Hindawi
Human Behavior and Emerging Technologies
Volume 2023, Article ID 9009791, 14 pages
https://doi.org/10.1155/2023/9009791

Research Article

Impact of Monitoring Requests on Publics' Assignment of Blame and Praise towards ADS Level 3 Vehicles

Liam Kettle, Madeleine M. McCarty, Kassidy L. Simpson, and Yi-Ching Lee

Psychology Department, George Mason University, Fairfax, Virginia, USA

Correspondence should be addressed to Liam Kettle; lkettle@gmu.edu

Received 27 April 2023; Revised 12 October 2023; Accepted 6 November 2023; Published 17 November 2023

Academic Editor: Mirko Duradoni

As vehicle automation capabilities increase, driving control shifts from the human to the vehicle system. However, concerns arise regarding responsibility following critical events and the publics' trust and acceptance of vehicles equipped with automated driving systems (ADS). The current study examined how participants assigned blame and praise to ADS-equipped vehicles and human drivers following collisions or near-misses and how these attributions were impacted by a virtual driving assistant that administered monitoring requests. Based on literature, our primary hypothesis was that more blame would be assigned to the human and more praise assigned to the ADS when the driving assistant was present. Additionally, we hypothesized greater reported trust towards ADS-equipped vehicles when the driving assistant was present. Participants read vignettes of automated driving, watched corresponding videos, and then self-reported trust, acceptance, anthropomorphism, and assignment of blame and praise. All hypotheses were supported indicating that significant effects were observed: participants assigned greater blame to the human when asked to actively monitor the driving environment and assigned greater praise to the ADS when it alerted the human driver. Additionally, participants reported greater trust and anthropomorphism of the ADS when the driving assistant was present. These findings suggest that explicitly communicating monitoring responsibility through a driving assistant significantly impacts the publics' opinion of responsibility following critical events. These findings provide initial support for a solution to improve driver safety as well as policy implications regarding positive perceptions and the adoption of ADS-equipped vehicles.

1. Introduction

Advancements in vehicle technology by various car manufacturers, such as Ford, Audi, and Tesla, have led to an increase in the incorporation of automated and autonomous features [1]. However, there have been almost 1,000 collisions involving vehicles equipped with automated driving system (ADS) features since 2021 with 18 resulting in fatal outcomes [2]. The first reported fatality involving an ADS-equipped vehicle occurred in Arizona in 2018, whereby a level 5-operated vehicle (i.e., full self-driving ability) struck and killed a pedestrian, with both human error and systemic failure contributing to this incident. That is, the driver, who was distracted by their cell phone, was unaware that the ADS recognized but failed to identify a pedestrian in harm's way. After an extensive investigation, the vehicle manufacturer was not found legally responsible for the death of the pedestrian, even though the ADS failed to identify them as a pedestrian [3]. More recently, a driver of a level 2-operated ADS was charged with vehicle manslaughter after a collision in which the autopilot feature was engaged [4]. The driver was legally held responsible for the collision due to the excessive speed it was traveling with autopilot engaged. These deadly collisions caused by failed automation and human error contributes to the ongoing debate surrounding how much responsibility does the general public attribute to the driver or the ADS-equipped vehicle involved in collisions and near-misses. Additionally, one open question is the degree to which explicitly requesting that the human driver actively monitors the road environment influences attributed responsibility.

A monitoring request (MR) is a salient alert intended to reorient the human drivers' attention away from any non-driving-related activities back to the driving environment

when the ADS encounters an unexpected scenario or is susceptible to automation failure [5]. Lu et al. [5] discovered that drivers preferred to engage in MRs so that they had ample time to return their attention to the road prior to resuming manual control. Additionally, incorporating MRs leads to higher situation awareness, trust, and acceptance of ADS-equipped vehicles [5, 6]. Generally, MRs were issued 7–12 s before the potential system failure [5, 7, 8]. Indeed, studies showed that drivers prefer to have perceived control over the vehicle system rather than being forced to quickly regain manual control [9]. For example, drivers could be prompted to return their attention to the road by a MR when approaching a zebra crossing where pedestrian movement is unclear to the vehicle system. Should any pedestrian decide to run across the road, the driver would have more time to assess the situation and react appropriately. A successful response to an MR requires a driver to understand and rationalize the purpose of the prompt; otherwise, the driver may distrust or ignore advisory alerts [10].

Providing drivers with information about the vehicle's ADS status increases driver awareness [11]. Commonly, auditory, visual, and haptic alerts are used to quickly cue drivers about a critical environmental event [12]. However, these alerts can be ambiguous in meaning. For example, a beep in a car may convey several different types of warnings depending on the situation, vehicle, or manufacturer, and drivers may even interpret the same alert differently under different circumstances. In conjunction with MRs, virtual driving assistants (DA) may help alleviate confusion and misinterpretation of alerts. Virtual DAs communicate critical information about the surrounding driving scene through natural language (e.g., "road closure ahead" and "construction zone ahead") [13]. Communicating critical information via DAs can mitigate issues of high mental workload that the complex task of driving requires [14], thus allowing drivers to safely focus their attention where needed to the surrounding environment. Moreover, previous research has shown that drivers would prefer a virtual DA with a sociable, likable, competent, and assertive female voice [15–17].

Shifting towards responsibility, attribution is a cognitive process in which individuals identify the cause of events and the factors that lead to the events [18]. Typically, individuals attribute responsibility to the cause of events. For instance, backward-looking responsibility relates to events that have occurred in the past and involves attributing blame or praise to an agent, depending on the successful or failed outcome, respectively [19, 20]. Although some argue that machines cannot hold responsibility for event outcomes because they lack intentionality or controllability of the situation [21], people do attribute responsibility, in the forms of blame and praise, to nonhuman agents for both positive and negative outcomes [22–24].

Research relevant to human-machine teaming indicates that responsibility for an outcome is associated with an agents' intentionality to engage in behavior. When asked to explain why various agents engaged in certain behaviors, participants tended to use similar mental models to explain the behaviors for both human and machine agents [25], suggesting that individuals infer intentionality towards nonhuman agents. As individuals infer a reason behind intentional acts, if an agent is believed to act with intention, then they could be perceived as being more responsible for their actions [26, 27].

In the context of driving, research examining blame assignment following collisions utilized text vignettes outlining hypothetical situations involving a human driver and an ADS-equipped vehicle to determine whether participants blame the human drivers or the vehicle system [28]. Bennet et al. [29] found that as the levels of automation increased, participants placed less blame on the human driver and more blame on the vehicle system. Therefore, when the human is less in control of the vehicle, people tend to shift the blame towards the automation, the vehicle system, and the manufacturers [30, 31]. This shift in blame is attributed to the blame attribution asymmetry [28], in which individuals place more blame on automation and judge ADS-equipped vehicle collisions more harshly when compared to human-caused collisions.

On the other hand, Awad et al. [32] argue that the public is more inclined to place less blame on the ADS during an error when the human and the vehicle system have shared control of the vehicle. Additionally, differences in assigned blame could be influenced by the extent that individuals perceive the driving agents as engaging in intentional behaviors as individuals tend to perceive intentionality and intended responses differently between ADS-equipped vehicles and human drivers [33].

Similar to blame assignment, drivers can be attributed praise following hypothetical critical events. For instance, McManus and Rutchick [34] examined praise and blame attributions towards drivers following hypothetical text vignettes conveying moral dilemmas (i.e., variations of the trolley problem) involving ADS-equipped vehicles. Results found that greater driver control agency led to greater praise for positive consequences and blame for negative consequences. However, these hypothetical situations focused on whether drivers made selfish or selfless decisions across moral dilemmas rather than responsibility following typical collisions or near-misses. Additionally, blameworthiness and praiseworthiness were assessed on a single scale and were only assigned to the driver. More broadly in the field of human-robot interaction, previous literature suggests that highly anthropomorphized agents are associated with greater or equal praise than humans for successful outcomes and blamed less for negative outcomes [23, 35]. Anthropomorphism in this context relates to the attribution of human-like mental capacities of agency to nonhuman agents [36]. Thus, these results suggest that blame and praise should be assessed as separate constructs.

Additional to assessing blame and praise as separate constructs, it is important to understand how individuals attribute these constructs towards both human and vehicle system agents across positive and negative critical events involving ADS-equipped vehicles. Typically, blame is attributed towards agents following task failures (e.g., collisions), and praise is attributed towards agents following successful outcomes (e.g., successful near-miss). However, past

research shows inconsistent patterns of attribution of blame or praise to human and virtual agents in task failures or success. Awad et al. [32] observed less blame attributed to the ADS than the human during errors when both the human and vehicle system share control. Similarly, Bartneck et al. [35] observed less blame attributed to highly anthropomorphized agents in negative outcomes and greater or equal praise in successful outcomes than humans. In contrast, Copp et al. [30] highlighted increased blame towards automation, vehicle system, and manufacturers when humans have less control over the vehicle during negative outcomes. However, in cases of shared control of the vehicle, there lacks examination of the degree to which individuals hold positive perceptions of vehicle agents even in the case of negative events (i.e., collisions). Thus, it is important to understand how individuals attribute both blame and praise across collision (negative outcome) and near-miss (positive outcome) critical events.

To our knowledge, only one early study examined blame attribution in scenarios involving ADS-equipped vehicles and conventional driving using a driving simulation [37]. Although they found that both the low and high anthropomorphic ADS were blamed more for a collision than manual driving, the highly anthropomorphic ADS (characterized by name, gender, and voice features) was blamed less and trusted more than the low anthropomorphic ADS (no anthropomorphic features) following a collision. Waytz et al. [37] further state that if the ADS were able to avoid the collision, they would predict that participants' tendency to assign praise to the ADS would increase with higher anthropomorphic features.

Individuals' perceptions of ADS-equipped vehicles tend to influence their trust and overall willingness to adopt ADS-equipped vehicles. For instance, individuals who assign greater blame to the ADS following critical events tend to trust the ADS-equipped vehicles less [38]. Although little research has examined attributed praise towards ADS-equipped vehicles, research suggests that those who hold optimistic and more favorable perceptions towards the technology tend to have stronger intentions to use and adopt ADS-equipped vehicles [39, 40]. Having positive perceptions of ADS technology can benefit the intention to use and adoption of ADS-equipped vehicles especially in emerging countries where intelligent transportation systems are less available [41]. Similarly, familiarity and exposure to ADS-equipped vehicles tend to positively correlate with technological optimism and intention to use [39]. Additionally, perceiving the ADS as having greater agency tends to reduce attributed blame, fosters greater trust, and predicts individuals' willingness to ride or purchase ADS-equipped vehicles [37, 42].

1.1. Overall Objectives. The current study is aimed at building upon the previous research that utilized hypothetical text vignettes (i.e., [28–32, 34]) through the inclusion of corresponding video stimuli and the consideration of gender and voice features through the inclusion of a DA to mimic a highly anthropomorphic ADS-equipped vehicle. Therefore, the overall objective of the current study was to examine the impact of MRs, delivered through a DA, in collisions and near-miss situations on people's blame and praise attributions. Specifically, participants experienced scenarios involving a level 3-operated vehicle with either the DA present (DAP) or absent (DAA) in both collision and near-miss critical outcomes. The near-miss outcome was further separated into either the human or ADS agent appropriately responding in the corresponding scenario. Participants then assigned blame and praise attributions to both the human driver and the ADS agent. Additionally, participants self-reported their level of trust, acceptance, and perceived anthropomorphism of the ADS. As current ADS technology does not have the capacity to integrate DAs with MR functionality, a Wizard-of-Oz approach was used to simulate the scenarios.

Previous research indicated that as ADS-operated vehicles have greater control of the driving situation, individuals assign more blame to the ADS than to the human [28–31], yet this may not hold true during shared control [32]. However, higher anthropomorphic vehicles are blamed less and trusted more than lower anthropomorphic vehicles [37, 42]. Although there is a lack of research examining praise assignment in ADS-equipped vehicles, research related to human-machine teaming suggests that higher perceived anthropomorphism leads to greater praise assignment towards a nonhuman agent than a human agent [23, 35]. Based on the findings from these previous studies, the primary hypotheses relevant to blame and praise assignment include the following:

(1) For the collision condition, (a) there will be an interaction effect whereby more blame will be assigned to the human than to the ADS in the DAP condition than in the DAA condition, and (b) there will be an interaction effect whereby more praise will be assigned to the ADS than to the human in the DAP condition than in the DAA condition

(2) For the near-miss vehicle responding condition, (a) there will be an interaction effect whereby more blame will be assigned to the human than to the ADS in the DAP condition than in the DAA condition, and (b) there will be an interaction effect whereby more praise will be assigned to the ADS than to the human in the DAP condition than in the DAA condition

(3) For the near-miss human responding condition, (a) there will be an interaction effect whereby more blame will be assigned to the human than to the ADS in the DAP condition than in the DAA condition, and (b) there will be interaction effect whereby more praise will be assigned to the ADS than to the human in the DAP condition than in the DAA condition

The secondary hypotheses include the following:

(1) Trust towards the ADS-equipped vehicle is significantly higher in the DAP condition than in the DAA condition

(2) Perceived anthropomorphism of the ADS-equipped vehicle is significantly higher in the DAP condition than in the DAA condition

2. Method

2.1. Design. The current study used a 2 (autonomous feature) × 2 (event type) × 3 (outcome severity) design. As a between-subject variable, the autonomous feature included either the presence or absence of a DA. As another between-subject variable, the event type was either a vehicle or a pedestrian event. A between-subject design was used as the vehicles with and without the DA used identical videos with the only difference being the DA present. If participants experienced both DA conditions, they would realize it was the same video, possibly disrupting the experimental immersion. As a within-subject variable, the outcome severity included a collision or a near-miss which was further separated into either the human or ADS agent responding appropriately. The dependent variables included attributed blame towards the human or ADS-equipped vehicle, attributed praise towards the human or ADS-equipped vehicle and trust and acceptance in the ADS-equipped vehicle, and perceived anthropomorphism of the ADS-equipped vehicle.

2.2. Participants. In total, 525 participants were recruited. However, 26 were removed due to completion of the study within 8 minutes, which was considered unrealistic. Therefore, there were 499 eligible participants. Participants were, on average, 26.4 years of age (SD = 10.0) and had low experience with vehicle automation features ($M = 1.4$, SD = 0.9). A preliminary analysis was performed and found little to no differences in blame assignment between men and women [43]. Participants were expected to have working Internet and the ability to watch and listen to videos. Participants were recruited through either George Mason University Department of Psychology's Research Portal, SONA, or Amazon's Mechanical Turk (MTurk). Between the two methods of recruitment, 316 participants were recruited through SONA and were, on average, 20.6 years of age (SD = 3.4); 183 participants were recruited through MTurk and were, on average, 36.4 years of age (SD = 9.8). A t-test indicated significant differences in age between recruitment methods ($t(207.36) = 21.12$, $p < 0.005$). Regardless, a Pearson's chi-square test showed no significant differences for recruitment method between the four experimental groups ($\chi^2(3) = 1.48$, $p = 0.687$) indicating that each recruitment method was similarly represented in each group. Those recruited through SONA received course credit, while those recruited through MTurk received a $4 compensation upon eligible completion. This research complied with the American Psychology Association Code of Ethics and was approved by the Institutional Review Board at George Mason University (IRB# 1866476-1). Informed consent was obtained from each participant.

Key demographic data, separated by experimental group, and test on sample distribution across groups are shown in Table 1. Between the four experimental groups, a Kruskal-Wallis test showed no significant differences for age ($\chi^2(3) = 2.88$). Pearson's chi-square tests showed no significant differences for ethnicity ($\chi^2(18) = 12.82$), license type ($\chi^2(6) = 3.95$), or vehicle ownership ($\chi^2(3) = 6.37$). For gender, Pearson's chi-square test showed significant differences between groups ($\chi^2(12) = 29.84$, $p = 0.003$); however, this included 11 participants (2%) who self-identified as nonbinary/third gender, other, or preferred not to say, leading to a potentially unreliable chi-square test due to the small sample. Thus, an additional Pearson's chi-square test only including those who self-identified as man or woman was conducted leading to no significant differences between groups ($\chi^2(3, N = 488) = 1.80$).

2.3. Study Materials and Apparatus

2.3.1. Driving Video Stimuli. Video recordings were created from a driving simulation software, City Car Driving. A Wizard-of-Oz methodology was used to simulate a level 3-equipped vehicle outfitted with a DA that could issue monitoring requests. Videos were created for each event type and outcome severity conditions; however, the same videos were used for the DAA and DAP conditions with the only difference being the addition of voice clips for imitating the driving assistant in the DAP condition. In each video, the vehicle drove normally on the road following legal traffic rules without driver input until the critical event occurred. Voice and audio files were synchronized according to the driving environments. Monitoring requests occurred 10 seconds before the critical event whereby participants could see the event in the distance once the alert sounded. Each video showed a Tesla Model 3 vehicle simulated with level 3 features driving without the driver's hands in view and lasted approximately two minutes. A Tesla vehicle was displayed as the City Car Driving software simulates real-world vehicles, and we wanted to display a vehicle that could realistically be perceived as an ADS-equipped vehicle. An example image is shown in Figure 1. In total, there were 12 videos corresponding to their specific hypothetical vignette scenarios, six distinct videos each having a duplicate with the DA voice overlaid.

2.3.2. DA Voice Agent. The DA was created using Amazon Polly (https://aws.amazon.com/polly/) for text-to-speech generation. In particular, the DA used the US English, female, Joanna voice, which has been previously used in [16]. The DA informed drivers of monitoring requests and included other information such as the driving environment and vehicle actions. An example script is shown in Table 2.

2.3.3. Text Vignettes. Text vignettes were written for the 12 videos. Vignettes were similar across all conditions but differed by DA presence or absence, outcome severity, event type, and human or ADS agent responding to traffic (for near-misses). An example script for a collision with a vehicle in the DAP condition stated:

Imagine that you are watching the dashcam footage of an autonomous vehicle before a collision. The autonomous driving system can perform all aspects of driving tasks, including controlling speed, steering, and lane control. In addition, the autonomous vehicle is fitted with an AI driving assistant to communicate its actions while driving along a route. However, this system cannot handle all possible situations. The human driver is playing games on their phone. The vehicle's object identification system detects a potential

Table 1: Summary of participants' age group, gender, ethnicity, and driving-related variables further separated by experimental group.

Demographic variable	DAA-pedestrian (*N*)	DAA-vehicle (*N*)	DAP-pedestrian (*N*)	DAP-vehicle (*N*)	*p* value
Age					0.410
18-24	73	76	72	74	
25-34	25	30	34	29	
35-44	15	15	16	13	
45-54	6	4	1	2	
55-64	2	2	2	4	
65-74	1	0	1	2	
Gender					0.616[†]
Man	53	53	50	55	
Woman	66	71	73	67	
Nonbinary/third gender	2	3	2	0	
Prefer not to say	1	0	0	2	
Other	0	0	1	0	
Ethnicity[††]					0.802
American Indian/Alaska Native	1	3	3	2	
Asian	26	19	26	23	
Black	21	17	19	17	
Native Hawaiian/other Pacific Islander	1	1	1	0	
Hispanic/Latino/Spanish origin	15	24	12	13	
White	65	69	69	72	
Other	4	5	5	8	
Driver's license type					0.684
Full driver's license	104	111	108	108	
Learner's permit	12	12	13	13	
None	6	4	5	3	
Vehicle ownership					0.095
Yes	99	97	97	91	
No	23	30	29	33	

Note: [††]Participants could select multiple ethnicities. [†]Chi-square test only including those who self-identified as man or woman. Kruskal-Wallis's test was used for age. Chi-square test was used for ethnicity, gender, driver's license type, and vehicle ownership.

Figure 1: Example screen of the simulated video shown to participants.

vehicle collision in the distance blocking the lane ahead and alerts the driver with a Monitoring Request – asking the driver to momentarily monitor the environment and resume manual vehicle control if needed. The vehicle crashes into the other vehicles. You will now watch the dashcam footage for this scenario.

An example script for a near-miss with pedestrian jaywalking with the human appropriately responding in the DAP condition stated:

Imagine that you are watching the dashcam footage of an autonomous vehicle before a near miss. The autonomous driving system can perform all aspects of driving tasks, including controlling speed, steering, and lane control. In addition, the autonomous vehicle is fitted with an AI driving assistant to communicate its actions while driving along a route. However, this system cannot handle all possible situations. The human driver is playing games on their phone. The vehicle's object identification system detects erratic pedestrian movement and alerts the driver with a Monitoring Request – asking the driver to momentarily monitor the environment and resume manual vehicle control if needed. The pedestrian does not notice the vehicle approaching and illegally runs across the road. The human driver sees the pedestrian and responds appropriately leading to a successful near miss. You will now watch the dashcam footage for this scenario.

2.3.4. Demographic Questionnaire. Participants were asked demographic information including age, gender, ethnicity, and work status.

Table 2: Example DA scripts.

Event	Script
Vehicle collision ahead	"Warning! Potential vehicle collision ahead. Please monitor the environment and takeover if needed"
Potential illegal pedestrian jaywalker ahead	"Warning! Pedestrian movement unclear! Please monitor the environment and takeover if needed"
Approaching pedestrian crossing	"Checking pedestrian crossing"; "No pedestrians found"
Turning right at red traffic light	"I am checking for oncoming vehicles"; "No oncoming vehicles"; "I am now turning right"

2.3.5. Driving and Virtual Assistant Behaviors. Driving behaviors measured included driver's license type (full driver's license, license permit, and none), vehicle ownership, and experience using vehicle automation, including adaptive cruise control, lane keeping assist, blind spot warning, parking assist, forward collision warning, automatic emergency braking, and lane departure warning using a 4-point Likert-like scale (0 = none, 3 = high).

2.3.6. Trust. Participants' trust in the ADS was assessed using the six-item Situational Trust Scale for Automated Driving [44]. The items asked questions relevant to trust, performance, non-driving-related tasks, risk, judgment, and reaction. Participants rated their perception of the ADS on a 7-point Likert-like scale (1 = strongly disagree, 7 = strongly agree). All items were aggregated into one composite score. McDonald's Omega reliability for each condition score ranged from 0.56 to 0.79.

2.3.7. Acceptance. Acceptance of the ADS was assessed using the nine-item, two-dimensional scale developed by [45]. Items were scored on a 5-point semantic differential scale for two dimensions: usefulness (useful–useless, bad–good, effective–superfluous, assisting–worthless, and raising alertness–sleep-inducing) and satisfaction (pleasant–unpleasant, nice–annoying, irritating–likable, and undesirable–desirable). Each set of items was aggregated into respective overall scores. McDonald's Omega reliability for each condition ranged from 0.79 to 0.84 for usefulness and 0.81 to 0.90 for satisfaction.

2.3.8. Anthropomorphism. In the context of this study, anthropomorphism was defined as the attribution of human-like mental capacities of agency, and each item was derived from [37]. The items included "how well could the vehicle system plan a route," "how well could the vehicle system feel what was happening," "how well could the vehicle system anticipate what was about to happen," and "how smart was the vehicle system." Each item was rated from 0 (not at all) to 10 (very much). All items were aggregated into one overall score. Additionally, perceived anthropomorphism was used as an indirect manipulation check between the DAP and DAA conditions. McDonald's Omega reliability for each condition ranged from 0.88 to 0.93.

2.3.9. Blame and Praise Attributions. In the current literature, there does not seem to be a psychometrically validated scale for blame attribution towards ADS-equipped vehicles. Therefore, we adapted items used in previous research [28, 31, 46]. In particular, the survey included "to what extent do you think the [driver/autonomous vehicle system] should be blamed for the [collision/near-miss]?" on a 7-point Likert-like scale from 1 (no blame) to 7 (a lot of blame). Past research has yet to examine praise attributions towards ADS-equipped vehicles; therefore, we altered the latter question to assess praise, i.e., "to what extent do you think the [driver/autonomous vehicle system] should be praised for [collision/near-miss]?" on a 7-point Likert-like scale from 1 (no praise) to 7 (a lot of praise).

2.4. Procedure. Individuals were presented information about the study through either SONA or MTurk. If individuals chose to participate, they were directed to Qualtrics and asked to provide informed consent. Following consent, participants were asked to complete the pretest questionnaire (demographics, driving and virtual assistant behaviors). Before beginning the first experimental condition, participants were asked to verify that they could see and hear an initial video using the audio and visual verification process. Failure to verify resulted in immediate discontinuance of the study. Next, participants were shown an introductory driving simulation video indicating the type of video content to expect. Participants were then randomly assigned to one of four conditions: DAP-vehicle, DAP-pedestrian, DAA-vehicle, and DAA-pedestrian. For each condition, participants read the text vignette and watched the corresponding simulated video and then completed the posttest questionnaire (trust, acceptance, anthropomorphism, and blame and praise attributions). This process was repeated for each of the three outcome severity conditions. Each of these scenarios was randomly presented to participants to eliminate any priming effects. Before watching each video, they were reminded to keep their audio on and to watch the video as if they were watching a dashcam recording before a critical event. Upon completion, compensation or course credits were provided depending on the recruitment method. A diagram of the procedure and experimental design is shown in Figure 2.

2.5. Data Analysis. Self-reported scores for trust, usefulness, satisfaction, anthropomorphism, and blame and praise attributions were aggregated across their respective measures. Pearson's correlations were conducted to examine the relationship between age, trust, usefulness, satisfaction, anthropomorphism, blame attribution, praise attribution, and total experience with ADS features.

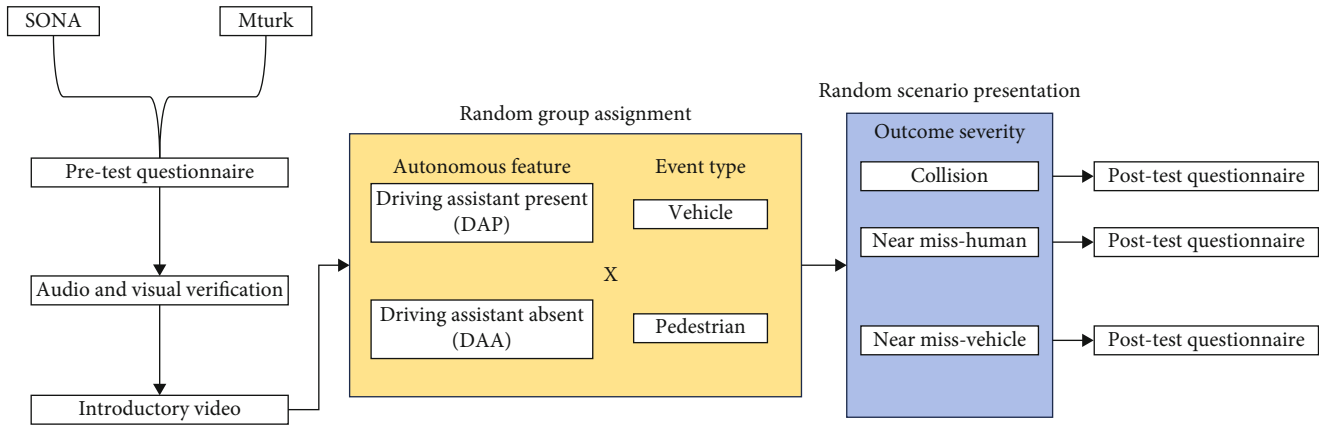

Figure 2: Procedure and experimental design.

Table 3: Correlations between variables of interest.

	Trust	Use	Sat	Anth	B (H)	B (ADS)	P (H)	P (ADS)	Exp
Age	0.10*	0.01	0.08*	0.14*	-0.06*	0.07*	**0.22** *	**0.20** *	0.11*
Trust	1.00	**0.64** *	**0.56** *	**0.72** *	**-0.23** *	**-0.43** *	**0.20** *	**0.60** *	0.06*
Use		1.00	**0.74** *	**0.62** *	-0.15*	**-0.47** *	0.06*	**0.38** *	0.00
Sat			1.00	**0.58** *	-0.13*	**-0.35** *	0.14*	**0.38** *	0.12*
Anth				1.00	-0.09*	**-0.32** *	**0.24** *	**0.60** *	0.10*
B (H)					1.00	**0.25** *	-0.13*	-0.12*	0.10*
B (ADS)						1.00	0.15*	-0.16*	0.11*
P (H)							1.00	**0.40** *	0.20*
P (ADS)								1.00	0.18*

Note: *$p < 0.05$. Bold indicates correlations > 0.2. Use = usefulness; Sat = satisfaction; Anth = anthropomorphism; B = blame; P = praise; H = human; ADS = automated driving system; Exp = experience.

For the main analysis, an initial $2 \times 2 \times 3$ mixed model ANOVA was conducted for both trust and anthropomorphism scores using DA, event type, and outcome severity as the independent variables. A 2×2 mixed model ANOVA was conducted for blame scores following each of the critical events and praise scores following each of the critical events using DA and driver agent (human or ADS agent) as the independent variables. Upon model significance, ad hoc pairwise comparisons using Bonferroni's corrections were conducted to determine differences between individual conditions.

Prior to the analyses, normality and homogeneity of variance assumptions were checked for each analysis. The Shapiro-Wilk test indicated that the data was nonnormal for all analyses. However, given the large sample size, it is generally accepted to proceed with analysis even with slightly nonnormal data. For the assumption of homogeneity of error variances, Levene's test was nonsignificant for the trust and anthropomorphism analyses, indicating that this assumption was satisfied. As both normality and homogeneity assumptions were violated for the blame and praise analyses, a two-way nonparametric robust standard error mixed ANOVA was performed with heterogeneity correction instead.

3. Results

3.1. Correlations. Pearson's correlations were conducted between continuous variables. Shown in Table 3, anthropomorphism was strongly associated positively with trust and was moderately correlated with higher perceived usefulness, satisfaction, and praise of the ADS. Higher self-reported trust was moderately correlated with higher perceived usefulness, satisfaction, and praise of the ADS and less blame to the ADS. Although there were significant correlations between age and the other variables, these were found to be very weak or weak associations. Similarly, weak or very weak associations were found between experience with vehicle automation and the other variables.

3.2. Blame

3.2.1. Collision. A two-way nonparametric robust standard error mixed ANOVA was performed to analyze the effect of the presence or absence of the DA and the driver agent (human or ADS) on blame assignment following a collision. Shown in Figure 3, the two-way ANOVA revealed a statistically significant interaction between the effects of DA presence and driver agent ($F(1, 497) = 48.47$, $p < 0.001$), in that there was no

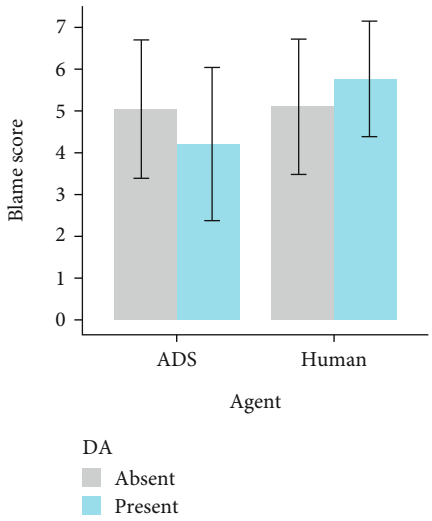

Figure 3: Blame assignment following a collision across driver agents and DA conditions. Note: error bars indicate standard deviation.

difference between assigned blame for the human ($M = 5.1$, SD = 1.6) and the ADS ($M = 5.0$, SD = 1.7) in the DAA condition, yet more blame was assigned to the human ($M = 5.8$, SD = 1.4) than the ADS ($M = 4.2$, SD = 1.8) in the DAP condition. No main effect was found for DA in that assigned blame did not significantly differ between the presence ($M = 5.0$, SD = 1.8) and absence ($M = 5.1$, SD = 1.6) of the DA ($p = 0.392$). However, there was a main effect for driver agent in that the human driver ($M = 5.4$, SD = 1.5) was assigned more blame overall than the ADS ($M = 4.6$, SD = 1.8) ($p < 0.001$).

3.2.2. Near-Miss Vehicle Responding. A two-way nonparametric robust standard error mixed ANOVA was performed to analyze the effect of the presence or absence of the DA and the driver agent (human or ADS) on blame assignment following a near-miss where the vehicle responded appropriately. Shown in Figure 4, the two-way ANOVA revealed a statistically significant interaction between the effects of DA presence and driver agent ($F(1, 497) = 37.16$, $p < 0.001$), in that there was a greater difference in assigned blame between the human ($M = 4.6$, SD = 1.9) and the ADS ($M = 3.2$, SD = 1.8) in the DAP condition than between the human ($M = 3.9$, SD = 1.9) and the ADS ($M = 3.7$, SD = 1.8) in the DAA condition. No main effect was found for DA in that assigned blame did not significantly differ between the presence ($M = 3.9$, SD = 2.0) and absence ($M = 3.8$, SD = 1.9) of the DA ($p = 0.584$). However, there was a main effect for driver agent in that the human driver ($M = 4.2$, SD = 1.9) was assigned more blame overall than the ADS ($M = 3.4$, SD = 1.9) ($p < 0.001$).

3.2.3. Near-Miss Human Responding. A two-way nonparametric robust standard error mixed ANOVA was performed to analyze the effect of the presence or absence of the DA and the driver agent (human or ADS) on blame assignment following a near-miss where the human responded appropriately. Shown in Figure 5, the two-way ANOVA revealed a statistically significant interaction between the effects of

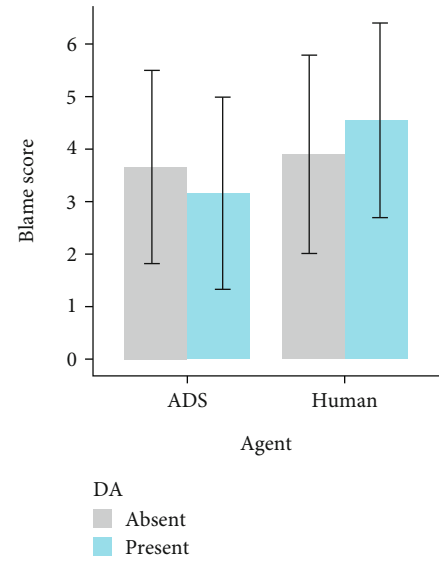

Figure 4: Blame assignment following a near-miss where the vehicle responded appropriately across driver agents and DA conditions. Note: error bars indicate standard deviation.

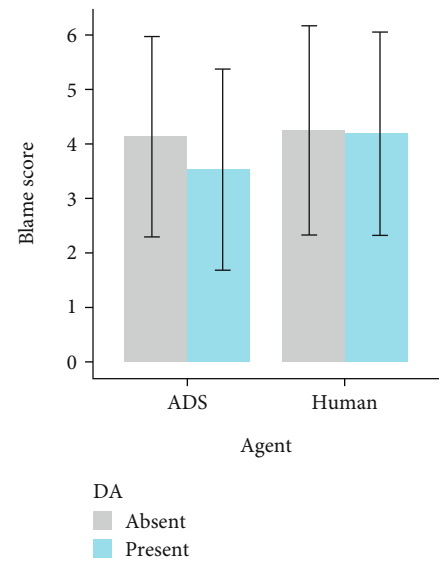

Figure 5: Blame assignment following a near-miss where the human responded appropriately across driver agents and DA conditions. Note: error bars indicate standard deviation.

DA presence and driver agent ($F(1, 497) = 7.73$, $p = 0.006$), in that there was a greater difference in assigned blame between the human ($M = 4.2$, SD = 1.9) and the ADS ($M = 3.5$, SD = 1.9) in the DAP condition than between the human ($M = 4.3$, SD = 1.9) and the ADS ($M = 4.1$, SD = 1.8) in the DAA condition. A main effect was identified for DA in that the presence ($M = 3.9$, SD = 1.9) of the DA was assigned less blame overall than the absence ($M = 4.2$, SD = 1.9) of the DA ($p = 0.015$). Additionally, there was a main effect for driver agent in that the human driver ($M = 4.2$, SD = 1.9) was assigned more blame overall than the ADS ($M = 3.9$, SD = 1.9) ($p < 0.001$).

3.3. Praise

3.3.1. Collision. A two-way nonparametric robust standard error mixed ANOVA was performed to analyze the effect of the presence or absence of the DA and the driver agent (human or ADS) on praise assignment following a collision. Shown in Figure 6, the two-way ANOVA revealed a statistically significant interaction between the effects of DA presence and driver agent ($F(1, 497) = 27.13$, $p < 0.001$), in that there was no difference between assigned praise for the human ($M = 2.4$, SD = 1.9) and the ADS ($M = 2.3$, SD = 1.9) in the DAA condition, yet more praise was assigned to the ADS ($M = 2.7$, SD = 1.9) than the human driver ($M = 2.2$, SD = 1.8) in the DAP condition. No main effect was found for DA as assigned praise did not significantly differ between the presence ($M = 2.5$, SD = 1.9) and absence ($M = 2.4$, SD = 1.9) of the DA ($p = 0.489$). A main effect was found for driver agent in that the ADS system ($M = 2.5$, SD = 1.9) was assigned more praise overall than the human driver ($M = 2.3$, SD = 1.9) ($p < 0.001$).

3.3.2. Near-Miss Vehicle Responding. A two-way nonparametric robust standard error mixed ANOVA was performed to analyze the effect of the presence or absence of the DA and the driver agent (human or ADS) on praise assignment following a near-miss where the vehicle responded appropriately. Shown in Figure 7, the two-way ANOVA revealed a statistically significant interaction between the effects of DA presence and driver agent ($F(1, 497) = 5.61$, $p = 0.018$), in that there was a greater difference in assigned praise between the ADS ($M = 5.0$, SD = 1.7) than the human ($M = 3.0$, SD = 1.8) in the DAP condition than between the ADS ($M = 4.7$, SD = 1.7) and the human ($M = 3.2$, SD = 1.8) in the DAA condition. No main effect was found for DA as assigned praise did not significantly differ between the presence ($M = 4.0$, SD = 2.0) and absence ($M = 3.9$, SD = 2.0) of the DA ($p = 0.630$). However, a main effect was found for driver agent in that the ADS system ($M = 4.8$, SD = 1.7) was assigned more praise overall than the human driver ($M = 3.1$, SD = 1.9) ($p < 0.001$).

3.3.3. Near-Miss Human Responding. A two-way nonparametric robust standard error mixed ANOVA was performed to analyze the effect of the presence or absence of the DA and the driver agent (human or ADS) on praise assignment following a near-miss where the human driver responded appropriately. Shown in Figure 8, the two-way ANOVA revealed a statistically significant interaction between the effects of DA presence and driver agent ($F(1, 497) = 29.87$, $p < 0.001$), in that more praise was assigned to the human ($M = 4.3$, SD = 1.9) than the ADS ($M = 3.6$, SD = 1.9) in the DAA condition, yet more praise was assigned to the ADS ($M = 4.5$, SD = 1.6) than the human driver ($M = 4.0$, SD = 1.8) in the DAP condition. A main effect was found for DA as the presence ($M = 4.3$, SD = 1.7) of the DA was assigned more praise overall than the absence ($M = 3.9$, SD = 1.9) of the DA ($p = 0.016$). However, no main effect was found for driver agent as assigned praise did not significantly differ between the ADS ($M = 4.0$, SD = 1.8) and the human driver ($M = 4.2$, SD = 1.9) ($p = 0.261$).

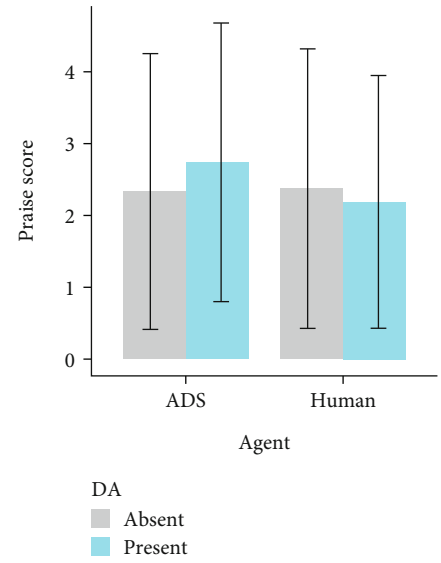

Figure 6: Praise assignment following a collision across driver agents and DA conditions. Note: error bars indicate standard deviation.

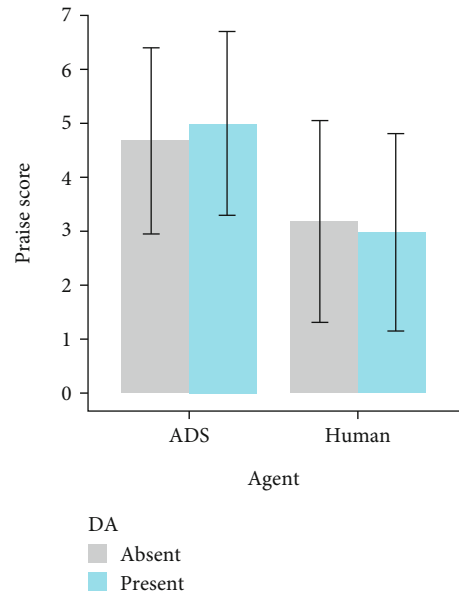

Figure 7: Praise assignment following a near-miss where the vehicle responded appropriately across driver agents and DA conditions. Note: error bars indicate standard deviation.

3.4. Trust. To understand the impact of a DA on trust scores across event type and outcome severity, a 2 (DA) × 2 (event type) × 3 (outcome severity) mixed ANOVA was performed. Results of the three-way ANOVA identified no significant three-way interaction between DA, event type, and outcome severity ($F(1.73, 854.21) = 0.56$, $p = 0.548$). As there was no main effect of event type ($F(1,495) = 3.55$, $p = 0.060$, or two-way interaction effect with event type and DA ($F(1,495) = 0.11$, $p = 0.746$), a further two-way

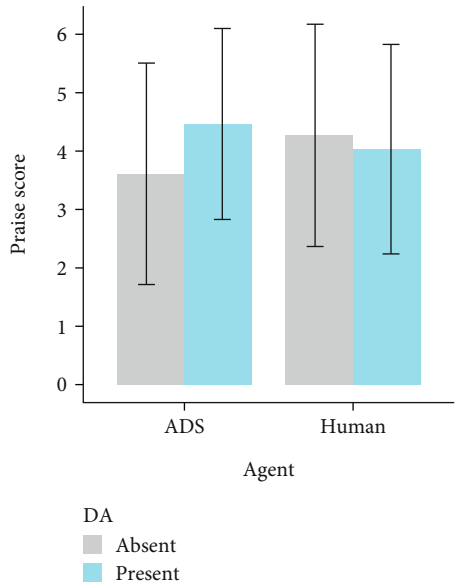

Figure 8: Praise assignment following a near-miss where the human responded appropriately across driver agents and DA conditions. Note: error bars indicate standard deviation.

mixed ANOVA was conducted to determine the effect of DA and outcome severity on total trust score. Shown in Figure 9(a), the two-way ANOVA revealed that there was a statistically significant interaction between DA and outcome severity ($F(1.74, 865.09) = 16.61$, $p < 0.001$). Main effect analysis showed that the presence of the DA ($M = 23.9$, SD = 6.8) resulted in significantly higher trust than the absence of the DA ($M = 20.9$, SD = 7.4) ($p < 0.001$). Additionally, a main effect of outcome severity was found for total trust ($p < 0.001$). Pairwise comparisons indicated that there were significant differences between each outcome severity on total trust with the highest trust for near-miss vehicle responding ($M = 26.7$, SD = 5.1), followed by near-miss human responding ($M = 23.5$, SD = 5.7) and then collisions ($M = 17.1$, SD = 7.1).

3.5. Anthropomorphism. To understand the impact of a DA on perceived anthropomorphism of the ADS across event type and outcome severity, a 2 (DA) × 2 (event type) × 3 (outcome severity) mixed ANOVA was performed. Results of the three-way ANOVA identified no significant three-way interaction between DA, event type, and outcome severity ($F(1.72, 851.34) = 0.83$, $p = 0.422$). As there was no main effect of event type ($F(1,495) = 3.07$, $p = 0.080$), or any two-way interaction effect with event type as an independent variable, a further two-way mixed ANOVA was conducted to determine the effect of DA and outcome severity on perceived anthropomorphism. Shown in Figure 9(b), the two-way ANOVA revealed that there was a statistically significant interaction between DA and outcome severity ($F(1.72, 855.08) = 23.09$, $p < 0.001$). A main effect was found for DA as the presence of the DA ($M = 28.8$, SD = 8.5) resulted in significantly higher anthropomorphism than the absence of the DA ($M = 22.7$, SD = 10.8) ($p < 0.001$). Additionally, outcome severity had a statistically significant main effect on total anthropomorphism ($p < 0.001$). Pairwise comparisons indicated that there were significant differences between each outcome severity on perceived anthropomorphism, with the highest perceived anthropomorphism for near-miss vehicle responding ($M = 30.3$, SD = 7.2), followed by near-miss human responding ($M = 27.2$, SD = 8.9) and then collisions ($M = 19.8$, SD = 11.1).

4. Discussion

The current study examined the assignment of blame and praise towards ADS-equipped vehicles and human drivers following a collision or near-miss as well as how an ADS-equipped vehicle integrated with a DA that administers MRs influenced these attributions. The results supported all hypotheses. Following a collision, participants blamed the ADS proportionately less and praised the ADS proportionately more when the DA was present. It should be noted that each agent was blamed and praised equally when the DA was absent. Following a near-miss where the vehicle appropriately responded, participants blamed the ADS proportionately less and praised the ADS proportionately more when the DA was present. Additionally, the ADS was significantly assigned less blame and more praise than the human regardless of the DA presence. Following a near-miss where the human appropriately responded, participants blamed the ADS proportionately less and praised the ADS proportionately more when the DA was present. When the DA was absent, each agent was blamed equally. Furthermore, more praise was assigned to the ADS than the human when the DA was present, yet more praise was assigned to the human when the DA was absent. Across the scenarios, we found that perceived anthropomorphism of the ADS was higher when the DA was present than when the DA was absent, thus indicating evidence of a valid manipulation for DA presence. We found that higher perceived anthropomorphism was strongly associated with higher trust, usefulness, and satisfaction and that self-reported trust and anthropomorphism of the ADS were significantly higher when the DA was present. Finally, it seems that driving experience with ADS features did not strongly or moderately influence blame or praise attributions towards the human or ADS, or even trust, usefulness, satisfaction, or perceived anthropomorphism of the ADS-equipped vehicle. This could be due to the fact that participants, on average, self-reported low experience with vehicle automation features.

The results were as expected and align with past research. When the vehicle control and oversight were ambiguous, no differences in blame assignment between the human and the ADS were identified, aligning with Copp et al. [30]. Furthermore, when the responsibility of the driving was more clear-cut with the presence of the DA, there was a distinct difference in blame assignment with the ADS being assigned less blame than the human across all conditions. This aligns with Waytz et al.'s [37] findings where higher anthropomorphism of the vehicle led to reduced blame attribution compared to a low anthropomorphized vehicle. Additionally, the assignment of blame supports Awad et al. [32] and Pöllänen et al.'s [31] findings in

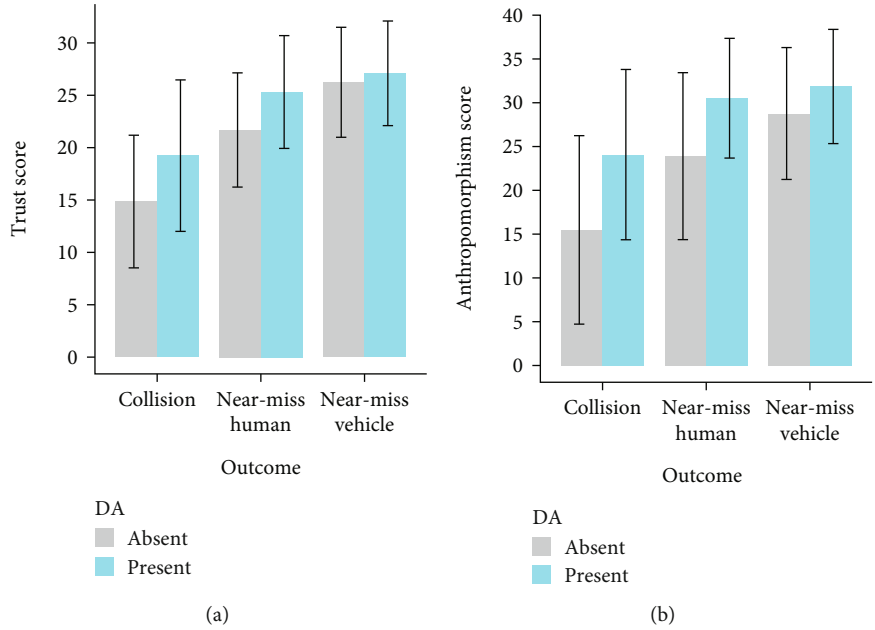

Figure 9: Total trust (a) and anthropomorphism (b) scores across outcome severity conditions and DA conditions. Note: error bars indicate standard deviation.

that less blame is assigned to the ADS when there is greater shared control of the vehicle. Furthermore, our results align with Bennet et al. [29] in that the human driver never received significantly less blame than the ADS. The current results extend the previous research by strengthening the homogeneity of participant perceptions and better unifying expectations of the ADS through the incorporation of corresponding video stimuli.

However, there were conflicting findings between the current results and past research. Liu and Du [28] proposed the blame attribution asymmetry bias whereby individuals have the tendency to blame the vehicle automation more than the human driver in equivalent crashes. This blame attribution asymmetry bias was not found in the current findings; the opposite occurred in which the ADS was blamed equal to or less than the human driver across events. This conflict could be due to Liu and Du [28] focusing on sole driving responsibility as opposed to shared cooperation of the driving task utilized in the current study. Furthermore, our results conflict with Young and Monroe's [42] two main findings. First, they concluded that participants blamed ADS-equipped vehicles more than humans when our findings indicated the reverse. Second, they found that when incorporating driving assistants with ADS-equipped vehicles, participants attributed equal or more blame towards the ADS, dependent on the human likeness of the driving agent. However, these conflicts could be due to the methodological differences as Young and Monroe [42] focused on hypothetical trolley dilemmas whereas the current study used more realistic driving scenarios individuals could come across.

As previous research has yet to examine the assignment of praise towards driving agents following critical events, our results can be considered novel findings and provide an initial step towards understanding positive perceptions of ADS-equipped vehicles. We identified a consistent pattern of praise assignment across the conditions in that the ADS was assigned more praise than the human when the DA was present. The human was assigned more praise only in the near-miss human responding condition without the DA which could be due to the human driver appropriately responding to a critical event without being prompted. With the DA presence, the ADS was praised more possibly due to the participants praising the capacity that the DA could identify and communicate uncertainty with the human, allowing the human to respond appropriately. Further, we wanted to examine whether participants attributed any semblance of positive perceptions towards the ADS in the context of negative collision events. Here, similar results were identified in which participants attributed significantly more praise towards the ADS than the human. Participants may have praised the ADS simply for communicating relevant actions and road elements, suggesting that positive perceptions of advanced ADS features may possibly be resilient to negative outcomes.

Given the current findings, we provide support for the inclusion of an advanced virtual driving assistant in ADS-equipped vehicles to better clarify monitoring expectations. With this, we provide three considerations when developing vehicles equipped with level 3 or higher ADS features. First, explicitly clarify assumed monitoring responsibility of the driving to reinforce perceived control and responsibility of the vehicle. Second, consider communicating the system uncertainty to calibrate human drivers' expectations when driving. Third, consider portraying the ADS as its own entity when including a DA. This may lead to the human driver trusting the ADS more and holding more favorable perceptions of the ADS-equipped vehicle. These recommendations can potentially create clarity for assumed control of the vehicle when the ADS faces system uncertainty. Additionally, it could potentially foster better relationships and perceptions with ADS-equipped vehicles.

4.1. Practical Implications. Given the recent vehicle manslaughter charges [4], the pattern of blame attribution from the current study when responsibility is ambiguous conflicts with the current legal ramifications. The findings indicated that individuals assigned more blame towards the ADS when monitoring is not explicitly stated, which is where the current level of ADS technology is at, suggesting that the public's perception of liability falls unfavorably on the ADS. Yet, when actively communicating which agent is responsible for monitoring the road environment, individuals tended to blame the human driver more for the collision, aligning more with the current legal framework. Having greater legal distinction in driver responsibility when the level of ADS technology increases may close the gap between public and legal liability perceptions. However, it should be noted that assignment of blame may not correlate with assignment of legal liability in situations with resulting injuries; thus, generalizations should be taken with caution.

More broadly, our findings have potentially significant policy implications for vehicle manufacturers and the development of ADS technology. Unfortunately, deceptive marketing of the capabilities of ADS-equipped vehicles can subsequently lead to inappropriate expectations [47] as how ADS-equipped vehicles are described significantly alters publics' capability understanding [48]. Therefore, stricter regulations on the development and deployment of ADS-equipped vehicles may involve strongly enhancing communicative features through actively clarifying human monitoring expectations while operating ADS technology. Additionally, stronger regulations towards how ADS-equipped vehicles are to be marketed to the public as the findings indicate that more advanced ADS technology significantly influences responsibility assignment following critical events. Understanding blame assignment following critical events for various ADS capabilities is critical for determining the appropriate policy responses and for ensuring the safe and responsible development of ADS-equipped vehicles.

In contrast, while understanding praise assignment towards ADS-equipped vehicles is unlikely to be as important as understanding blame assignment when it comes to policy implications, it can still play an important role in shaping public perceptions, attitudes, and, ultimately, adoption of the technology. For example, if an ADS-equipped vehicle is involved in successful operation during critical events, and the public praises the ADS capabilities, it could help increase public confidence in the ADS technology and encourage widespread adoption. This, in turn, could influence policy decisions related to the regulation and deployment of ADS-equipped vehicles, though more so for ensuring ADS-equipped vehicles are safe and reliable.

4.2. Limitations. One limitation of the current study was that participants did not experience both conditions with and without the DA presence. To minimize potential confounding factors between the ADS-equipped vehicles with and without the DA, we needed to use identical videos for each outcome and event type and overlay verbal audio on the videos to imitate a DA. This meant, however, that if participants experienced both DA conditions, they would realize it was the same video, disrupting the driving realism and experimental immersion. Secondly, although complementing text vignettes with a corresponding video stimulus better unified participants' perceptions and expectations of the ADS, these scenarios were not real driving behaviors or experiences and as such did not contain any real perceived risk or personal injury. Therefore, caution is required when generalizing results to real critical events as participants may not have perceived any real risk or negative effect from watching driving simulations, which may influence attributions in naturalistic settings. Thirdly, there is the potential for brand bias with video clips showing a Tesla vehicle which may have influenced participants' opinions regarding trust, praise, and blame. Unfortunately, the vehicles in the City Car Driving software were based on real-world vehicles; thus, removing any branding from vehicles was unattainable at the given time. Fourthly, the large majority of participants were identified as young or white individuals, so caution should be taken when generalizing other demographics. Finally, the trust measure was identified to have varying reliability across all the conditions, with many below the 0.7 threshold. The low reliability scores indicate that the measure was not answered consistently across items among participants. This may be due to the researchers utilizing a trust measure focused on situational rather than general trust. Thus, subsequent inferences regarding trust should be taken with caution.

4.3. Future Research. Given that the current study indicates that integrating an ADS-equipped vehicle with a DA better distinguishes responsibility assignment, future research should focus on identifying elements of the DA and observing the impact of a DA on real-world driving behaviors. First, future research should identify which key road-related environmental factors drivers would like system uncertainty communicated or which environmental feedback participants would like communicated during highly autonomous driving. Second, future research should observe the impact of an ADS utilizing a DA with monitoring requests on driver behaviors. More specifically, assess driving performance metrics, such as response time, as well as eye-tracking to understand where drivers direct their attention following monitoring requests. Third, future research should integrate both verbal (DA) and visual (augmented reality) stimuli to understand responsibility assignment when vehicles are equipped with advanced communication features. Augmented reality enhances drivers' situation awareness during ADS engagement but may direct visual attention away from critical events if drivers are overloaded with information [49].

5. Conclusion

Overall, as vehicle technology's self-driving capacity is increased, more technologically advanced driver systems are required for driver safety. Additionally, it is imperative to understand who is perceived as being at blame for a collision or praised for safe maneuvers in ADS-equipped vehicles

integrated with DAs, the driver or the vehicle? The current study identified the public's perceptions of blame and praise assignment towards ADS-equipped vehicles involved in critical events and found that responsibility assignment is clearly distinguished when monitoring expectations are explicitly stated. That is, individuals tend to blame the human driver more when they are asked to actively monitor the driving environment, whereas individuals tend to assign greater praise to the ADS system when it alerts the human driver. Furthermore, the current study provided support for a potential solution to improve driving safety, mitigate the risk of fatal collisions, and enhance widespread adoption of ADS-equipped vehicles.

Data Availability

The datasets generated during and/or analyzed during the current study are available in the OSF repository (https://osf.io/4ngse/?view_only=3c8485c576ca4933bd2a8c0c20f49262).

Ethical Approval

This research complied with the American Psychology Association Code of Ethics and was approved by the Institutional Review Board at George Mason University (IRB# 1866476-1).

Consent

Informed consent was obtained from all individual participants included in the study.

Conflicts of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] I. Goncharov, "Autonomous vehicle companies and their ML," 2022, Weights & Biases. https://wandb.ai/ivangoncharov/AVs-report/reports/Autonomous-Vehicle-Companies-And-Their-ML–VmlldzoyNTg1Mjc1.

[2] National Highway Traffic Safety Administration, "Standing general order on crash reporting|NHTSA [Text]," 2023, https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting.

[3] K. Conger, "Driver Charged in Uber's Fatal 2018 Autonomous Car Crash," 2020, The New York Times. https://www.nytimes.com/2020/09/15/technology/uber-autonomous-crash-driver-charged.html.

[4] Burke, "Tesla driver charged with manslaughter in deadly Autopilot crash raises new legal questions about automated driving tech," 2022, https://www.nbcnews.com/news/us-news/tesla-driver-charged-manslaughter-deadly-autopilot-crash-raises-new-le-rcna12987.

[5] Z. Lu, B. Zhang, A. Feldhütter, R. Happee, M. Martens, and J. C. F. De Winter, "Beyond mere take-over requests: the effects of monitoring requests on driver attention, take-over performance, and acceptance," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 63, pp. 22–37, 2019.

[6] S. Ma, W. Zhang, Z. Yang et al., "Take over gradually in conditional automated driving: the effect of two-stage warning systems on situation awareness, driving stress, takeover performance, and acceptance," *International Journal of Human–Computer Interaction*, vol. 37, no. 4, pp. 352–362, 2021.

[7] B. Zhang, Z. Lu, R. Happee, J. de Winter, and M. Martens, "Compliance with monitoring requests, biomechanical readiness, and take-over performance: video analysis from a simulator study," in *13th ITS European Congress*, Brainport, the Netherlands, 2019.

[8] W. Zhang, Y. Zeng, Z. Yang et al., "Optimal time intervals in two-stage takeover warning systems with insight into the drivers' neuroticism personality," *Frontiers in Psychology*, vol. 12, article 601536, 2021.

[9] M. Walch, T. Sieber, P. Hock, M. Baumann, and M. Weber, "Towards cooperative driving: involving the driver in an autonomous vehicle's decision making," in *Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*, pp. 261–268, New York, 2016.

[10] F. Naujoks, A. Kiesel, and A. Neukum, "Cooperative warning systems: the impact of false and unnecessary alarms on drivers' compliance," *Accident Analysis & Prevention*, vol. 97, pp. 162–175, 2016.

[11] J. Koo, D. Shin, M. Steinert, and L. Leifer, "Understanding driver responses to voice alerts of autonomous car operations," *International Journal of Vehicle Design*, vol. 70, no. 4, p. 377, 2016.

[12] J. H. Yang, S. C. Lee, C. Nadri, J. Kim, J. Shin, and M. Jeon, "Multimodal displays for takeover requests," in *User Experience Design in the Era of Automated Driving*, A. Riener, M. Jeon, and I. Alvarez, Eds., vol. 980, pp. 397–424, Springer International Publishing, 2022.

[13] N. Du, F. Zhou, D. Tilbury, L. P. Robert, and X. J. Yang, "Designing alert systems in takeover transitions: The effects of display information and modality," in *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 173–180, New York, NY, USA, 2021.

[14] D. R. Large, G. Burnett, B. Anyasodo, and L. Skrypchuk, "Assessing cognitive demand during natural language interactions with a digital driving assistant," in *Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*, pp. 67–74, New York, 2016.

[15] J. Dong, E. Lawson, J. Olsen, and M. Jeon, "Female voice agents in fully autonomous vehicles are not only more likeable and comfortable, but also more competent," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 64, no. 1, pp. 1033–1037, 2020.

[16] S. Lee, R. Ratan, and T. Park, "The voice makes the car: enhancing autonomous vehicle perceptions and adoption intention through voice agent gender and style," *Multimodal Technologies and Interaction*, vol. 3, no. 1, p. 20, 2019.

[17] P. N. Y. Wong, D. P. Brumby, H. V. R. Babu, and K. Kobayashi, "Voices in self-driving cars should be assertive to more quickly grab a distracted driver's attention," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 165–176, New York, 2019.

[18] V. Manusov and B. Spitzberg, "Attribution Theory: Finding Good Cause in the Search for Theory," in *In Engaging Theories*

in Interpersonal Communication: Multiple Perspectives, pp. 37–50, SAGE Publications, Inc, 2008.

[19] S. Botti and A. L. McGill, "When choosing is not deciding: the effect of perceived responsibility on satisfaction," *Journal of Consumer Research*, vol. 33, no. 2, pp. 211–219, 2006.

[20] I. van de Poel and M. Sand, "Varieties of responsibility: two problems of responsible innovation," *Synthese*, vol. 198, no. S19, pp. 4769–4787, 2021.

[21] R. Hakli and P. Mäkelä, "Moral responsibility of robots and hybrid agents," *The Monist*, vol. 102, no. 2, pp. 259–275, 2019.

[22] B. Friedman, ""It's the computer's fault": reasoning about computers as moral agents," in *Conference Companion on Human Factors in Computing Systems*, pp. 226-227, New York, 1995.

[23] X. Lei and P.-L. P. Rau, "Should I blame the human or the robot? Attribution within a human–robot group," *International Journal of Social Robotics*, vol. 13, no. 2, pp. 363–377, 2021.

[24] B. F. Malle, M. Scheutz, T. Arnold, J. Voiklis, and C. Cusimano, "Sacrifice one for the good of many? People apply different moral norms to human and robot agents," in *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, pp. 117–124, New York, NY, USA, 2015.

[25] M. M. A. de Graaf and B. F. Malle, "People's explanations of robot behavior subtly reveal mental state inferences," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 239–248, Daegu, Korea, 2019.

[26] Y. E. Bigman, A. Waytz, R. Alterovitz, and K. Gray, "Holding robots responsible: the elements of machine morality," *Trends in Cognitive Sciences*, vol. 23, no. 5, pp. 365–368, 2019.

[27] A. E. Monroe and B. F. Malle, "Two paths to blame: intentionality directs moral information processing along two distinct tracks. Journal of Experimental Psychology," *General*, vol. 146, no. 1, pp. 123–133, 2017.

[28] P. Liu and Y. Du, "Blame attribution asymmetry in human–automation cooperation," *Risk Analysis*, vol. 42, no. 8, pp. 1769–1783, 2022.

[29] J. M. Bennett, K. L. Challinor, O. Modesto, and P. Prabhakharan, "Attribution of blame of crash causation across varying levels of vehicle automation," *Safety Science*, vol. 132, p. 104968, 2020.

[30] C. J. Copp, J. J. Cabell, and M. Kemmelmeier, "Plenty of blame to go around: attributions of responsibility in a fatal autonomous vehicle accident," *Current Psychology*, vol. 42, no. 8, pp. 6752–6767, 2023.

[31] E. Pöllänen, G. J. M. Read, B. R. Lane, J. Thompson, and P. M. Salmon, "Who is to blame for crashes involving autonomous vehicles? Exploring blame attribution across the road transport system," *Ergonomics*, vol. 63, no. 5, pp. 525–537, 2020.

[32] E. Awad, S. Levine, M. Kleiman-Weiner et al., "Drivers are blamed more than their automated cars when both make mistakes," *Nature Human Behaviour*, vol. 4, no. 2, pp. 134–143, 2020.

[33] Y.-C. Lee, A. Momen, and J. LaFreniere, "Attributions of social interactions: driving among self-driving vs. conventional vehicles," *Technology in Society*, vol. 66, article 101631, 2021.

[34] R. M. McManus and A. M. Rutchick, "Autonomous vehicles and the attribution of moral responsibility," *Social Psychological and Personality Science*, vol. 10, no. 3, pp. 345–352, 2019.

[35] C. Bartneck, J. Reichenbach, and J. Carpenter, "The carrot and the stick," *Social Behaviour and Communication in Biological and Artificial Systems*, vol. 9, no. 2, pp. 179–203, 2008.

[36] H. M. Gray, K. Gray, and D. M. Wegner, "Dimensions of mind perception," *Science*, vol. 315, no. 5812, pp. 619–619, 2007.

[37] A. Waytz, J. Heafner, and N. Epley, "The mind in the machine: anthropomorphism increases trust in an autonomous vehicle," *Journal of Experimental Social Psychology*, vol. 52, pp. 113–117, 2014.

[38] Q. Zhang, C. Wallbridge, D. Jones, and P. Morgan, "Judgements of Autonomous Vehicle Capability Determine Attribution of Blame in Road Traffic Accidents," 2022, Available at SSRN 4093012.

[39] S. O'Hern and R. S. Louis, "Technology readiness and intentions to use conditionally automated vehicles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 94, pp. 1–8, 2023.

[40] J. Xiao and K. G. Goulias, "Perceived usefulness and intentions to adopt autonomous vehicles," *Transportation Research Part A: Policy and Practice*, vol. 161, pp. 170–185, 2022.

[41] F. Alonso, M. Faus, C. Esteban, and S. A. Useche, "Is there a predisposition towards the use of new technologies within the traffic field of emerging countries? The case of the Dominican Republic," *Electronics*, vol. 10, no. 10, p. 1208, 2021.

[42] A. D. Young and A. E. Monroe, "Autonomous morals: inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas," *Journal of Experimental Social Psychology*, vol. 85, article 103870, 2019.

[43] L. Kettle and Y.-C. Lee, "Gender differences in responsibility assignment towards level 3-ADS vehicles," in *Proceedings of the human factors and ergonomics society annual meeting*, SAGE Publications, Los Angeles, CA, 2023.

[44] B. E. Holthausen, P. Wintersberger, B. N. Walker, and A. Riener, "Situational trust scale for automated driving (STS-AD): development and initial validation," in *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 40–47, New York, NY, USA, 2020.

[45] J. D. Van Der Laan, A. Heino, and D. De Waard, "A simple procedure for the assessment of acceptance of advanced transport telematics," *Transportation Research Part C: Emerging Technologies*, vol. 5, no. 1, pp. 1–10, 1997.

[46] J.-W. Hong, Y. Wang, and P. Lanz, "Why is artificial intelligence blamed more? Analysis of faulting artificial intelligence for self-driving car accidents in experimental settings," *International Journal of Human–Computer Interaction*, vol. 36, no. 18, pp. 1768–1774, 2020.

[47] S. Cao, "Tesla's Claim That Its Cars Are Self-Driving May Cross the Line from Permitted 'Puffery' to False Advertising," 2022, Observer. https://observer.com/2022/09/tesla-self-driving-software-face-false-advertising-elon-musk/.

[48] E. Kassens-Noor, M. Wilson, M. Cai, N. Durst, and T. Decaminada, "Autonomous vs. self-driving vehicles: the power of language to shape public perceptions," *Journal of Urban Technology*, vol. 28, no. 3–4, pp. 5–24, 2021.

[49] L. Kettle and Y.-C. Lee, "Augmented reality for vehicle-driver communication: a systematic review," *Safety*, vol. 8, no. 4, 2022.