CrossRef]
10. Gangadharaiah, R.; Mims, L.; Jia, Y.; Brooks, J.O. Opinions from Users Across the Lifespan about Fully Autonomous and Rideshare Vehicles with Associated Features. *SAE Int. J. Adv. Curr. Pract. Mobil.* **2024**, *6*, 309–323. [CrossRef]
11. Parasuraman, R.; Riley, V. Humans and automation: Use, misuse, disuse, abuse. *Hum. Factors* **1997**, *39*, 230–253. [CrossRef]
12. Bonnefon, J.F.; Shariff, A.; Rahwan, I. The social dilemma of autonomous vehicles. *Science* **2016**, *352*, 1573–1576. [CrossRef] [PubMed]
13. Shariff, A.; Bonnefon, J.F.; Rahwan, I. Psychological roadblocks to the adoption of self-driving vehicles. *Nat. Hum. Behav.* **2017**, *1*, 694. [CrossRef]
14. Nojoumian, M. Adaptive Mood Control in Semi or Fully Autonomous Vehicles. U.S. Patent 10,981,563, 20 April 2021.
15. Nojoumian, M. Adaptive Driving Mode in Semi or Fully Autonomous Vehicles. U.S. Patent 11,221,623, 11 January 2022.
16. Hartwich, F.; Hollander, C.; Johannmeyer, D.; Krems, J.F. Improving passenger experience and trust in automated vehicles through user-adaptive HMIs: "The more the better" does not apply to everyone. *Front. Hum. Dyn.* **2021**, *3*, 38. [CrossRef]
17. Hartwich, F.; Witzlack, C.; Beggiato, M.; Krems, J.F. The first impression counts–A combined driving simulator and test track study on the development of trust and acceptance of highly automated driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *65*, 522–535. [CrossRef]
18. Shahrdar, S.; Park, C.; Nojoumian, M. Human Trust measurement using an immersive virtual reality autonomous vehicle simulator. In *Proceedings of the 2nd AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)*; ACM: New York, NY, USA, 2019; pp. 515–520. [CrossRef]
19. Cegarra, J.; Unrein, H.; Andre, J.M.; Mouton, O.; Navarro, J. Driving among autonomous vehicles: The effect of initial trust and driving style on driving behaviors. *Transp. Res. Part F Traffic Psychol. Behav.* **2025**, *112*, 99–110. [CrossRef]
20. Ittner, S.; Mühlbacher, D.; Weisswange, T.H. The discomfort of riding shotgun–Why many people don't like to be co-driver. *Front. Psychol.* **2020**, *11*, 584309. [CrossRef] [PubMed]
21. Mühl, K.; Strauch, C.; Grabmaier, C.; Reithinger, S.; Huckauf, A.; Baumann, M. Get ready for being chauffeured: Passenger's preferences and trust while being driven by human and automation. *Hum. Factors* **2020**, *62*, 1322–1338. [CrossRef] [PubMed]
22. Kolekar, S.; de Winter, J.; Abbink, D. Human-like driving behaviour emerges from a risk-based driver model. *Nat. Commun.* **2020**, *11*, 4850. [CrossRef]
23. Basu, C.; Yang, Q.; Hungerman, D.; Singhal, M.; Dragan, A.D. Do you want your autonomous car to drive like you? In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 417–425. [CrossRef]
24. Hajiseyedjavadi, F.; Boer, E.R.; Romano, R.; Paschalidis, E.; Wei, C.; Solernou, A.; Forster, D.; Merat, N. Effect of environmental factors and individual differences on subjective evaluation of human-like and conventional automated vehicle controllers. *Transp. Res. Part F Traffic Psychol. Behav.* **2022**, *90*, 1–14. [CrossRef]
25. Dettmann, A.; Hartwich, F.; Rossner, P.; Beggiato, M.; Felbel, K.; Krems, J.; Bullinger-Hoffmann, A. Comfort or Not? Automated Driving Style and User Characteristics Causing Human Discomfort in Automated Driving. *Int. J. Hum.-Comput. Interact.* **2021**, *37*, 331–339. [CrossRef]
26. Schlüter, J.; Hellmann, M.; Weyer, J. Identifikation von Fahrertypen im Kontext des automatisierten Fahrens. *Forsch. Ingenieurwesen* **2021**, *85*, 945–955. [CrossRef]
27. Bellem, H.; Thiel, B.; Schrauf, M.; Krems, J.F. Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, *55*, 90–100. [CrossRef]
28. Craig, J.; Nojoumian, M. Should Self-Driving Cars Mimic Human Driving Behaviors? In *Proceedings of the 3rd International Conference on HCI in Mobility, Transport and Automotive Systems (MobiTAS)*; Springer: Berlin/Heidelberg, Germany, 2021; LNCS 12791; pp. 213–225. [CrossRef]
29. Park, C.; Nojoumian, M. Social Acceptability of Autonomous Vehicles: Unveiling Correlation of Passenger Trust and Emotional Response. In *Proceedings of the 4th International Conference on HCI in Mobility, Transport and Automotive Systems (MobiTAS)*; Springer: Berlin/Heidelberg, Germany, 2022; LNCS 13335; pp. 402–415. [CrossRef]
30. Park, C.; Shahrdar, S.; Nojoumian, M. EEG-based classification of emotional state using an autonomous vehicle simulator. In *Proceedings of the 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*; IEEE: New York, NY, USA, 2018; pp. 297–300. [CrossRef]
31. Awad, E.; Levine, S.; Kleiman-Weiner, M.; Dsouza, S.; Tenenbaum, J.; Shariff, A.; Bonnefon, J.F.; Rahwan, I. Drivers are blamed more than their automated cars when both make mistakes. *Nat. Hum. Behav.* **2020**, *4*, 134–143. [CrossRef]

32. Deloitte. 2020 Global Automotive Consumer Survey. 2020. Available online: https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/consumer-industrial-products/Global_Automotive_Consumer_Study_EN_Web_AODA.pdf (accessed on 15 August 2022).
33. Kyriakidis, M.; Happee, R.; de Winter, J.C.F. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transp. Res. Part F Traffic Psychol. Behav.* **2015**, *32*, 127–140. [CrossRef]
34. Muzammel, C.S.; Spichkova, M.; Harland, J. Cultural Influence on Autonomous Vehicles Acceptance. In *Proceedings of the Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous 2023)*; Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering; Springer: Berlin/Heidelberg, Germany, 2024; Volume 594. [CrossRef]
35. de Oña, R.; Garach, L.; de Oña, J. Social acceptance of autonomous vehicles: A cross-country model validation. *Transp. Res. Part F Traffic Psychol. Behav.* **2025**, *115*, 103329. [CrossRef]
36. Yang, Y.; Peng, L.; Wan, D. A Comparative Study on the Acceptance of Autonomous Driving Technology by China and Europe: A Cross-Cultural Empirical Analysis Based on the Technology Acceptance Model. *World Electr. Veh. J.* **2025**, *16*, 589. [CrossRef]
37. Tolbert, S.; Nojoumian, M. Cross-Cultural Expectations from Self-Driving Cars. In *Proceedings of the Human Factors in Design, Engineering, and Computing, Honolulu, HI, USA, 20–22 July 2025*; Ahram, T., Karwowski, W., Kalra, J., Eds.; AHFE Open Access: New York, NY, USA, 2025; Volume 199, pp. 137–146. [CrossRef]
38. Carlier, M. New Vehicle Sales in Central America by Country. 2021. Available online: https://www.statista.com/statistics/892278/central-america-new-vehicle-sales-registration-number/ (accessed on 15 August 2022).
39. Martinez, S.; Sanchez, R.; Yanez-Pagans, P. Road safety: Challenges and opportunities in Latin America and the Caribbean. *Lat. Am. Econ. Rev.* **2019**, *28*, 17. [CrossRef]
40. Du, N.; Robert, L.P.; Yang, X.J. Cross-cultural investigation of the effects of explanations on drivers' trust, preference, and anxiety in highly automated vehicles. *Transp. Res. Rec.* **2023**, *2677*, 554–561. [CrossRef]
41. Marroquin, A.; Sadd, L.; Saravia, A. Trust and perceptions of autonomous vehicles in Latin America. *Econ. Bull.* **2021**, *41*, 1461–1470.
42. Tolbert, S.; Nojoumian, M. *Analysis of Cross-Cultural Trust and Vehicle Operation Metrics for Self-Driving Cars Datasets*; Zenodo: Geneva, Switzerland, 2026. Available online: https://zenodo.org/records/18458900 (accessed on 15 August 2022). [CrossRef]
43. Flora, D.B.; Curran, P.J. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Methods* **2004**, *9*, 466–491. [CrossRef] [PubMed]
44. Kenny, D.A.; Kaniskan, B.; McCoach, D.B. The Performance of RMSEA in Models with Small Degrees of Freedom. *Sociol. Methods Res.* **2015**, *44*, 486–507. [CrossRef]
45. Hu, L.t.; Bentler, P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. A Multidiscip. J.* **1999**, *6*, 1–55. [CrossRef]
46. Sijtsma, K. On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika* **2009**, *74*, 107–120. [CrossRef]
47. Ten Berge, J.M.F.; Sočan, G. The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika* **2004**, *69*, 613–625. [CrossRef]
48. Flora, D.B. Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates. *Adv. Methods Pract. Psychol. Sci.* **2020**, *3*, 484–501. [CrossRef]
49. McDonald, R.P. *Test Theory: A Unified Treatment*; Psychology Press: Hove, UK, 2013.
50. Chen, F.F. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Model. Multidiscip. J.* **2007**, *14*, 464–504. [CrossRef]
51. Putnick, D.L.; Bornstein, M.H. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Dev. Rev.* **2016**, *41*, 71–90. [CrossRef]
52. Chien, S.Y.; Lewis, M.; Hergeth, S.; Semnani-Azad, Z.; Sycara, K. Cross-Country Validation of a Cultural Scale in Measuring Trust in Automation. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2015**, *59*, 686–690. [CrossRef]
53. Li, C.H. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* **2016**, *48*, 936–949. [CrossRef] [PubMed]
54. Mayer, R.C.; Davis, J.H.; Schoorman, F.D. An integrative model of organizational trust. *Acad. Manag. Rev.* **1995**, *20*, 709–734. [CrossRef]
55. Henseler, J.; Ringle, C.M.; Sarstedt, M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* **2015**, *43*, 115–135. [CrossRef]
56. Lakens, D. Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Soc. Psychol. Personal. Sci.* **2017**, *8*, 355–362. [CrossRef]
57. Norman, G. Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ.* **2010**, *15*, 625–632. [CrossRef] [PubMed]

58. Carifio, J.; Perla, R. Resolving the 50-year debate around using and misusing Likert scales. *Med. Educ.* **2008**, *42*, 1150–1152. [CrossRef]
59. Dunn, O.J. Multiple Comparisons Among Means. *J. Am. Stat. Assoc.* **1961**, *56*, 52–64. [CrossRef]
60. Hodges, J.L., Jr.; Lehmann, E.L. Estimates of Location Based on Rank Tests. *Ann. Math. Stat.* **1963**, *34*, 598–611. [CrossRef]
61. Cliff, N. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychol. Bull.* **1993**, *114*, 494–509. [CrossRef]
62. Byrne, B.M.; Shavelson, R.J.; Muthén, B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychol. Bull.* **1989**, *105*, 456–466. [CrossRef]
63. International Bank for Reconstruction and Development. *World Development Report 2016: Digital Dividends*; World Bank: Washington, DC, USA, 2016. [CrossRef]
64. Schwab, K. The Global Competitiveness Report 2019. 2019. Available online: http://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf (accessed on 15 August 2022).
65. Ma, Z.; Zhang, Y. Drivers' trust, acceptance, and takeover behaviors in fully automated vehicles: Effects of automated driving styles and driver's driving styles. *Accid. Anal. Prev.* **2021**, *159*, 106238. [CrossRef]
66. Manchon, J.B.; Bueno, M.; Navarro, J. Calibration of Trust in Automated Driving: A Matter of Initial Level of Trust and Automated Driving Style? *Hum. Factors* **2023**, *65*, 1613–1629. [CrossRef] [PubMed]
67. Hohenberger, C.; Sporrle, M.; Welpe, I.M. How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups. *Transp. Res. Part A Policy Pract.* **2016**, *94*, 374–385. [CrossRef]
68. Charness, N.; Yoon, J.S.; Souders, D.; Stothart, C.; Yehnert, C. Predictors of attitudes toward autonomous vehicles: The roles of age, gender, prior knowledge, and personality. *Front. Psychol.* **2018**, *9*, 2589. [CrossRef] [PubMed]
69. Edelmann, A.; Stumper, S.; Petzoldt, T. Cross-cultural differences in the acceptance of decisions of automated vehicles. *Appl. Ergon.* **2021**, *92*, 103346. [CrossRef]
70. Wolf, E.J.; Harrington, K.M.; Clark, S.L.; Miller, M.W. Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educ. Psychol. Meas.* **2013**, *73*, 913–934. [CrossRef] [PubMed]
71. Nidamanuri, J.; Nibhanupudi, C.; Assfalg, R.; Venkataraman, H. A progressive review: Emerging technologies for ADAS driven solutions. *IEEE Trans. Intell. Veh.* **2022**, *7*, 326–341. [CrossRef]
72. Ye, L.; Yamamoto, T. Evaluating the impact of connected and autonomous vehicles on traffic safety. *Phys. A Stat. Mech. Appl.* **2019**, *526*, 121009. [CrossRef]