Multimodal Technologies and Interaction

Article

"Warning!" Benefits and Pitfalls of Anthropomorphising Autonomous Vehicle Informational Assistants in the Case of an Accident

Christopher D. Wallbridge [1], Qiyuan Zhang [2], Victoria E. K. Marcinkiewicz [2], Louise Bowen [2], Theodor Kozlowski [2], Dylan M. Jones [2,†] and Phillip L. Morgan [2,3,*]

[1] School of Computer Science and Informatics, Cardiff University, Cardiff CF10 3AT, UK; wallbridgec@cardiff.ac.uk

[2] School of Psychology, Cardiff University, Cardiff CF10 3AT, UK; Zhangq47@cardiff.ac.uk (Q.Z.); marcinkiewiczv@cardiff.ac.uk (V.E.K.M.); bowenl7@cardiff.ac.uk (L.B.); kozlowskitr@cardiff.ac.uk (T.K.)

[3] Psychology, Division of Health, Medicine and Rehabilitation, Luleå University of Technology, Regnbagsallen 5, 977 54 Lulea, Sweden

* Correspondence: morganphil@cardiff.ac.uk

† Deceased author.

Abstract: Despite the increasing sophistication of autonomous vehicles (AVs) and promises of increased safety, accidents will occur. These will corrode public trust and negatively impact user acceptance, adoption and continued use. It is imperative to explore methods that can potentially reduce this impact. The aim of the current paper is to investigate the efficacy of informational assistants (IAs) varying by anthropomorphism (humanoid robot vs. no robot) and dialogue style (conversational vs. informational) on trust in and blame on a highly autonomous vehicle in the event of an accident. The accident scenario involved a pedestrian violating the Highway Code by stepping out in front of a parked bus and the AV not being able to stop in time during an overtake manoeuvre. The humanoid (Nao) robot IA did not improve trust (across three measures) or reduce blame on the AV in Experiment 1, although communicated intentions and actions were perceived by some as being assertive and risky. Reducing assertiveness in Experiment 2 resulted in higher trust (on one measure) in the robot condition, especially with the conversational dialogue style. However, there were again no effects on blame. In Experiment 3, participants had multiple experiences of the AV negotiating parked buses without negative outcomes. Trust significantly increased across each event, although it plummeted following the accident with no differences due to anthropomorphism or dialogue style. The perceived capabilities of the AV and IA before the critical accident event may have had a counterintuitive effect. Overall, evidence was found for a few benefits and many pitfalls of anthropomorphising an AV with a humanoid robot IA in the event of an accident situation.

Keywords: autonomous vehicle; informational assistant; robot; dialogue style; anthropomorphism; trust; blame; accident outcome

Citation: Wallbridge, C.D.; Zhang, Q.; Marcinkiewicz, V.E.K.; Bowen, L.; Kozlowski, T.; Jones, D.M.; Morgan, P.L. "Warning!" Benefits and Pitfalls of Anthropomorphising Autonomous Vehicle Informational Assistants in the Case of an Accident. *Multimodal Technol. Interact.* **2024**, *8*, 110. https://doi.org/10.3390/mti8120110

Academic Editors: Andreas Riener, Myounghoon Jeon (Philart) and Ronald Schroeter

Received: 21 October 2024
Revised: 22 November 2024
Accepted: 23 November 2024
Published: 5 December 2024

1. Introduction

Highly autonomous vehicles (AVs) could offer major benefits to road safety, including significant reductions in injuries and fatalities. There were an estimated 1.19 million driving-related fatalities in 2021 [1], falling only by 5% since 2010 compared to the 50% reduction target set for the Decade of Action for Road Safety (2011–2020). In the US alone, there were 38,824 lives lost in road traffic crashes during 2020, reflecting an increase of 6.8% compared to 2019 [2], and in the UK, there were 1645 fatalities (29,643 including those seriously injured) during 2023 [3]. Human drivers-as well as pedestrians and other road users are often cited as key causes in as high as 90% of cases [4]. While some technological advances to road vehicles seem to be having a positive effect, marked benefits are not yet being realised.

Unlike human drivers, highly autonomous vehicles that can drive themselves most (SAE, Level 4 [5]) or all (SAE, Level 5 [5]) of the time will be able to constantly monitor the environment through multiple sensors, know the exact location, speed and accelerative velocity of other vehicles (e.g., through direct vehicle-to-vehicle/V2V or vehicle to infrastructure/V2I communication) and adjust their driving dynamics through advanced data-driven, reinforced learning (RL) control systems [6,7], always obey the rules of the road, and not suffer from lapses of judgement, effects of fatigue, and so on. Those in the 'driving seat' will essentially become passengers and may no longer be expected to pay full attention to the vehicle operations, the road ahead or around them, and should be able to perform other non-driving related tasks (NDRTs). These capabilities and benefits of highly autonomous vehicles are positive and should seem very appealing to potential users. Yet, some might be potential barriers to adoption and continued use.

The likely benefits of highly autonomous vehicles will only be realised if they are trusted (with links to safety and reliability, among other factors), accepted by users and then adopted at scale [8]. Based on National Safety Council [9] guidance and recommendations by the USDOT [10], it estimated that AVs could result in reductions to crash and injury rates caused by human drivers by up to 50% following 10% market penetration and, up to 90% following 90% penetration. They also predicted that accidents typically caused by pedestrians and cyclists would be halved. Ref. [11] revisited 453 non-AV crashes using the Baidu Apollo AV counterfactual simulation and determined that 60.91% could have been avoided and some injuries mitigated in the remaining 233 scenarios. While these findings indicate potential significant increases in road safety, they suggest that accidents will not be eliminated.

Despite promising findings (see also [12]), ref. [11] noted at least seven scenarios where an AV would have likely not been able to avoid a crash, including when an agent suddenly emerges from a blind spot. They stressed that under such circumstances, 'even if the perception system [of the AV] detects the object [e.g.,] vehicle in time, there is no sufficient time and space for AVs to develop an informed driving decision' (p. 10). Drivers and other road users, including pedestrians, are also prone to errors of judgement that even sophisticated, highly autonomous vehicle systems will not always be able to detect and/or avoid. Ultimately, accidents involving highly autonomous vehicles–including those that are beyond the capabilities of the technology to predict and react to in time will still occur, albeit to a much lesser extent than vehicles with no or lower levels of autonomy. These are likely to continue to negatively impact trust, potential adoption and perhaps even continued usage following adoption. Therefore, it is crucial to investigate the efficacy of methods that might alleviate reductions in trust following an accident involving a highly autonomous vehicle.

Many potential AV users have concerns about the technology. These are negatively impacting attitudes, trust and intentions to adopt and use them. Relinquishing driving control is a key issue, particularly amongst females, although prior knowledge of the technology is a predictor of willingness to use an AV [13]. Worries about handing over driving control are perhaps understandable given that despite the increasing sophistication of AV technology, mass market penetration has not yet occurred, with most people not having experienced a journey in one. Media coverage alone of as few as one or two incidents involving an AV can erode trust in this potentially revolutionary technology [14–17]. For example, sentiment analysis of over 1.7 million tweets 15 days before and after two fatal crashes involving AVs revealed a 32 per cent point increase in negative comments [18]. However, ref. [18] suggested that opportunities to interact with the technology and for the technology to interact with users will help to increase positivity, even under such negative circumstances. Ref. [19] noted that confidence in AVs will develop through a better understanding of the expectations of the systems, which will, in turn, impact trust and public intention to accept and potentially adopt them. Ref. [20] also found that system transparency, technical competence and situation management all affect trust in an AV.

Maybe then, keeping users 'in the loop' about AV intentions and actions at least some of the time needs to be seriously considered.

Situation awareness [21,22]–including the perception and understanding of what the AV is doing (actions) and planning on doing (intentions)–will likely be significantly impoverished in vehicles with higher levels of automation, especially Levels 4 and 5. This could–on occasion–result in unintended consequences. For example, adaptive cruise control (ACC) systems, which will be an integral component of Level 4 and 5 autonomous vehicles, have become increasingly sophisticated with intelligent control systems that ensure safety and comfort, such as fussy control [23,24] which can imitate the behaviours of skilled human drivers. Should we as users be made aware of these systems' actions and intentions, as well as the underlying reasoning when taking a journey in an AV, or else, as [25] suggested–we will have to beware of the unexpected with potential consequences to trust. Lowered levels of SA might be desirable for users of highly autonomous vehicles when agents (e.g., the AV, other vehicles, pedestrians) are performing optimally and safely, but could have counterintuitive effects when one or more agents are not: for example, in a situation that culminates in an accident. Even with significant advances in AV technological capabilities, there are factors beyond control that could lead to an accident or other negative experiences. For example, cyber security has become a major concern for connected and autonomous vehicles. The latest developments in the field of data-driven, model-free control, such as H-infinity control for ACC [26,27], show great promise and resilience in the handling of uncertainties and external disturbances, such as in the event of a cyber-attack. However, a cyber incident, even when successfully prevented from causing tangible damage, will damage trust and have consequences for other factors, such as attribution of blame, especially when the nature of the incident or the actions of the system are not sufficiently communicated to the users [28,29].

When an accident that is not directly attributable to the AV itself occurs, a level of blame will be assigned to it, and trust will be diminished [17,30], affecting the potential acceptance, adoption and continued usage of the technology. Such ironies of automation are not new: they were predicted more than 40 years ago [31], with the effects extended by others [32] and recently some AV sceptics [33]. Trust is a key enabler to the adoption and continued usage of many technologies, which have been stressed within Human Factors and related fields for decades, and often in response to advances in automation [34], including AVs [35]. Minor errors involving AVs may not erode trust significantly compared to the same errors caused by human drivers–such as taking a long to park or to drive when a traffic light turns from red to green [36], though this may not hold when an AV is involved in a situation with more major consequences, such as a collision. Loss of trust could lead to the disuse of Avs, which could have a major impact on the future of this potentially transformative approach to road transport. It is, therefore, important to understand how trust is affected after an accident involving a highly automated vehicle, as well as how blame might be apportioned. With a strong prediction that the effects on both will be negative, it is also important to explore the efficacy of interventions that might limit the loss of trust and reduce blame on the technology in the event of such accident situations.

Some research has already shown that different patterns of trust and blame exist between humans and AI agents. Ref. [30] demonstrated that AVs and human drivers are blamed differently for the same incident. In five out of six cases (e.g., involving pedestrians or animals stepping onto the road), AVs were blamed to a greater extent than human drivers. In one scenario, this effect reversed: where the AV or human-driven vehicle was overtaking a stopped bus, and a pedestrian stepped out from in front of it, violating the UK Highway Code. In this case, the human driver was blamed more than the AV. It was postulated that the 'bus' scenario differed from the others in that it had stronger potential causal cues that an accident could happen (e.g., the parked bus obstructed the view of the pavement, it had stopped at a bus stop such that passengers may have been disembarking) and therefore human intuition about the scenario and possible risks (despite the Highway Code rules) would have potentially resulted in a non-overtake decision. The other five scenarios involved immediate actions by a third party (e.g., a person suddenly stepping out into the road of a falling tree) and required immediate reactions by the AV. Despite the AV not being able to safely stop in time to avoid the incidents, the time to react should be a strength of this technology over a human driver. Further investigations into causal cues confirmed this is a contributing factor to the difference in assignment of blame and trust [17].

Refs. [17,30] concluded that the differences in trust and blame are, in part, caused by the expectation of the actions and capabilities of AV technology compared to a human driver. One way to further explore this is by providing users with information about the AVs' intentions and actions during a journey. Such information could reinforce to participants that the AV (i) did not directly cause the eventual accident and (ii) that it could not have responded in time to avoid the outcome. Our focus within the current paper is on accident scenarios involving an AV and its effects on trust and blame–taking a similar experimental approach [17,30]. One overarching question following [30] is on what happens to user trust in and blame on AVs in the event of an accident that is not directly caused by the vehicle and is extended by exploring the effects of having received information on the Avs' actions and intentions during the journey, prior to the critical accident outcome. Within the current paper, it is examined whether and to what extent voice-based informational assistants (IAs) designed to keep users informed about the actions and intentions of the AV during journeys can help to limit the degradation of trust in and blame on an AV in the event of an accident. The findings will have important implications for the design and regulation of autonomous vehicles with regards to the ways of improving transparency and interpretability of complex control systems (such as H-infinity control), both with their intentions and actions, to increase the resilience of trust in negative situations. Two types of AV IA are compared though manipulations of anthropomorphism and embodiment: physical embodiment with a humanoid robot and non-physical embodiment with a voice-only system. Also manipulated is the nature of the information (actions and intentions) communicated via each type of AV IA using a personable first-person conversational dialogue style compared to a more informational third-person style.

Some researchers have explored ways to increase passenger trust (as well as factors such as SA) during journeys in AVs. Thus, we look to such key studies to build predictions here. For example, augmented reality allows passengers to visualise internal information processed by an AV (including dynamic and static scenario factors, such as other vehicles, pedestrians and road signs) and has been found to improve trust, SA and user experience. These positive effects were captured during animation-based simulated settings, but only SA improved when using video footage [37], perhaps because the animations were more engaging. Also, such approaches rely on attracting and capturing visual attention as well as understanding the augmented feedback provided. Other approaches include the use of verbal communication and manipulation of styles and feedback/explanations [38] and are perhaps more desirable for some, given that highly autonomous vehicles should allow users opportunities to not have to visually engage with the AV technology or driving environment. Ref. [39], in a simulation study, manipulated three verbal explanation types (none, simple, attributional) and perceived risk. They found that at low levels of perceived risk, attributional explanations led to the highest trust ratings and no explanation for the lowest effects. This effect was, however, reversed in the highest perceived risk condition where, as well as experiencing adverse weather and high driving speeds, pedestrians occasionally ran into the road. Explanation did not help under such high-risk AV driving conditions. With such mixed findings, other methods also need to be examined.

Anthropomorphism using robot agents has been shown to increase trust, including within the context of AVs. Ref. [40] demonstrated that anthropomorphising an AV IA (or AVIA) can increase perceptions of competence. Participants experienced a simulation of a Level 5 AV under three informational assistant conditions: voice-only agent, conversational voice-only agent, and conversational with robot embodiment. The latter condition involved a Nao humanoid robot used commonly in human-robot interaction and especially in social robotics studies [41]. Ref. [40] defined conversational style as involving phrases that replicate more natural interaction than an informative style (i.e., the voice-only agent condition). For example, in the informative style condition, participants heard 'tunnel ahead', whereas in the conversational style condition, participants heard 'we are entering a tunnel'. After eight events, the conversational robotic agent was rated significantly more competent than the other two conditions. It was also rated significantly higher in terms of warmth (an attitude towards the robot). However, these findings were observed in situations with no negative outcomes, where the AV and other road agents were performing optimally. Thus, they cannot be generalised to situations involving unexpected and/or unavoidable incidents resulting in an accident.

Ref. [42] designed and conducted a similar experiment. The voice, generated by Amazon Polly, was considered to be female and from the USA. Participants were shown a series of nine events within each condition. The presence of the Nao robot significantly increased the perception of competence (another attitude towards the robot). The conversational dialogue style was considered less annoying than informational and resulted in a lowered self-reported workload. A further study by [43] involved the same 2 × 2 design (conversation style: informative vs. conversational; embodiment: NAO robot vs. no robot) within the context of a Level 3 partially autonomous AV. There were four take-over requests (where the participant would be expected to drive the vehicle) and four events where the automation was able to handle all driving aspects. The authors reported a significant effect of embodiment with the robot trusted more.

While the findings from the [40,42,43] studies are promising in terms of robot IAs within AVs to support, e.g., perceptions of competence and warmth, all involved the AV successfully performing manoeuvres. However, they did not examine the efficacy of different types of IA or dialogue styles (e.g., conversational, informative) in situations where the AV was involved in an incident or accident. The results were also mixed. For example, in some cases, dialogue style was found to be the most important factor, and in other cases, it was embodiment. Also, potential interaction effects were not considered and perhaps not possible due to elements of the experimental designs.

Context is also important to the way anthropomorphism affects trust. Ref. [44] found that robots can have a detrimental effect on perceptions of reliability and attention in an industrial environment. The authors suggested that this was in part based on the context, e.g., an industrial robot involved in a collaborative task rather than a social robot. Also, while perceived trust in the robot did not appear to interact with anthropomorphism, trust towards the robot did interact with the level of anthropomorphism. In the case of a more anthropomorphic robot, participants took longer in hand-over events. As with industrial contexts, AVs are safety-critical systems that, like non-autonomous vehicles, can cause harm or even fatalities in the event of an accident. Therefore, it is crucial to better understand the potential role that a robot (and other types of) IA(s) might play within AVs, as well as boundary conditions associated with any potential benefits or downsides. This is especially important given that robot IAs might be perceived as having capabilities beyond that of providing a narrative on AV intentions and actions, such as being able to intervene in the event of an unfolding accident situation [17,30].

Current Experiments

Three experiments were designed to investigate the effects of speech-based AV IAs on trust and blame in a highly autonomous vehicle following an accident that occurs at the end of a scenario. Similar to research by [40,42,43], the presence of an embodied robot agent (vs. a speech-only system) and dialogue style as a means to vary levels of anthropomorphism were examined. Agent embodiment was manipulated through physical presence with a humanoid robot compared to non-physical presence without a humanoid robot. To expand further upon previous research, a third no dialogue (control) condition was included to allow comparisons to baseline. The accident scenario chosen was one in which the AV could not have detected or stopped in time to prevent a pedestrian from walking into the road to the front of a parked bus (violating the Highway Code) while the AV was committed to an overtake manoeuvre (as in [30]). This is not an uncommon accident scenario. For example, ref. [45] revisited data on the contribution of vehicle manoeuvres to road accidents in the UK from [46] and determined that 20,310 of 340,728 accidents (almost 7%) occurred during an overtaking situation and was the fourth most likely cause of twelve. As in, e.g., [42,43], dialogue style has three levels: conversational, informational, and no-speech (control). In Experiment 1, the AVs actions and intentions communicated by the IAs are assertive, but more cautious in Experiment 2. Within Experiment 3, participants experienced instances of AV overtaking and not overtaking buses with no negative outcomes before the critical accident event.

Hypotheses

Previous research has established that anthropomorphism through robot presence and manipulations of dialogue style can increase perceptions of competence and trust in AVs [36–38], and thus, the following predictions are made:

Trust

HTrust-1–Agent: Anthropomorphising an AV through embodiment with a humanoid robot IA will result in higher trust following an accident than in conditions with a speech-only system.

HTrust-2–Dialogue Style: An AV with an IA communicating intentions and actions using first-person conversational style will be trusted more than when using informational third-person dialogue, and both conditions will result in higher trust than no-speech (control) conditions.

HTrust-3–Agent*Dialogue Style interaction: Trust will be higher in the robot-embodied conversational condition than in the no-robot conversational condition. The difference will be reduced with informational speech. There will be no difference between agents within the no-speech dialogue style conditions, and overall, trust will be lowest in these conditions.

Blame

Based on blame being closely related to trust, the following hypotheses are made:

HBlame-1–Agent: As in HTrust-1, though with blame lower in the robot agent conditions.

HBlame-2–Dialogue Style: As in HTrust-2, blame is lowest in the conversational dialogue style conditions and highest in the no-speech conditions.

HBlame-3–Agent*Dialogue Style interaction: As in HTrust-3, though with blame lower rather than higher in the robot agent embodied conditions, markedly so between the conversational dialogue style conditions and also between the informational dialogue style conditions with the highest blame in the no-speech conditions that will not differ between agent types.

2. Experiment 1

2.1. Experiment 1 Materials and Methods

2.1.1. Participants

A *G*Power* [47] calculation was conducted. With a power of 0.95 and medium effect size predictions (f = 0.25), a minimum of 251 participants were required. Participants were recruited via the online data collection platform Prolific and randomly assigned to one of six conditions until each was filled. They were renumerated £3.75 for taking part. Three hundred and forty-two participants were recruited, and after withdrawals, sound test failures and failing both attention checks (see Materials), there were 296 valid datasets. Ages ranged from 18–73 years (M = 40.33, SD = 13.72). One hundred and fifty-seven were female, one hundred and thirty-five male, three other, and one preferred not to say. Participants spoke English as a first language or were highly proficient as a second language. Participants reported having normal/normal corrected vision and hearing.

Out of the 296 participants, two-hundred and thirty-two held a full driving license, thirty-three had a provisional license, six did not have a license but had had one previously, five were in the process of applying for a provisional license, 18 did not have a license and were not in the process of getting one, and two preferred not to say. Of the 232 full driving license holders, the mean number of years qualified was 21.22 (*SD* = 13.71), ranging from 1- to 55- years. Qualified drivers reported on average 6986.17 miles driven per year before the COVID-19 pandemic (*SD* = 5659.73, *Max* = 40,000, *Min* = 0) with mileage within 12 months of taking part averaging 3840.26 miles (*SD* = 3407.77, *Max* = 25,000, *Min* = 0).

2.1.2. Design and Materials

A 2 (Agent embodiment: physical presence with a humanoid robot, non-physical presence without a humanoid robot) × 3 (Dialogue Style: no speech, informational speech, conversational speech) between participants design was employed. Conversational speech was used in two of the conditions (one with the robot agent), and informational speech was used in another two conditions (one with the robot agent). There were two no-speech conditions: one with an agent and one without (control). The two chosen IA designs were selected based on previous research by [40,42,43] and to represent different ends of the anthropomorphic spectrum: with high anthropomorphism in the physical presence of a humanoid robot (see below) condition and low anthropomorphism in the non-physical presence/no humanoid robot/speech-only condition.

The driving scenario was generated using the Simulation Software Generated Animation (SSGA) method [17,30]. The SSGA was created using cutting-edge driving simulation software *SCANeR Studio* (2021.2) and within a bespoke driving simulator designed, developed, and built by *AV Simulation* (Figure 1).

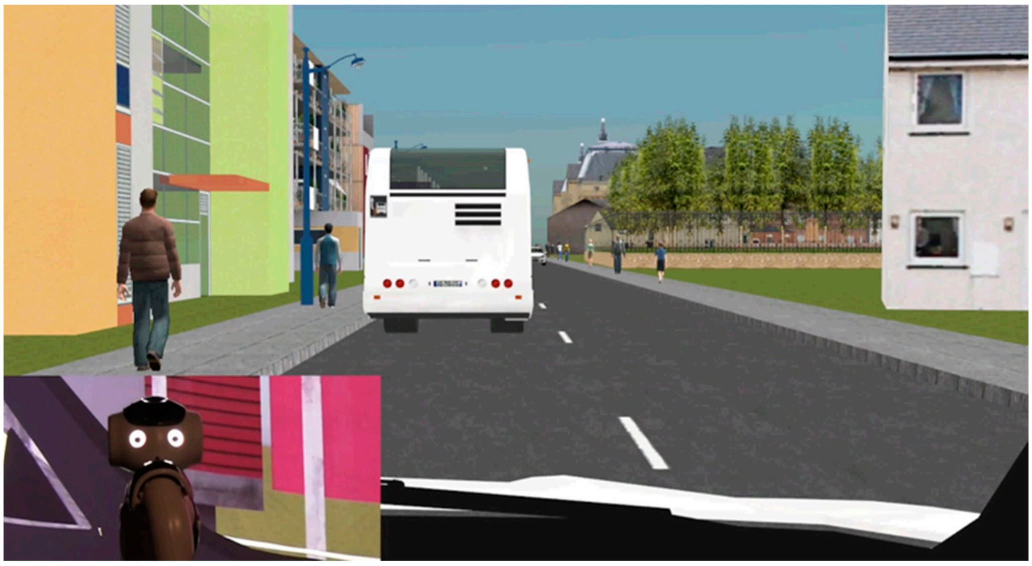


Figure 1. Screenshot of one of the videos created using the SSGA method for Experiment 1. In this–a physical embodiment agent condition-the Nao robot was always positioned to the bottom left of the video image. The view of the robot is from the right-hand seat perspective, with the robot positioned on the dashboard on the passenger side. Within this example (a speech condition), the robot turns to face the passenger as it speaks and turns back and faces the road ahead at all other times.

The SSGA used within all conditions was identical apart from the manipulations of agent embodiment and dialogue style. The SSGA depicted a ~1-min 40-s scenario with a view from where the driving seat would be in a left-side car, looking ahead. The AV is driving along a single carriage road with a low-moderate level of traffic moving in the opposite direction in the other lane. The journey starts in the countryside before entering an urban town setting with buildings, other structures and pedestrians walking along pavements. Early in the scenario (after ~10 s), the AV approaches a moving bus and drives behind it at the speed limit (30 mph/48.28-kph) and at a safe distance for 1 min 14 s before the bus indicator lights are switched on and the bus gradually comes to a stop at a clearly marked and identifiable bus stop. The AV slows down in response and stops at a safe distance behind the bus. After 5-s, the overtake attempt cautiously starts when there is clearly no traffic in the opposite lane or pedestrians visibly on pavements on either side of the road. As the AV is fully committed to the overtake action (side-by-side with the bus), the critical event occurs: a pedestrian suddenly walks out in front of the bus, performing an unsafe action that violates the UK Highway Code. This is only visible when the AV is already parallel to the front of the bus. The AV cannot stop in time. The video stops on a freeze frame before the accident, and the text appears to inform the participant that the AV could not stop in time: it hit the pedestrian, who sustained minor injuries.

Videos of the NAO robot were recorded using an HD video camera and edited and overlayed onto the SSGA footage using iMovie software (9.0.1). The NAO was programmed using *Choregraphe* (2.5.5). In the physical embodiment agent conditions, the NAO was full-screen at the beginning of the SSGA (during the introduction) and then appeared in a smaller box in the bottom left corner of the screen for the remainder of the scenario (Figure 1). When delivering dialogue, the NAO's head would turn to face the participant, accompanied by servo movement audio–to enhance the sense of robot presence in the AV.

In all dialogue conditions, the IA (with the same voice used across all conditions) introduces itself (before the main video starts) as an Autonomous Vehicle Informational Assistant ('AVIA'). In the physical embodiment agent conditions, an NAO v6 (*Softbank Robotics*) is first shown standing and moving its limbs using an animated dialogue style for the introduction. Note that the introduction is the same in the non-physical embodiment agent conditions with dialogue, albeit with speech only. In the physical embodiment agent conditions, and after the introduction, the robot appears to be on the vehicle dashboard on the left-hand side, facing the windscreen and road ahead. The introductory speech in the non-physical embodiment agent conditions is the same as in physical embodiment agent conditions. The following caption was presented before each scenario started, including within the non-speech conditions: "you are about to be shown a driving scenario already in progress".

The dialogue conditions (see Table 1) involved regular (every 10–20 s) updates about the AV actions (e.g., 'driving along a country lane') and intentions (e.g., '[I am] looking for opportunities to overtake'). The dialogue conditions mainly differed by use of language to suggest that the agent was conversing with the participant (passenger) in the first-person. Language such as 'we are', 'I am', 'me' was used in the conversational speech conditions versus e.g., 'the vehicle is' and 'this vehicle is' in the informational speech conditions. Google Text-to-Speech (gtts) was employed to create the voice so that it did not sound 'robot-like' (the default Nao voice), given that 50% of the conditions did not involve a robot agent.

The dependent variables were measured immediately after the SSGA had ended and involved Visual Analogue Scales (VAS) ranging from 0–100. There were three trust measures:

Single Question: Similar to [17], this question was adapted to our specific scenario: "Based on the video footage you just watched, how much would you trust the autonomous system that controls the vehicle you were a passenger in to operate safely on the road in the future?"

Trust in Automated Systems Survey (TiAS): The scale was initially called the Checklist for Trust Between People and Automation [48] and then the System Trust Scale [49]. It has recently been referred to as the Trust in Automated Systems Survey [50]. It is not to be confused with the Trust in Automation (TiA) scale developed by [51], which measures general trust in automation. TiAS was used by [43], and thus, we chose to use it within the current experiments. The TiAS is composed of 12 questions referring to trust in automation in general. The first five questions are framed such that higher scores represent distrust, and the remaining seven are framed with higher scores representing positive trust. TiAS response scales were changed from Likert style to Visual Analogue Scales (VAS). While Likert-format and VAS are similar in terms of the capability of producing valid results and are often highly correlated with each other, VAS is regarded by some as superior in terms of, e.g., test-retest reliability (reproducibility), can produce data with less variance (e.g., through minimising anchor polarisation and/or over-reliance), and allow for a fuller range of response options, e.g., [52–57].

Table 1. Informational and Conversational dialogue scripts within Experiment 1. Note. the final 'warning' speech was the same across all conditions, given that the AV could not stop in time to avoid the collision.

Time	Informational Dialogue	Conversational Dialogue
−0:05–0:00	This is an Autonomous Vehicle Informational Assistant, or AVIA for short. You are about to be shown a driving scenario already in progress.	I am an Autonomous Vehicle Information Assistant, but you can call me AVIA for short. You are about to be shown a driving scenario already in progress.
0:12	Vehicle is driving behind a bus on a country lane, looking for opportunities to overtake.	We are driving behind a bus on a country lane, I am looking for opportunities to overtake.
0:25	The traffic conditions are preventing vehicle finding an appropriate overtaking window.	The traffic conditions are preventing me from finding an appropriate overtaking window.
0:36	The high traffic density is still preventing vehicle from overtaking.	The high traffic density is still preventing me from overtaking.
0:54	Vehicle is still being prevented from overtaking.	I am still being prevented from overtaking.
1:14	The bus is stopping, providing this vehicle an opportunity to overtake.	The bus is stopping, providing me an opportunity to overtake.
1:25	Warning!	Warning!

Situational Trust Scale for Automated Driving (STS-AD [58]): This was used as a multi-item scale with the questions mapping onto a general measure of trust in automated driving. It is composed of six questions relating to trust in the situation. Questions 2, 4 and 5 are presented in a manner that means reverse scoring is needed, as the higher the rating, the lower the trust (i.e., the right anchor represents complete distrust). As with TiAS, responses were recorded using VAS scales.

In addition, to measure (1) propensity to trust and use AV technology and then (2) consider possible changes as a result of taking part in the experiment, participants' attitudes towards autonomous vehicles were measured on VAS scales both at the beginning of the experiment before they were exposed to the SSGA (i.e., pre-trial measures) and at the end of the experiment (i.e., post-trial measures). Two questions were asked pre- and post-: one concerning trust in AV technology in general ("*Imagine that fully autonomous vehicles will be deployed on a large scale on UK roads within the next 12-months. Please rate how much you trust autonomous vehicle technology*" 0 = Do not trust at all; 100 = Completely trust"); and the other concerning the likelihood of using an AV ("*Imagine that fully autonomous vehicles will be deployed on a large scale on UK roads within the next 12-months. Please rate how likely you would be to use an autonomous vehicle.*" 0 = Extremely unlikely; 100 = Extremely likely").

Participants were asked questions about blame on the AV and the pedestrian, the two main agents involved in the critical incident at the end of the scenario, as in [17]. Blame ratings were recorded using VAS scales, with higher scores (maximum 100) representing higher judgements of blame. Taking blame on the AV as an example, the question asked was: *based on the footage you just watched, to what extent do you think the autonomous system that controls the vehicle you were a passenger in should be blamed for the incident that just took place*? The anchor to the left of the VAS was 'not at all' and that on the right was 'completely'.

Given that the experiment was conducted online, an online sound test (based on [59,60]) was developed using *iMovie* software to ensure that participants could adequately hear the speech during the dialogue conditions. Participants were required to listen to five audio clips, each of which contained five words from the phonetic alphabet, and they had to select the quietest word, noting that one was recorded at 50% of the original volume.

2.1.3. Procedure

Participants signed up via Prolific to take part in an experiment on 'road transport'. They were not informed about other elements of the experiment until reading the instructions and when they were debriefed. When ready to start, participants initiated the Qualtrics survey by clicking on a link embedded in the invitation sent to them via Prolific. They were required to read an online information sheet with details on the experiment aims (with no reference to the collision critical incident component), requirements for them (including to remain focused and free from distractions during the entirety of the experiment), data anonymisation and storage, their rights (e.g., to withdraw from taking part, and so on), as well as to ensure that they were wearing headphones or using device (PC or laptop only) speakers set at a comfortably audible level. It was indicated that any elements shown within the vehicle were part of the AV.

An online consent form had to be electronically signed to proceed. Participants were asked to provide a *Prolific* ID to be used if they wished to have data withdrawn within 10-working days. They were given instructions on how to complete the sound tests–including setting the device volume to a comfortable level. They had to correctly identify the quietest word per clip and needed to pass at least three of the five tests to proceed to the main part of the experiment. Participants then completed optional demographic questions on age and gender and driving-related questions, including license status, frequency, and miles driven per year. Their pre-trial general attitudes towards autonomous vehicles were then measured via the trust and intention to use questions specified above.

The main experimental phase involved watching one version of the SSGA according to the conditions assigned. Participants were instructed to pay full attention and to try and eliminate all possible background distractions. Immediately after the SSGA had ended, participants were presented with questions on trust and blame. The single questions on trust and questions on blame were presented first, followed by the TiAS and STS-AD (counterbalanced). Following this, two attention and understanding check questions relating to what happened during the overtake and critical events were included. Then, participants' post-trial general attitudes towards AVs (trust, intention to use) were measured using the exact same questions as during the pre-trial phase. This was followed by an opportunity to add free text comments. Participants were provided with a debrief form that included information on the main aims of the experiment.

2.2. Experiment 1 Results

Data were analysed using SPSS. Unless otherwise stated, between-participant analysis of variance (ANOVA) was used, and post-hoc analysis was performed using Bonferroni with adjustment for multiple comparisons.

Trust

Single Measure: Trust ratings were generally low across all conditions, with only one surpassing 30/100 (physical embodiment + conversational speech), see Figure 2. There was a significant main effect of dialogue style, $F(2,290) = 3.37$, $p = 0.036$, $f = 0.15$, with trust higher in the informational than no dialogue conditions (though marginally non-significant: $p = 0.052$). There was a non-significant main effect of agent embodiment $F(1,290) = 0.54$, $p = 0.46$, and a non-significant interaction, $F(2,290) = 2.18$, $p = 0.12$.

TiAS: Answers to questions 1–5 were reverse-coded prior to all item scores being averaged. TIAS item reliability was very high ($\alpha = 0.91$). Compared to the single-item trust question, average TiAS scores are higher (Figure 3). However, there were non-significant main effects of agent embodiment, $F(1,290) = 0.96$, $p = 0.33$, dialogue style, $F(2,290) = 1.74$, $p = 0.18$, and a non-significant interaction, $F(2,290) = 1.81$, $p = 0.17$.

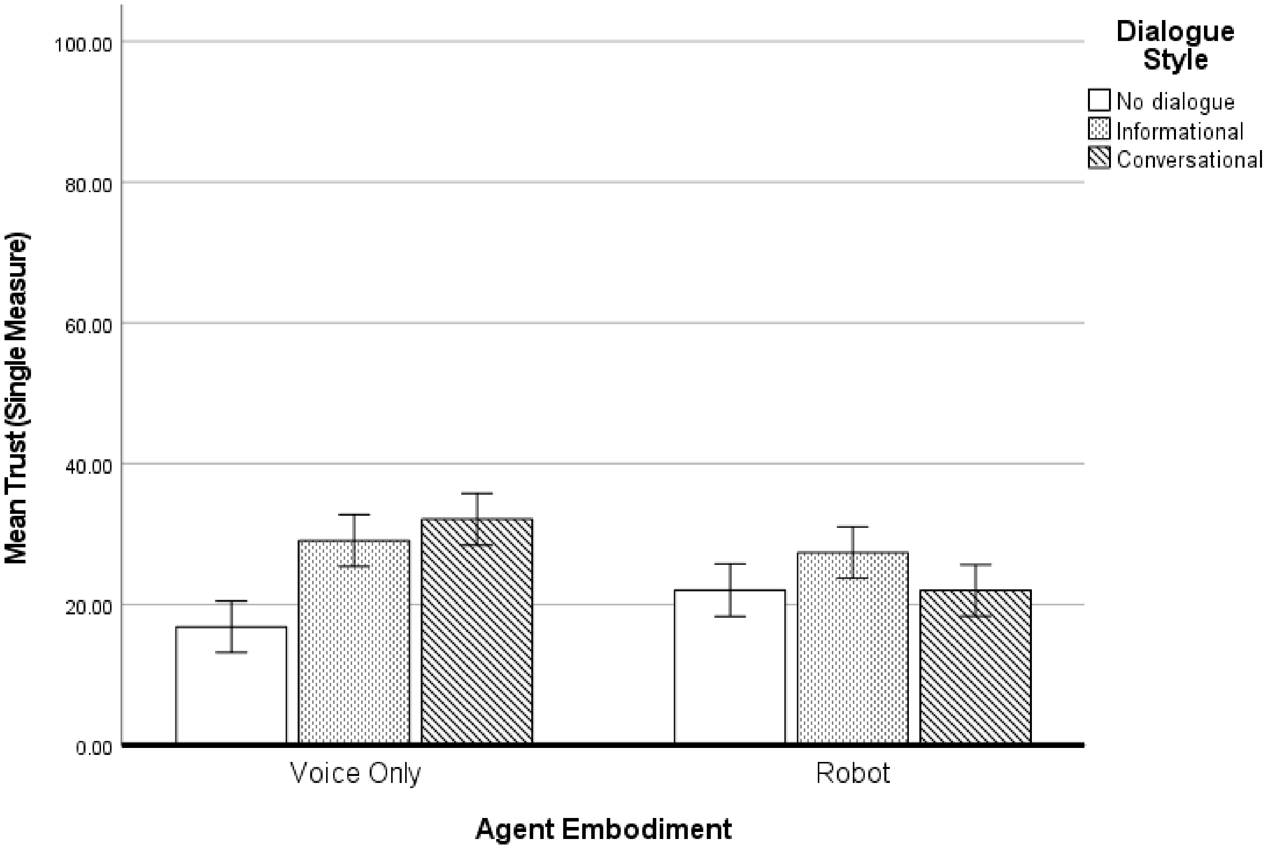


Figure 2. Mean ratings of trust (single measure) across agent embodiment and dialogue conditions. Error bars are ±SE.

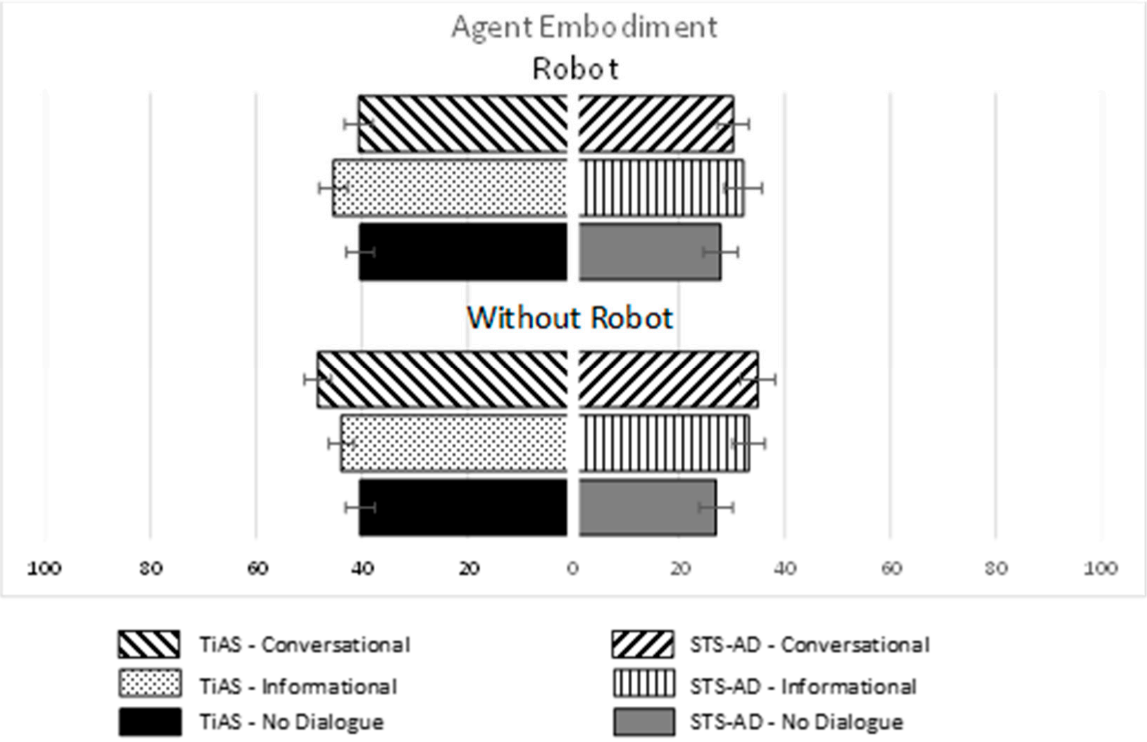


Figure 3. TiAS and STS-AD mean ratings across agent embodiment and dialogue conditions. Error bars are ±SE.

STS-AD: Answers to questions 2, 4 and 5 were reverse-coded prior to all item scores being averaged (Figure 3). Item reliability was high ($\alpha = 0.89$). There were non-significant main effects of agent embodiment, $F(1,290) = 0.38$, $p = 0.54$, or dialogue style, $F(2,290) = 1.70$, $p = 0.19$, and there was a non-significant interaction, $F(2,290) = 0.38$, $p = 0.69$.

Blame

On the AV: Average blame ratings ranged between ~50% and almost 70%, see Figure 4. There was a significant main effect of dialogue style, $F(2,290) = 5.91$, $p = 0.003$, $f = 0.20$, with blame lower in the informational condition than the no dialogue condition ($p = 0.002$). There was a non-significant main effect of agent embodiment, $F(1,290) = 0.11$, $p = 0.74$, and a non-significant interaction, $F(2,290) = 1.09$, $p = 0.34$.

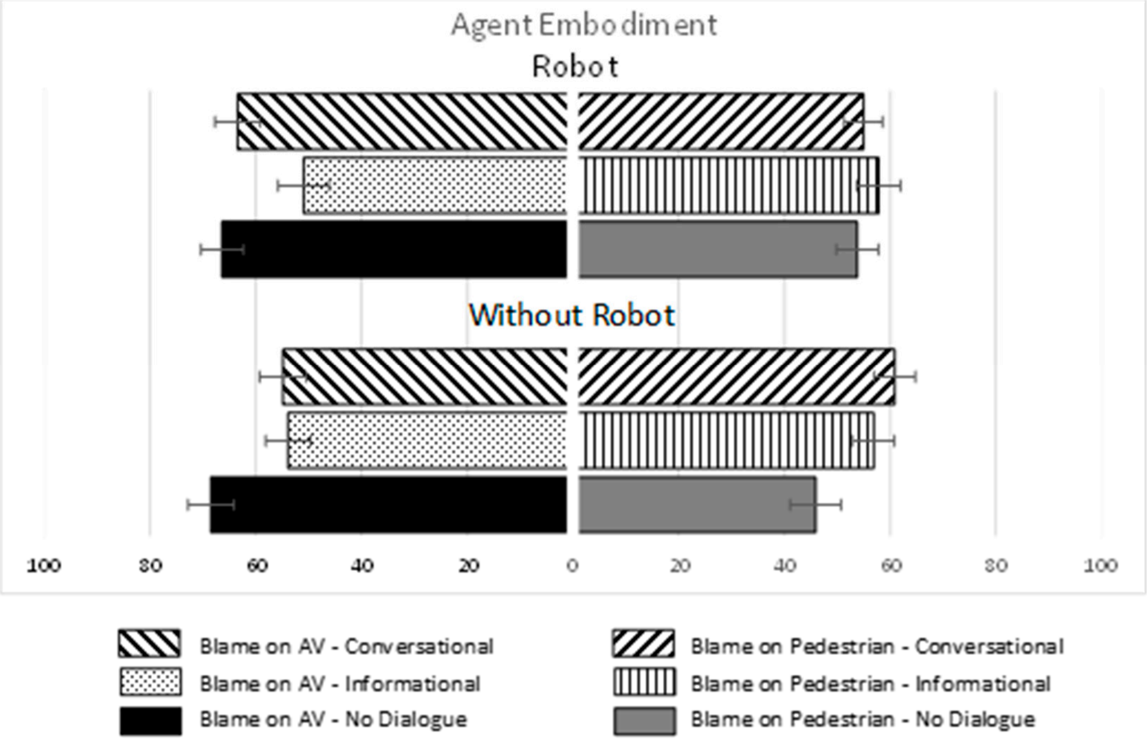


Figure 4. Mean levels of blame on the AV and the pedestrian across agent embodiment and dialogue conditions. Error bars are ±SE.

On the Pedestrian: Average ratings were slightly lower than those on the AV, ranging from just below ~50% to just above 60% (Figure 4). There was no significant main effect of agent embodiment, $F(1,290) = 0.09$, $p = 0.77$, or dialogue style, $F(2,290) = 2.43$, $p = 0.09$, and a non-significant interaction, $F(2,290) = 1.43$, $p = 0.24$.

Attitudes Towards AVs Pre- and Post- Scenarios

Participants' ratings on the two questions regarding their general attitudes towards autonomous vehicles at both temporal points (pre-trial and post-trial) were analysed using a 2 (Stage: pre-trial, post-trial) × 2 (Agent embodiment: physical presence with a humanoid robot, non-physical presence without a humanoid robot) × 3 (Dialogue Style: no speech, informational speech, conversational speech) mixed ANOVA, which revealed a main effect of Stage on trust in autonomous vehicles in general, $F(1,262) = 60.54$, $p < 0.001$, $f = 0.48$. Overall, after watching the SSGA, participants' trust in AVs decreased ($M = 28.85$, $SD = 26.10$) compared to before taking part in the main experimental phase ($M = 37.43$, $SD = 25.88$). This was expected since the SSGA featured an accident scenario. No significant main effects of Agent or Dialogue Style were found, $F(1,262) = 0.42$, $p = 0.52$, and $F(1,262) = 0.30$, $p = 0.742$, respectively, and there were no significant interactions ($ps > 0.05$). Ratings on the likelihood of using an AV in the future displayed a similar pattern. There was a significant main effect of Stage, $F(1{,}262) = 89.50$, $p < 0.001$, $f = 0.59$. That is, after watching the SSGA, participants likelihood of using an AV decreased ($M = 25.91$, $SD = 26.82$) compared to before ($M = 37.22$, $SD = 30.50$). All other main effects and interactions were non-significant ($ps > 0.05$).

2.3. Experiment 1 Discussion

Overall, most findings were non-significant. For trust, only one significant main effect of dialogue style was found. This was with the single-item measure and was not fully supportive of our prediction (HTrust-2): trust was higher only in the informational speech than in the no-speech condition, with no difference between conversational speech and any other condition. There were no significant findings to indicate that a physical embodiment as operationalised with a humanoid robot informational assistant is trusted more than an agent with no physical embodiment. If anything, the pattern of means (noting no significant interactions) across trust measures indicated an increase from no dialogue to informational to conversational in the non-physical embodiment conditions only, with no such patterns in the physical embodiment conditions. For blame, there was also only one significant main effect again of dialogue style and on the AV, with lower blame in the informational than conversational and no dialogue conditions. The HBlame-2 hypothesis is not supported. Additionally, participants' general trust in and intention to use AVs significantly decreased after having experienced the scenario, expectedly given that the outcome involved a negative effect. However, the agent and dialogue style had no effect.

Taken together, the trust and blame findings from Experiment 1 could indicate that our manipulation of agent embodiment and anthropomorphism using a physical robot did not result in the intended effects. In addition, an informational dialogue style seems to–for some measures at least–lead to higher trust and lower blame in the AV than in conversational speech. These findings appear to stand in contrast to previous research where robot presence (a manipulation of agent embodiment and anthropomorphism) has been shown to increase trust in an AV, especially with conversational dialogue style [42,43]. However, to our knowledge, no studies to date have examined trust and blame following an accident scenario. Thus, it could be the case that the potentially promising findings from studies such as those by [40,42,43] are restricted to situations where an AV and other road users and agents perform optimally without incident or accident.

However, it is important to check whether other aspects of the materials might have had an impact on findings and, consequently, conclusions that can be drawn. We examined qualitative open-text comments left by participants at the end of the experiment. A theme that emerged amongst some participants in the informational and conversational conditions is that they felt that the system was constantly trying to find an opportunity to overtake the bus before it pulled into the bus stop. For example, one participant stated that:

> *P36. 'The bit that concerned me was why the system was constantly trying to find an overtaking opportunity, even when the bus went into a 30 zone and obviously slowed down to the lower speed limit.'*

The agent (robot or no robot) was regularly communicating an intention to try and overtake the bus (when safe to do so), and this could have been perceived as risky even prior to the critical event (e.g., [informational] *vehicle is driving behind a bus on a country lane, looking for opportunities to overtake*; [conversational] *we are driving behind a bus on a country lane, I am looking for opportunities to overtake*). This may have impacted trust and blame ratings after the accident scenario (see also [30]) and was quite different to the speech materials used in past studies [40,42,43]. Numerous other studies have shown that a higher perception of risk in a robot can negatively impact trust [61,62], use [63], and satisfaction and intention to reengage [64]. This is also the case in the event of robots being perceived to perform poorly [65].

Experiment 2 was designed to determine whether perceived riskiness was having a confounding effect in Experiment 1. The IA agents (physically embodied with a humanoid robot, non-physically embodied–no robot) used in Experiment 2 communicated actions and intentions as in Experiment 1, but not an intent to overtake the moving bus until the end of the scenario, before the critical event.

3. Experiment 2

Within Experiment 1, there were indications of increased trust and reduced blame in the informational dialogue conditions, and there were no significant effects of physical embodiment using an embodied humanoid robot IA. These findings did not support the predictions. To determine whether perceived riskiness of the AV based on actions and intentions communicated via the IAs (i.e., looking for opportunities to overtake) was masking effects, the main objective of Experiment 2 was to use more cautious dialogue during the scenario and before the critical event. Also included was a question on the perceived riskiness of the AV. All Experiment 1 hypotheses are held for Experiment 2. In addition, it is predicted that there will be no differences in perceived risk across the different dialogue conditions irrespective of agent type but that the AV will be perceived to be riskier in the non-dialogue conditions, especially in the physical embodiment agent condition with higher anthropomorphism though the use of a humanoid robot informational assistant.

3.1. Experiment 2 Materials and Methods

3.1.1. Participants

Two-hundred and twenty-three participants signed up to take part. After withdrawals, participants failing the sound test, and those failing both attention checks, there were 199 valid datasets. Ages ranged from 18–80 years (*M* = 42.05, *SD* = 13.39). Ninety-one were female, and 108 were male. Of the 199 participants, one hundred and sixty-one held a full driving licence, nineteen a provisional licence, one held a license in the past, two were in the process of obtaining a provisional licence, fifteen did not have a license and were not in the process of getting one, and one preferred not to say. Of the 161 full driving license holders, the mean number of years was 21.45 (*SD* = 13.15), ranging from 1- to 59- years. Qualified drivers reported driving on average 7478.88 miles per year (*SD* = 8208.12, *Max* = 60,000, *Min* = 0) before the COVID-19 pandemic and 4191.91 miles per year (*SD* = 6335.59, *Max* = 60,000, *Min* = 0) within the 12-months prior to taking part.

3.1.2. Design, Materials and Procedure

The design was the same as in Experiment 1. The main change was to the dialogue. For example, "I am looking for opportunities to overtake" (Experiment 1) was changed in Experiment 2 to "I am not looking for opportunities to overtake" (i.e., less risky). Additionally, a single measure of perceived risk was added using a VAS. This question was included after all others on trust and blame. Everything else was the same as in Experiment 1, including the IA announcing the intention to overtake the bus prior to the critical event.

3.2. Experiment 2 Results

Trust

Single Measure: As in Experiment 1, mean trust ratings were low, ranging from ~10–~30/100 (Figure 5). Unlike Experiment 1, there appears to be an increase in trust from the no dialogue to informational to conversational conditions in the physical embodiment conditions, with the opposite pattern across the non-physical embodiment conditions (Figure 5). There was a significant main effect of physical embodiment, *F*(1,193) = 4.98, *p* = 0.027, *f* = 0.16, with higher trust in the robot agent conditions. Dialogue style was not significant, *F*(2,193) = 0.10, *p* = 0.90. There was a marginally non-significant interaction, *F*(2,193) = 3.01, *p* = 0.052, *f* = 0.18. Post-hoc analyses revealed significantly higher trust in the robot agent conversational dialogue condition than in the non-robot agent conversational condition (*MD* = 17.64, *p* = 0.002). There were no other significant post-hoc differences.

TiAS: As in Experiment 1, means were higher than the single trust measure, with the highest rating in the physical embodiment conversational condition and lowest in the non-physical embodiment conversational condition (Figure 6). Reliability between items was very high ($\alpha = 0.90$). There was a non-significant main effect of agent embodiment, $F(1,193) = 2.12$, $p = 0.15$, and dialogue style, $F(2,193) = 0.40$, $p = 0.67$, and a non-significant interaction, $F(2,193) = 2.17$, $p = 0.12$. (Despite the non-significant interaction and for exploratory purposes only, there was a significant post-hoc difference identified between conversational conditions with higher trust in the robot than non-robot agent conditions, $p = 0.015$.)

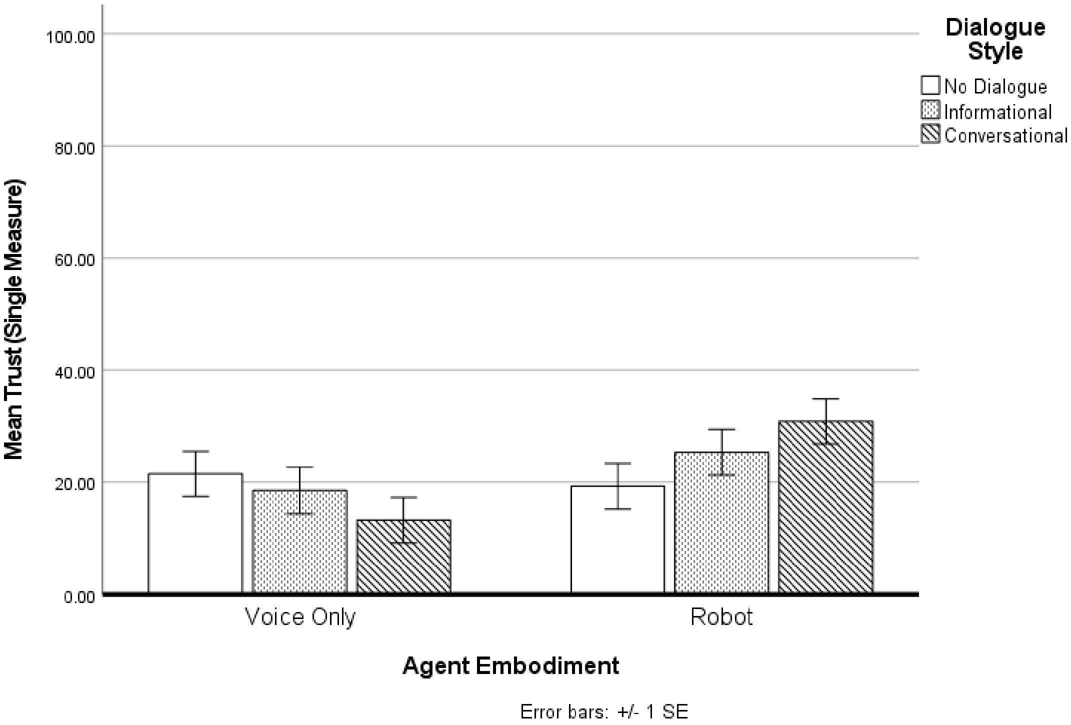


Figure 5. Mean ratings of trust (single measure) across agent embodiment and dialogue conditions. Error bars are ±SE.

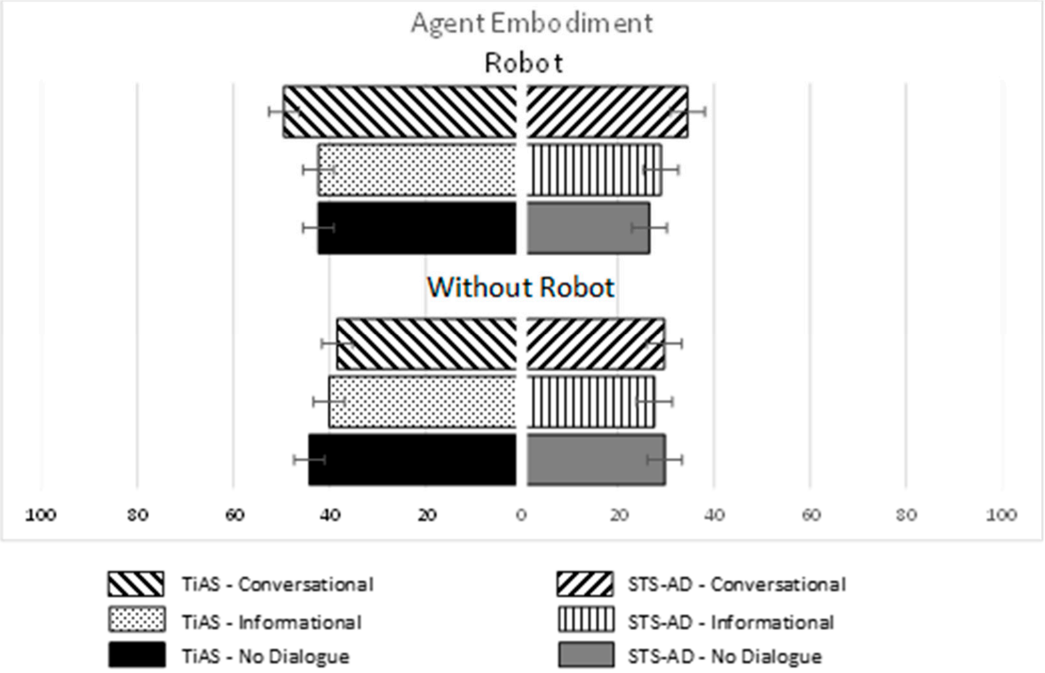


Figure 6. TiAS and STS-AD mean ratings across agent embodiment and dialogue conditions. Error bars are ±SE.

STS-AD (Figure 6): Reliability between items was high (α = 0.86). There was a non-significant main effect of robot presence, *F*(1,193) = 0.11, *p* = 0.74, and dialogue style, *F*(2,193) = 0.76, *p* = 0.47, and a non-significant interaction, *F*(2,193) = 0.62, *p* = 0.54.

Blame

On the AV: Mean ratings were largely above 60/100 and almost as high as 80/100 in the robot agent no dialogue condition (Figure 7). The was a non-significant main effect of physical embodiment, *F*(1,193) = 0.02, *p* = 0.88, and dialogue style, *F*(2,193) = 1.95, *p* = 0.16. There was, however, a significant interaction, *F*(2,193) = 3.78, *p* = 0.025, *f* = 0.20. Post-hoc analyses revealed that with a humanoid robot agent, blame on the AV was higher in no dialogue condition than in the other two speech conditions (*ps* < 0.05). Also, blame was higher in the physical embodiment with no dialogue condition than in the non-physical embodiment with no dialogue condition (*p* = 0.037).

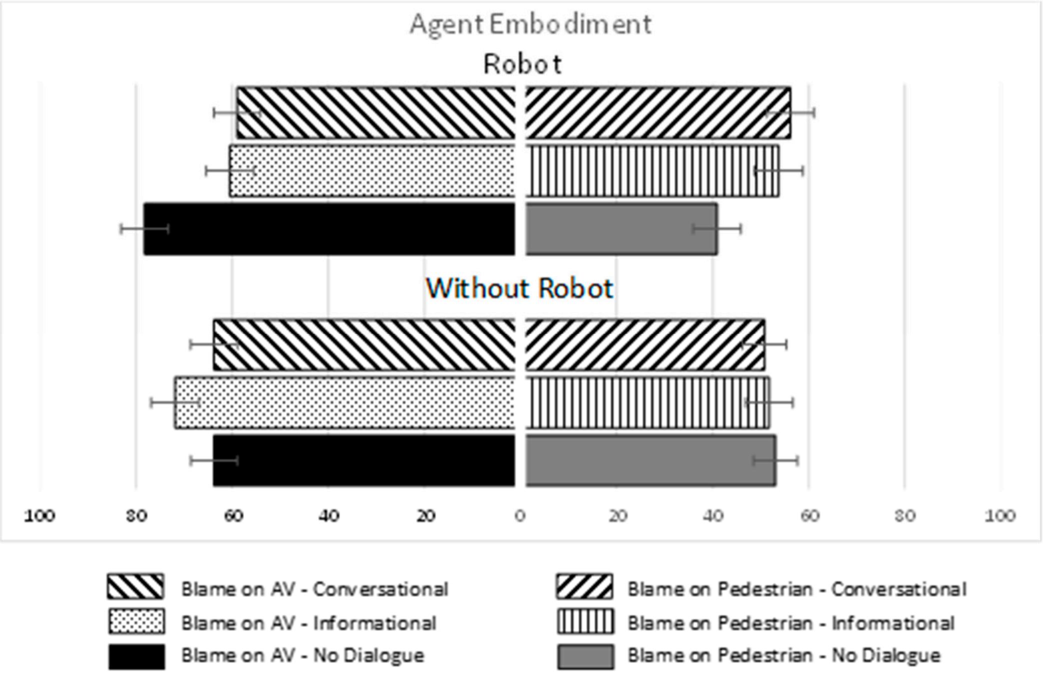


Figure 7. Mean levels of blame on the AV and the pedestrian across agent embodiment and dialogue conditions. Error bars are ±SE.

On the Pedestrian (Figure 7): There were non-significant main effects of agent embodiment, *F*(1,193) = 0.16, *p* = 0.69, and dialogue style, *F*(2,193) = 1.10, *p* = 0.34, as well as a non-significant interaction, *F*(2,193) = 1.87, *p* = 0.16.

Perceived Risk: Mean risk ratings were largely <40/100 and lowest in the robot agent conversational dialogue style condition (Figure 8). There was a significant main effect of agent embodiment, *F*(1,193) = 4.87, *p* = 0.028, *f* = 0.16, with lower risk in the physical embodiment than in non-physical embodiment agent conditions. There was a non-significant main effect of dialogue style, *F*(2,193) = 0.19, *p* = 0.83, and non-significant interaction, *F*(2,193) = 0.22, *p* = 0.80.

Attitudes Towards AVs Pre- and Post- Scenarios

The results were largely similar to that of Experiment 1. There was a main effect of Stage on trust in AVs in general, *F*(1,193) = 44.54, *p* < 0.001, *f* = 0.48. Overall, after watching the SSGA, participants' trust in AVs decreased (*M* = 28.10, *SD* = 25.94) compared to before (*M* = 36.87, *SD* = 26.33). However, as in Experiment 1, no significant main effects of Agent or Dialogue Style were found, *F*(1,193) = 0.01, *p* = 0.95, and *F*(1,193) = 0.14, *p* = 0.87, respectively, or significant interactions (*ps* > 0.05). Ratings on the likelihood of using AVs in the future displayed a similar pattern. Only the main effect of Stage was significant,

$F(1,193) = 69.04$, $p < 0.001$, $f = 0.60$. That is, after watching the SSGA, participants likelihood of using AVs decreased ($M = 26.03$, $SD = 27.83$) compared to before ($M = 36.57$, $SD = 31.34$). All other main effects and interactions were non-significant ($ps > 0.05$).

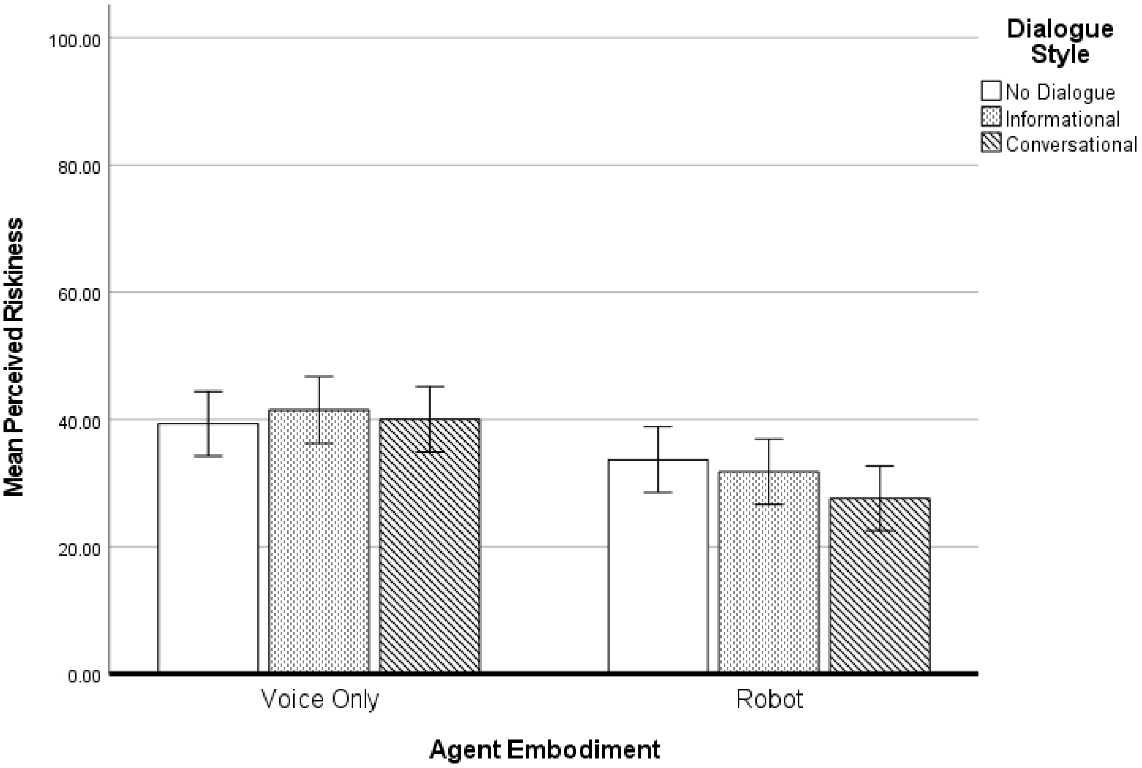


Figure 8. Mean ratings of perceived riskiness across agent embodiment and dialogue conditions. Error bars are ±SE.

3.3. Experiment 2 Discussion

The findings from Experiment 2 are, overall, more supportive of the predictions compared to Experiment 1. HTrust-1 was supported in the case of the single measure of trust: the AV was trusted more in conditions with a physically embodied robot agent. While the interaction with dialogue style was marginally non-significant ($p = 0.052$), trust was significantly higher in the physically embodied robot agent conversational style condition than in the same dialogue style non-physically embodied condition, partly supporting HTrust-3. This conclusion must be treated with some caution as there were no such effects based on TiAS and STS-AD ratings despite a similar pattern of mean differences.

There was also support for HBlame-3, with a significant interaction between agent embodiment and dialogue style, with higher blame attributed to the AV in the robot agent's no dialogue condition than both robot agents' dialogue conditions. In addition, blame was higher in the former condition than in the comparable non-physically embodied condition. Intuitively, this makes sense: having higher anthropomorphism in the form of a humanoid robot IA in an AV that is not providing any information could be interpreted as something like a technological failure and perhaps then, to some degree, could be perceived as being linked to the accident outcome. If a robot IA is to be used within an AV, it likely needs to be communicating with the passenger(s). Otherwise, during an accident situation, it could be perceived to be part of the cause.

A risk measure was also included, largely because of our rationale within Experiment 2 to only include dialogue relating to cautious intentions during the scenario and before the critical event. In general, and as predicted, it was found that the AV was perceived as less risky in physical agent embodiment conditions. However, there was no effect of dialogue style on risk and no interaction. While this risk measure was not included in

Experiment 1, the finding from Experiment 2 provides an indication that risk perception after an accident is lower with a humanoid robot IA than without one. Additionally, as in experiment 1, participants' general trust in and intention to use AVs significantly decreased after having experienced the scenario, most likely because the outcome again entailed a collision. As in experiment 1, agent and dialogue style had no effect on post-experiment trust in or intention to use AVs.

There are possible reasons–even with the cautionary dialogue style adopted before the critical event–why some of our hypotheses are still not fully supported. Participants only experienced the AV attempting an overtake manoeuvre that ended with an accident outcome. It is likely that most participants will not have experienced (even virtually) a journey in an AV and will not have had an opportunity to develop trust in such technology. Indeed, in related previous studies [40,42,43], journeys/scenarios were considerably longer than in our experiments, and this may have afforded participants an opportunity to develop higher levels of trust in the AV. It is important to determine whether having an opportunity to experience an AV performing manoeuvres successfully not only increases trust prior to the accident event but also whether such an increase is higher in a physically embodied robot agent condition (as in [40,42,43]). Moreover, to investigate whether such predicted increases in trust have an impact on trust and blame following the accident event, and if–and as predicted in Experiments 1 and 2–trust is higher and blame lower in the physically embodied robot agent conditions, markedly so with a conversational dialogue style.

4. Experiment 3

4.1. Experiment 3 Introduction

A limitation of the scenarios used in Experiments 1 and 2 is that participants will have had limited experience of AV driving safely and no experience of negotiating overtake attempts without a negative outcome. This may well have negatively impacted trust and blame (measured only after the accident) despite some significant findings within Experiment 2 that fit the hypotheses. In addition, it is not always the case that it is safe to overtake a parked bus: there will be instances where the traffic (e.g., density) and other factors/conditions (e.g., pedestrians) mean that it is too risky and potentially dangerous with the judgement for now made by human drivers rather than autonomous systems. Therefore, within Experiment 3, the scenario and paradigm were further developed to include the AV performing two successful overtake attempts (with no oncoming traffic or pedestrians in sight) and two instances where it does not commit to an overtake manoeuvre (due to oncoming traffic and pedestrians in sight). The former will likely be perceived as riskier despite no negative outcomes. Within Experiment 3, and after each of these events, trust and risk are measured, allowing comparisons prior to the critical accident event and then following that event (as in Experiments 1 and 2).

In addition to our original hypotheses, it is predicted that trust ratings will increase from one event to the other prior to the critical accident event. It is also predicted that this increase will be more marked in the physical embodiment robot IA condition (e.g., [42,43]). Finally, we included questions on attitudes towards the IAs using the Robotic Social Attributes Scale (RoSAS) [66] to determine if the high embodiment robot conditions are preferred compared to the low embodiment non-robot conditions as would be predicted, especially in terms of competence (as was found by [43]). Noting that warmth and discomfort are other attitudes measured using the RoSAS.

4.2. Experiment 3 Materials and Methods

4.2.1. Participants

Participants

The recruitment method was the same as in Experiments 1 and 2. One hundred and ninety-six participants were recruited. After withdrawals, participants failing the sound test, and those failing both attention checks, there were 192 valid datasets. Ages ranged from 18–80 years (*M* = 41.52 *SD* = 12.48). One hundred and ten were female, eight-one male, and one did not provide gender information. One hundred and forty-eight held a full driving licence, twenty-three a provisional licence, three held a license in the past, three were in the process of applying for a provisional licence, and fifteen did not have a license and were not in the process of getting one. Of the full driving license holders, the mean number of years was 19.61 years (*SD* = 13.56, ranging from 1–61 years). Qualified drivers reported driving on average 6587.49.19 miles per year (*SD* = 6343.44 *Max* = 50,000, *Min* = 0) before the COVID-19 pandemic and 4678.24 miles (*SD* = 5618.47 *Max* = 40,000, *Min* = 0) within the 12-months prior to taking part.

4.2.2. Design, Materials and Procedure

Design, Materials and Procedure

As in Experiments 1 and 2, the scenario begins with the AV approaching a moving bus and driving behind it at the speed limit and at a safe distance. After 1 min and 14 s, the bus comes to a stop at a bus stop. Event 1 involves the AV determining that conditions are safe (according to the UK Highway Code) to commit to overtaking the bus: there is a broken white line separating the two lanes, no evidence of oncoming traffic in the opposite lane, and no pedestrian(s) attempting to cross the road. AVIA communicates (in all but the no dialogue control conditions) that conditions are safe to overtake, and the manoeuvre takes place successfully, with the AV continuing on its journey afterwards. Soon after overtaking the bus, another bus comes into view, and the AV drives behind it at the speed limit and at a safe distance. Event 2 involves the second bus stopping at a bus stop. The AV detects oncoming traffic in the opposite lane and determines that the conditions are too unsafe to attempt to overtake (according to the UK Highway Code). The AV IAs communicate (in all but the no dialogue control conditions) that conditions mean it is not safe to overtake, and an overtake manoeuvre does not occur. Instead, the AV waits for the bus to pull off and drives behind it at the speed limit and at a safe distance before it stops at another bus stop. At this point, Event 3 is triggered and is the same as Event 1. Following Event 3, the AV drives behind another bus at the speed limit and at a safe distance before it stops at a bus stop, initiating Event 4, which is the same as Event 2. Event 5 is like events 1 and 3 and mimics the critical event in Experiments 1 and 2. As the AV is committed to the overtaking manoeuvre, a pedestrian steps out in front of the bus, violating the UK Highway Code. The AV is unable to stop in time. Text appears to inform participants that the SDC could not stop in time and hit the pedestrian, who sustained minor injuries (as in Experiments 1 and 2). The full series of events 1–5 are depicted in Figure 9.

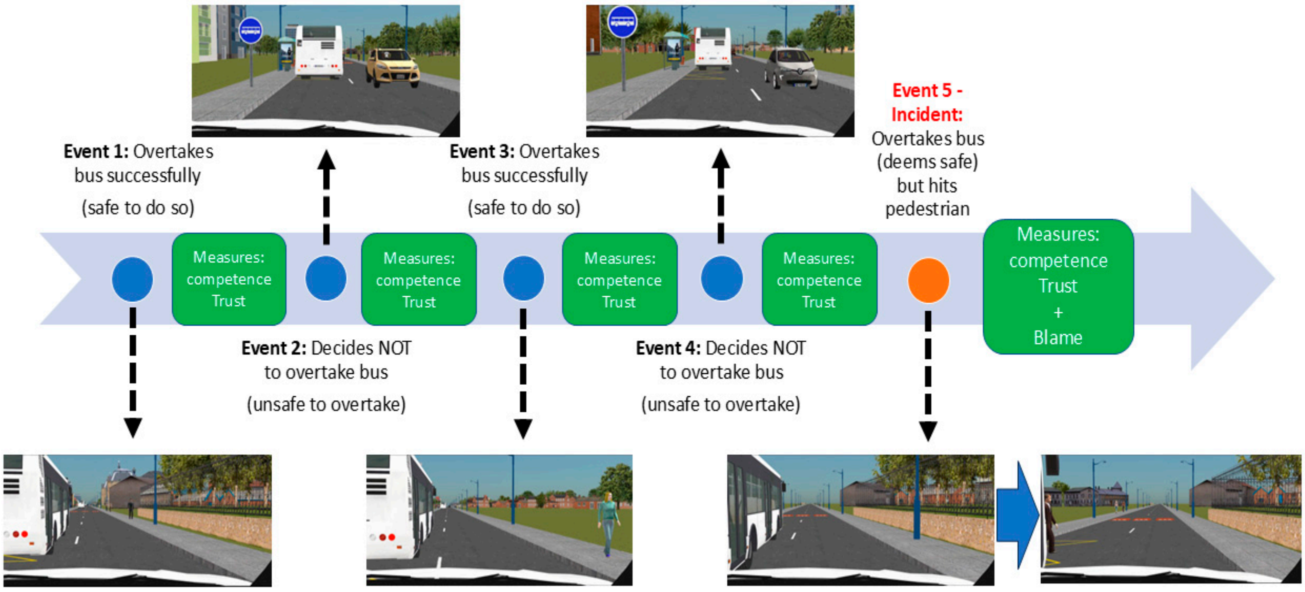

Event 1: Overtakes bus successfully
(safe to do so)
Measures: competence Trust
Event 2: Decides NOT to overtake bus
(unsafe to overtake)
Measures: competence Trust
Event 3: Overtakes bus successfully
(safe to do so)
Measures: competence Trust
Event 4: Decides NOT to overtake bus
(unsafe to overtake)
Measures: competence Trust
Event 5 - Incident: Overtakes bus (deems safe) but hits pedestrian
Measures: competence Trust + Blame

Figure 9. Event sequence in Experiment 3.

Following each of events 1–4, participants first rated trust using the TiAS. They then answered questions on the competence of IA using competence questions from the RoSAS (reliable, competent, knowledgeable, interactive, responsive, capable), followed by the question on the perceived riskiness of the AV. Then, an attention check multiple choice question was asked about the key actions taken by the AV after the bus had stopped. Following critical Event 5, questions were asked in the following order: TiAS (for consistency with events 1–4), single-item trust, blame on the AV and pedestrian, RoSAS (all three dimensions and questions: *competence*-reliable, competent, knowledgeable, interactive, responsive, capable; *warmth*-organic, sociable, emotional, compassionate, happy, feeling; and *discomfort*: awkward, scary, strange, awful, dangerous, aggressive), STS-AD, and finally perceived riskiness. A further attention check question was asked, followed by free text comments.

Finally, participants' attitudes (trust in and intention to use) towards AVs were measured twice in the same fashion as in experiments 1 and 2–i.e., before the experimental scenario started and at the end of the experiment before the free text comments option.

4.3. Experiment 3 Results

Trust

TiAS (Figure 10): Initially, trust across all five events is considered. Trust in the AV increased from Event 1, to 2, to 3, to 4 with a sharp decrease after Event 5, within the robot agent and non-robot agent conditions, although with higher means following events 2–5 in the non-robot agent conditions (Figure 10). TiAS item reliability was very high (α = 0.90–0.93). A 3-way ANOVA (robot presence × dialogue style × event) revealed a significant main effect of the event, $F(4{,}744) = 202.18$, $p < 0.001$, $f = 1.04$, with trust increasing from event 1 to 2, event 2 to 3, event 3 to 4 (all $ps \leq 0.001$) and decreasing from event 4 to 5 ($p < 0.001$). There were significant two-way interactions between agent embodiment and event, $F(4{,}744) = 3.01$, $p = 0.018$, $f = 0.13$, and dialogue style and event: $F(8{,}744) = 2.69$, $p = 0.006$, $f = 0.17$. There was also a significant three-way interaction between agent embodiment, dialogue style and event, $F(8{,}744) = 2.92$, $p = 0.003$, $f = 0.18$. Across events 1–4, there was a significantly greater increase in trust between events 2 (no overtake attempt) and 3 (second successful overtake attempt) in the robot agent conditions compared to the no-robot agent conditions ($ps < 0.05$), although no other differences were significant. After Event 5, trust fell significantly greater in the robot agent conversational style condition than in the speech-only (non-robot) conversational style condition ($MD = 13.06$, $p = 0.01$).

Focussing on Event 5 only (as in Experiments 1 and 2), a two-way ANOVA revealed non-significant main effects of agent embodiment, $F(1{,}186) = 3.30$, $p = 0.07$, dialogue style, $F(2{,}186) = 0.30$ $p = 0.74$, and a non-significant interaction, $F(2{,}186) = 2.41$, $p = 0.09$.

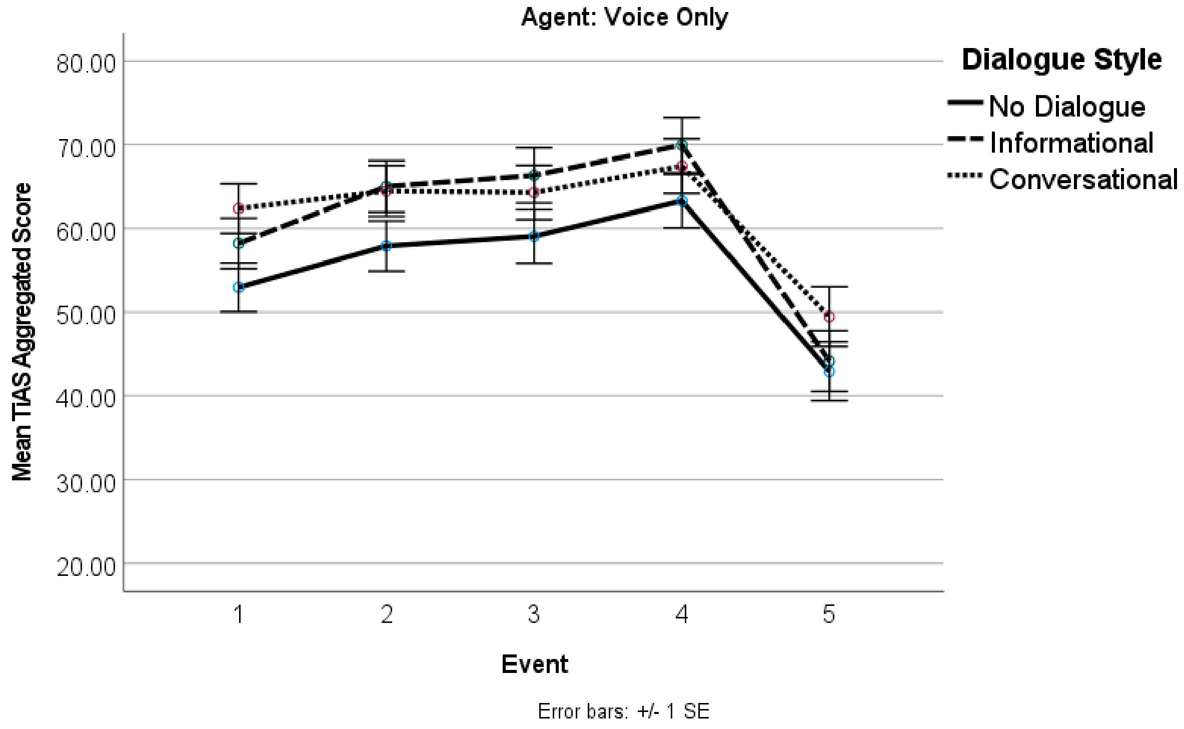
Single Measure (Figure 11): No significant main effects were found for agent embodiment, $F(1{,}186) = 3.13$, $p = 0.08$ or dialogue style, $F(2{,}186) = 0.13$, $p = 0.88$. There was a non-significant interaction, $F(2{,}186) = 0.66$, $p = 0.52$.

STS-AD (Figure 11): Item reliability was high (α = 0.85). No significant main effects were found for agent embodiment, $F(1{,}185) = 0.32$, $p = 0.57$, or dialogue style, $F(2{,}185) = 1.01$, $p = 0.37$, and there was a non-significant interaction, $F(2{,}185) = 0.59$, $p = 0.55$.

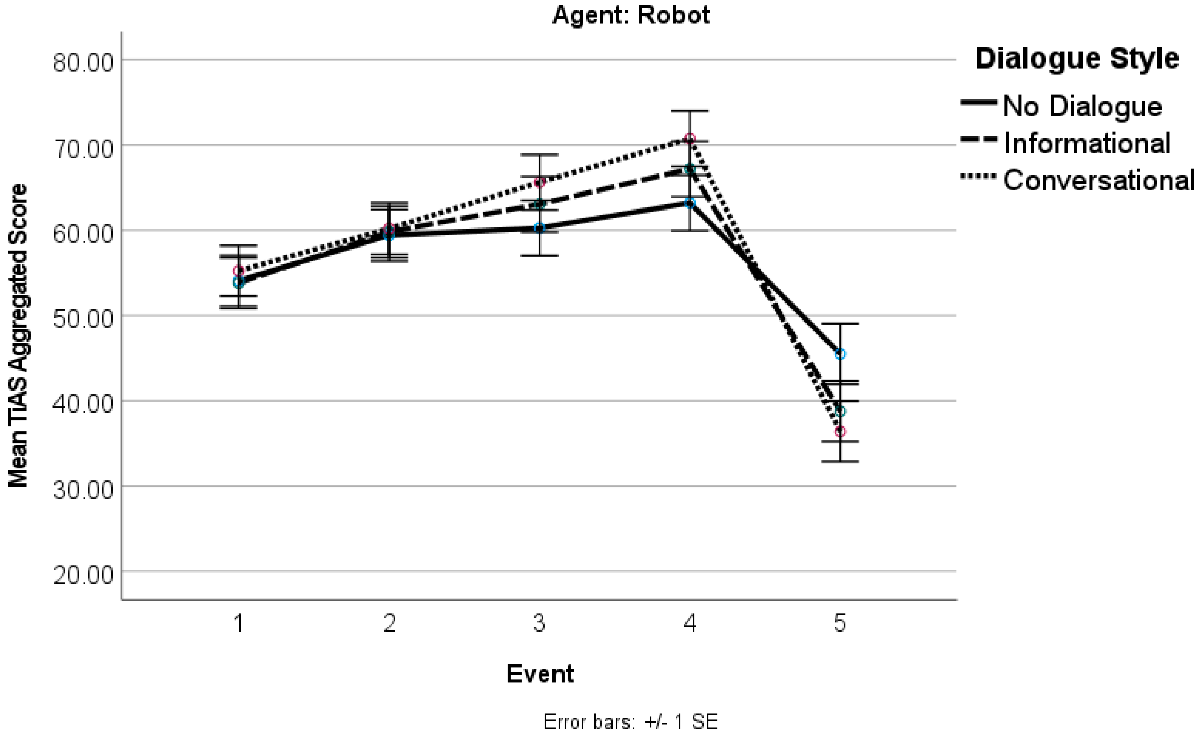
Blame (Measured after Event 5)

On the AV (Figure 12): There were non-significant main effects of agent embodiment: $F(1{,}185) = 1.85$, $p = 0.18$, or dialogue style: $F(2{,}185) = 0.63$, $p = 0.54$, and a non-significant interaction, $F(2{,}185) = 0.40$, $p = 0.67$.

On the pedestrian (Figure 12): There were non-significant main effects of agent embodiment: $F(1{,}185) = 0.003$, $p = 0.96$, and dialogue style: $F(2{,}185) = 0.45$, $p = 0.64$, and a non-significant interaction, $F(2{,}185) = 0.24$, $p = 0.78$.



(**a**)



(**b**)

Figure 10. Mean TiAS across agent embodiment and dialogue conditions after Events 1–5 ((**a**) Voice Only, (**b**) Robot). Error bars are ±SE.

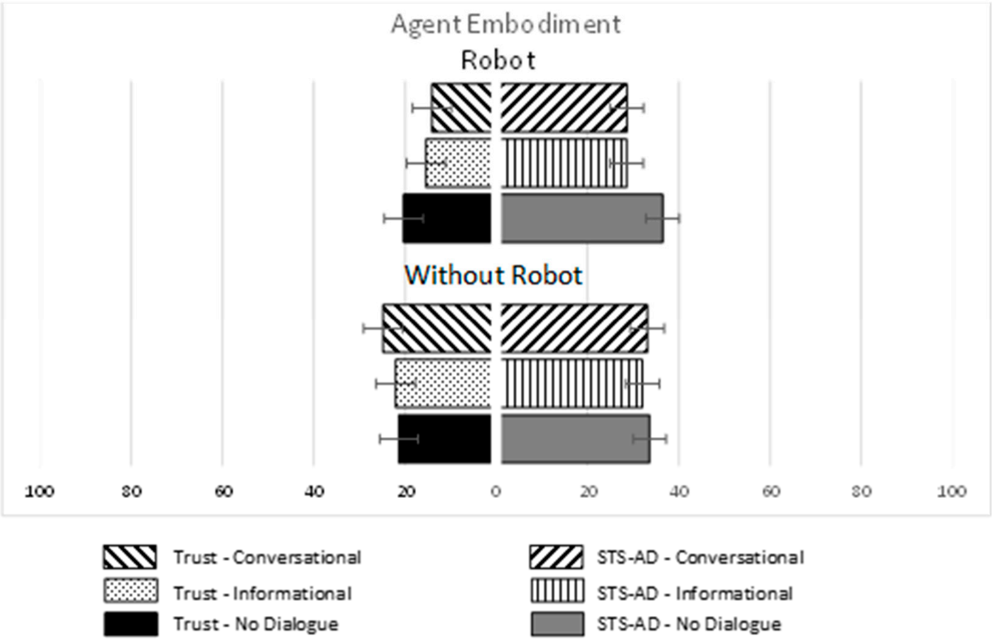


Figure 11. Mean single measure and STS-AD trust ratings across agent embodiment and dialogue conditions. Error bars are ±SE.

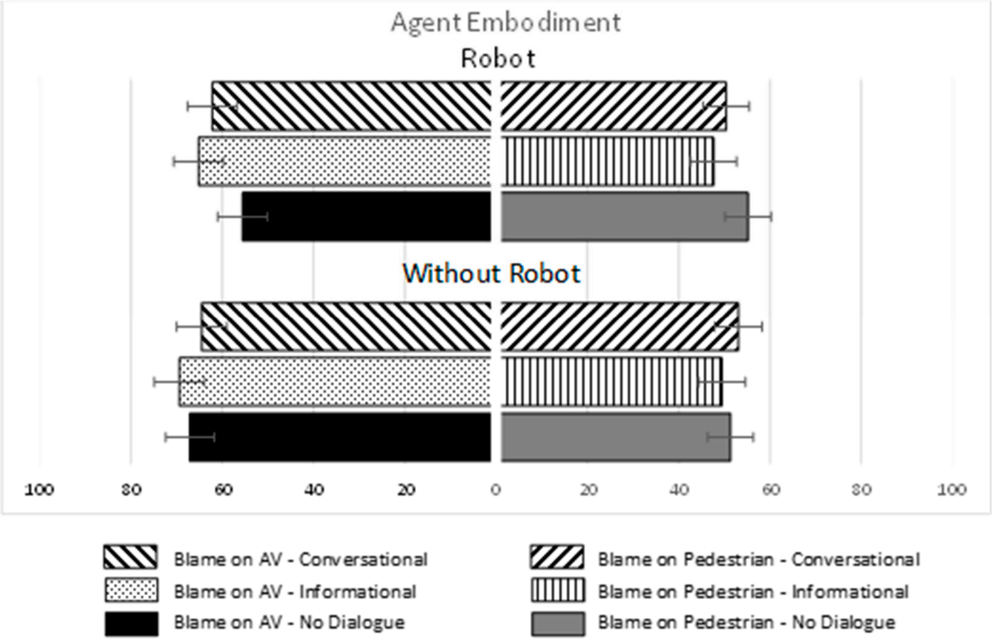
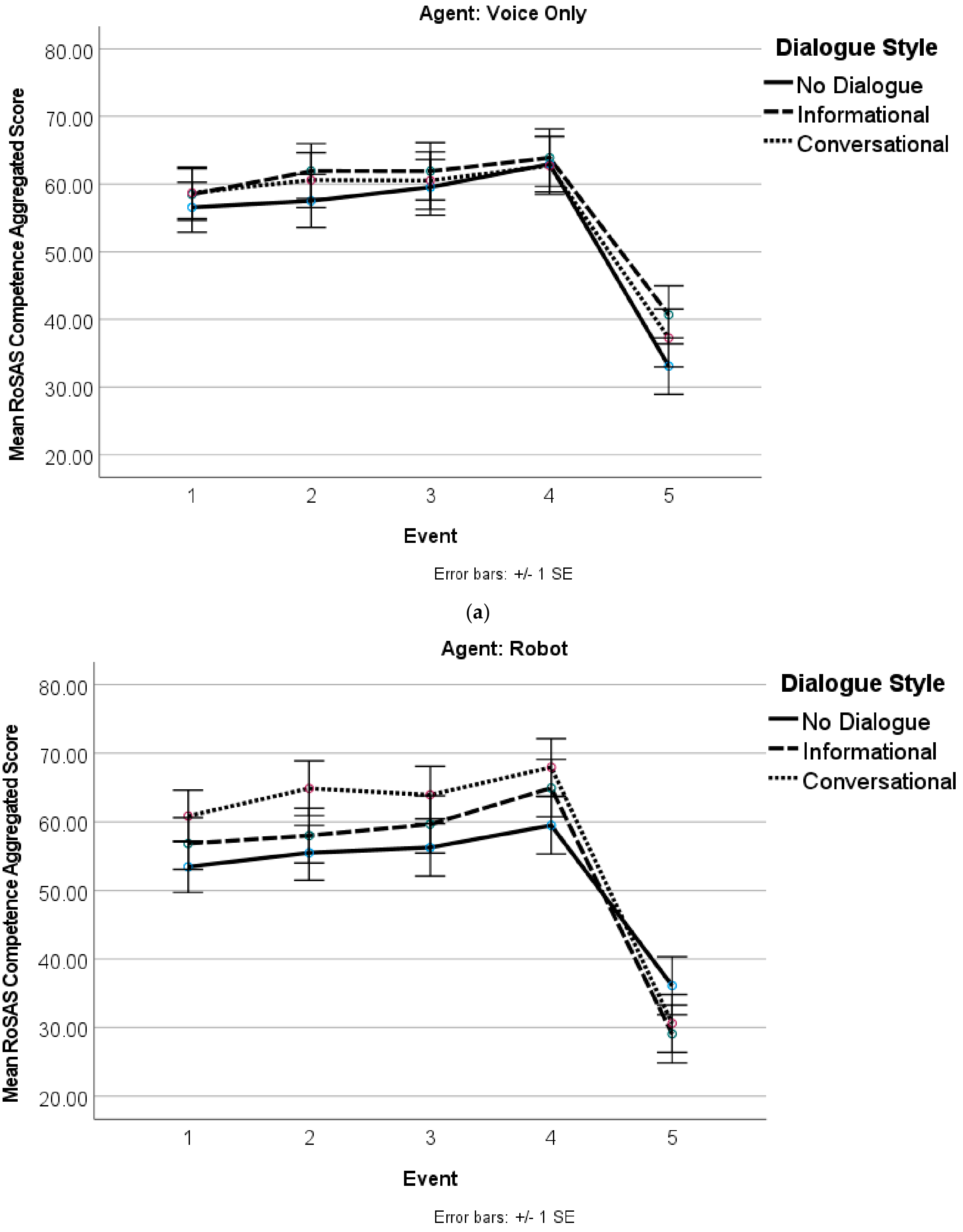


Figure 12. Mean levels of blame on the AV and the pedestrian across agent embodiment and dialogue conditions. Error bars are ±SE.

Social Attribution Measured using the RoSAS

Competence (Measured after all events, Figure 13): A three-way ANOVA revealed a significant main effect of the event, $F(4,740) = 212.72$, $p < 0.001$, $f = 1.07$ and a significant three-way interaction, $F(8,740) = 2.47$, $p = 0.012$, $f = 0.16$. Perceived competence significantly increased between events 1 and 2 ($p = 0.004$) and 3 and 4 ($p < 0.001$) but not 2 and 3 ($p = 0.99$). By Event 4, and with conversational dialogue, the robot agent condition was perceived as more competent than the speech-only condition, but this difference is non-significant ($MD = 5.24$), $p = 0.38$. Trust plummets after Event 5, but the direction of the difference at Event 4 is reversed: the non-robot agent condition is rated as more competent ($MD = 6.64$), but the difference is non-significant (p = 0.27). In the informational dialogue condition, both agent conditions have similar competence ratings after event 4. However, after Event 5, the non-robot agent speech-only condition is perceived as more competent than the robot agent condition (*MD* = 11.16), but the difference is marginally non-significant (p = 0.056). In the no dialogue condition, while competence ratings fall between events 4 and 5, there are no differences between agents.



(**a**)

(**b**)

Figure 13. Mean ratings of RoSAS competence across agent embodiment and dialogue conditions after Event 1–5 ((**a**) Voice Only, (**b**) Robot). Error bars are ±SE.

Warmth (Measured after Event 5, Figure 14): Reliability between items was high ($\alpha = 0.83$). There was a main effect of dialogue style, $F(2,185) = 4.68$, $p = 0.010$, $f = 0.23$, with higher warmth in the conversational than no dialogue condition ($MD = 6.74$), $p = 0.009$. There was a non-significant main effect of robot agent embodiment, $F(1,185) = 0.13$, $p = 0.72$, and a non-significant interaction, $F(2,185) = 0.001$, $p = 0.99$.

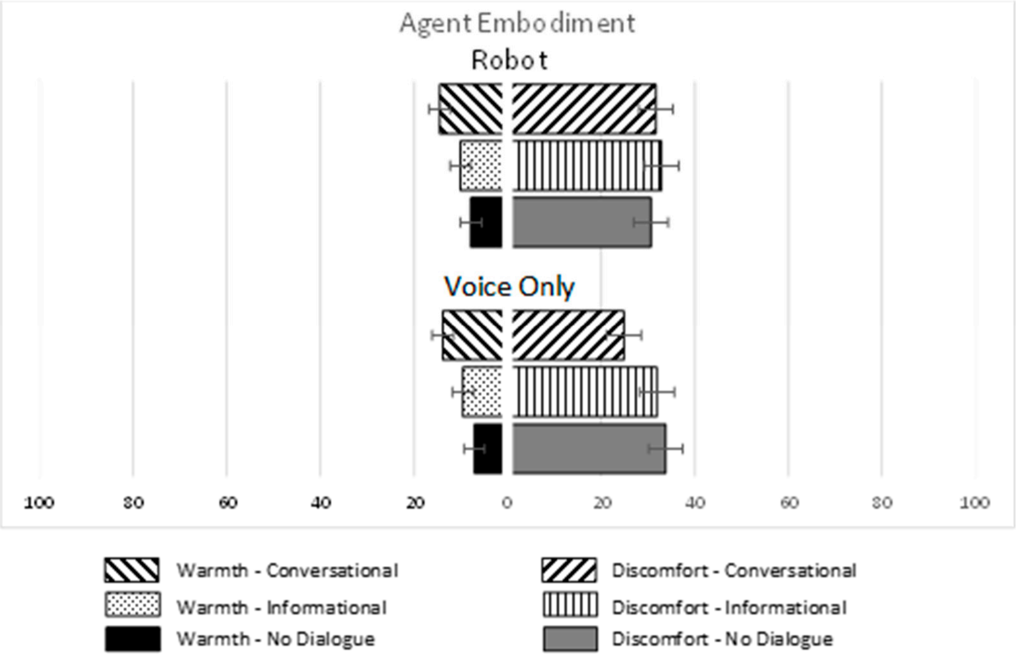


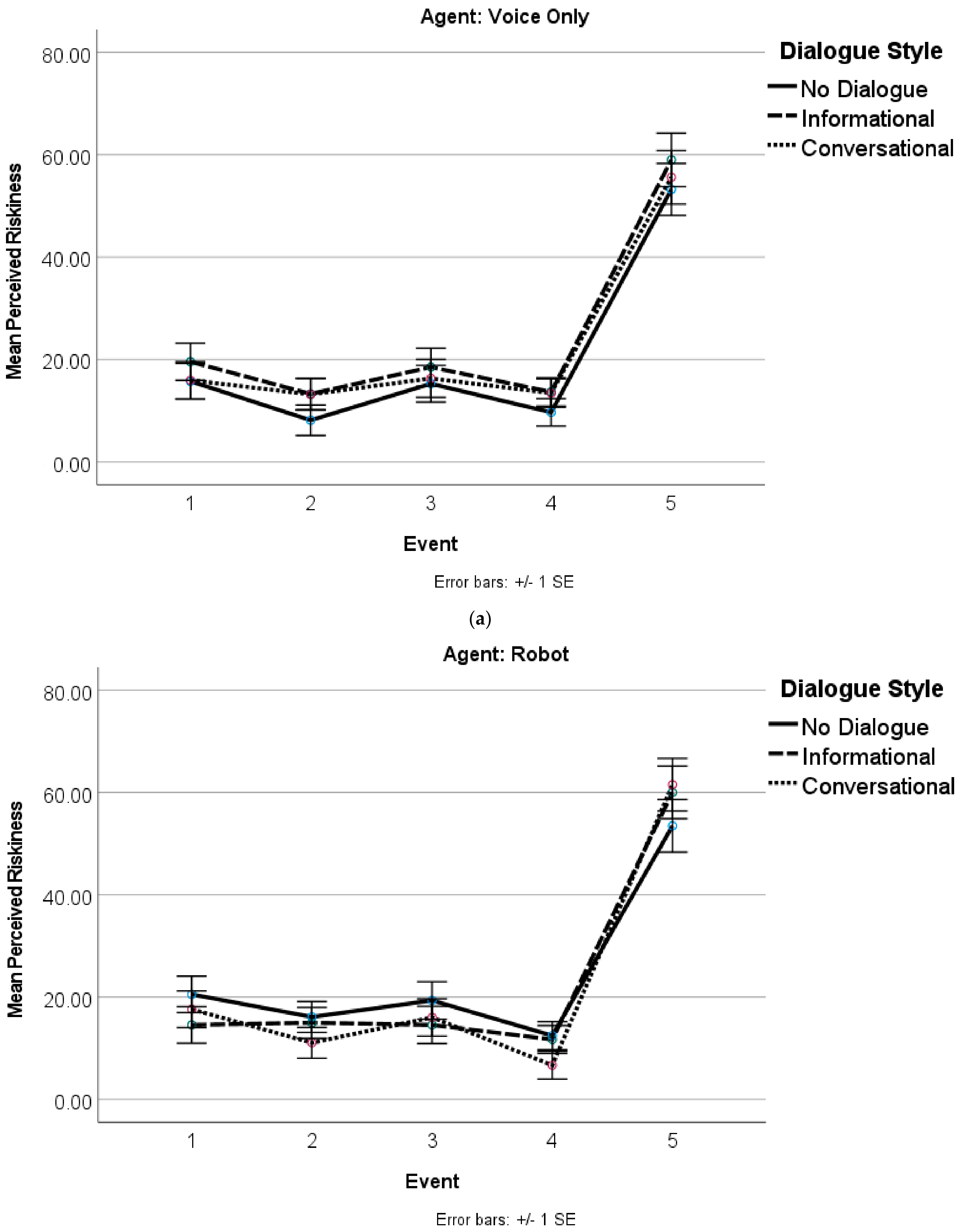
Figure 14. Mean ratings of RoSAS warmth and discomfort across agent embodiment and dialogue conditions after Event 5. Error bars are ±SE.

Discomfort (Measured after Event 5, Figure 14): Reliability between items was high ($\alpha = 0.83$). There was a non-significant main effect of agent embodiment, $F(1,185) = 0.27$, $p = 0.60$, and dialogue style, $F(2,185) = 0.81$, $p = 0.45$, and a non-significant interaction: $F(2,185) = 0.93$, $p = 0.40$.

Perceived Risk (Measured after all events): Mean risk ratings were very low (mostly less than 20/100) after events 1 to 4 in both the robot agent and non-robot agent conditions, although they increased after Event 5, surpassing 60/100 in some conditions (Figure 15). There was a significant main effect of the event, $F(4,740) = 271.70$, $p < 0.001$, $f = 1.21$, with significantly lower risk ratings after events 2 and 4 than 1 ($ps < 0.01$), and 4 than 3 ($p < 0.001$). There was a significant and marked increase in risk ratings between events 4 and 5 ($MD = 45.86$, $p < 0.001$). No other significant main effects were found (all $Fs < 1$), and none of the interactions were significant ($Fs < 1$).

Attitudes Towards AVs Pre- Event 1 and Post- Event 5

Consistent with experiments 1 and 2, there was a significant main effect of Stage on trust in AVS in general, $F(1,184) = 16.14$, $p < 0.001$, $f = 0.30$. Overall, after experiencing the extended SSGA with five events (including the collision after event five), trust in AVs decreased ($M = 26.58$, $SD = 25.49$) compared to before ($M = 32.84$, $SD = 27.20$). Note that the effect size was smaller in the current experiment, possibly due to the modified scenario featuring a series of successful manoeuvres before the collision event. As in experiments 1 and 2, there were non-significant main effects of agent and dialogue style, $F(1,184) = 2.04$, $p = 0.16$, and $F(1,184) = 0.50$, $p = 0.61$, respectively, and no significant interactions. The results were similar for the intention to use an AV. Only the main effect of Stage was significant, $F(1,184) = 45.12$, $p < 0.001$, $f = 0.50$; decreasing at the end ($M = 21.71$, $SD = 25.38$) compared to before ($M = 32.58$, $SD = 30.82$). All other main effects and interactions were non-significant ($ps > 0.05$).



(**a**)

(**b**)

Figure 15. Mean ratings of perceived riskiness across agent embodiment and dialogue conditions after Event 1–5 ((**a**) Speech Only, (**b**) Robot). Error bars are ±SE.

4.4. Experiment 3 Discussion

There were four main aims of Experiment 3. First, to give participants experience of the AV negotiating overtake attempts (by determining when conditions were safe or not) and thereafter examining effects on trust and some other measures prior to the critical (accident event). Second, to explore whether any increase in trust during these first four events (two overtakes, two non-overtakes) was higher in the robot agent embodiment condition (particularly with conversational dialogue style) than in the non-robot agent embodiment event, similar to [40,42,43]. The third aim was to examine whether these potential effects extended to trust and blame ratings after the fifth critical accident event with more robust and consistent findings than in Experiment 2. Finally, we included a measure of social attribution (RoSAS) to measure attitudes towards the IA agents that varied by embodiment (physical humanoid robot versus voice-only system).

TIAS ratings revealed consistent increases in trust across the first four events, which then plummeted after the critical Event 5. An interaction provided some evidence for the prediction linked to the second aim, in that trust (measured via the TAIS) was higher within the robot agent than non-robot agent between some (but not all) events (events 2 and 3). Examining post-event 5 TIAS ratings separately as well as STS-AD ratings, there were no differences between conditions or interaction, as was the case in Experiments 1 and 2. This was also the case for the single-item trust question, despite Experiment 2 findings that the robot agent was trusted more than the voice-only agent. Thus, our modified paradigm developed to increase trust in the AV through experience was successful (similar to [42,43]) and, to some extent, more so in the robot agent condition. However, it did not result in higher trust in the robot agent condition (when coupled with or without conversational dialogue) after an accident outcome.

Interestingly, blame ratings on the AV were quite high (in all but one condition >60/100), but blame on the pedestrian ratings were also higher than 50/100. In neither case did ratings significantly differ due to agent embodiment nor dialogue style. As with trust, it seems that blame assignments using our developed paradigm were not affected by the presence of a robot over a voice-only agent or by dialogue style. We also considered social attribution and, in particular, attitudes toward the IA across three dimensions using the RoSAS: competence (after all events), warmth (after Event 5) and discomfort (after Event 5). While perceived competence in an IA significantly increased between events 1 and 2 and 3 and 4, there were no statistical differences between agent type or effects of dialogue style. After Event 5, there was only a significant main effect for warmth towards the IA between the conversational and no dialogue conditions with no effects of agent type on this RoSAS dimension or for discomfort.

Finally, and as in experiments 1 and 2, participants general trust in and intention to use AVs significantly decreased after having experienced the entire scenario (after the five events vs. before part/event 1); again, most likely because the outcome of Event 5 entailed a collision. However, the effect size was smaller than in previous experiments giving an indication that the reductions observed were not as pronounced as in experiments 1 and 2, where participants did not have experience of the AV negotiating manoeuvres without any negative outcome, at least until the final critical event. As in experiment 1, agent and dialogue style had no effect on post-experiment trust in or intention to use AVs.

Taken together, the findings of Experiment 3 indicate that while providing participants with experience of a highly autonomous vehicle successfully negotiating overtake opportunities prior to an accident can increase trust and to some degree more so for a humanoid robot IA than a speech-only system, there are no other notable effects of either agent embodiment or dialogue style after an accident event. Perhaps, as in Experiment 1, the AV was perceived as being too assertive, although this time due to multiple overtake attempts, with the final attempt ending with an accident. A free text quote from two participants fits with this possibility:

> *P20. '[It] Seems to me that the risk the AV failed on was one that it should have been more aware of. Its priority in the journey was overtaking, which in a high street with many pedestrians and oncoming vehicles was maybe too aggressive?'*

> *P32. 'Seems to me that the risk the AV failed on was one that it should have been more aware of. Its priority in the journey was overtaking, which in a high street with many pedestrians and oncoming vehicles was maybe too aggressive?'*

However, these were the views of only a small number of participants who perceived the AV to be performing in a risky manner, yet perceived riskiness of the AV decreased in general across the first four events and, as expected, markedly increased after the final critical event. It could then be that expectations on the safety performance of the AV were heightened due to the positive 'incident free' first four events and then damaged by the critical accident event at the end of the scenario. Another participant noted

> *P63. 'The last scenario was the one that most worries me. For instance, if I am behind a school bus that pulls in before overtaking, I factor in that children may run in front of the bus and I adjust my actions accordingly. How do you programme human type intuition into an AV?'*

Participants may have felt that the chances of an accident occurring were increasing, given that the AV seemed too intent on trying to overtake parked buses. Thus, we have demonstrated under other conditions (i.e., with experience of the AV successfully negotiating overtake attempts) where the use of a robot IA and a more personable dialogue style does not have an impact on trust or blame and where attitudes towards a humanoid robot IA are the same as a speech-only system. Perhaps then, the benefits of a robot IA and, e.g., a conversational dialogue style are restricted to situations where a highly autonomous vehicle performs without experiencing an accident such that findings like those from [40,42] are representative of when AVs and other road users are performing optimally. Such situations are not guaranteed, and thus, our novel evidence points towards robot IAs not necessarily representing an intervention that can help minimise loss of trust and blame in an AV in the event of an accident not directly caused by the AV but instead due to another agent violating the Highway Code with the AV not being able to intervene in time. Clearly, then, further research is needed to evaluate whether humanoid robot IAs constitute a practicable and generally positive solution, as has been suggested by others.

5. General Discussion

The efficacy of AVIAs varying by embodiment (humanoid robot, speech only, nothing/control) and dialogue style (conversational, informational, no speech/control) on participant ratings of trust in and blame on a highly autonomous vehicle following an accident were investigated. The accident occurred when the AV was overtaking a parked bus after deeming it safe to do so based on traffic conditions and Highway Code rules. When committed to the overtaking manoeuvre, a pedestrian violated the Highway Code by walking in front of the bus and then onto the road without there being time for the AV to stop and avoid a collision. Such a scenario is entirely feasible [9] when highly autonomous vehicles are deployed and programmed to make decisions such as overtaking parked vehicles to, e.g., support efficient traffic flow and, more generally, when other road users violate the Highway Code.

In Experiment 1, the AVIA communicated intentions to try and overtake the bus while in motion and even before it had come to a halt at a bus stop. In Experiment 2, the AV was not looking for opportunities to overtake the bus until it came to a halt at a bus stop, and the AVIA dialogue reflected this throughout the journey. In Experiment 3, the AV negotiated a parked bus scenario four times (overtaking twice–deemed safe to do so; did not overtaking twice–deemed unsafe to do so) before the fifth critical overtake attempt that culminated in the same accident outcome as in Experiments 1 and 2. Participants also rated the perceived riskiness of the AV in Experiments 2 and 3, as well as attitudes towards the IAs in Experiment 3.

Based on findings from previous research [40,42,43], it was predicted that trust in the AV would be higher or lower with an anthropomorphic embodied humanoid robot IA than a speech-only system, markedly so with a conversational compared to an informational dialogue style. A key and novel factor within the current experiments was the critical incident accident outcome. These predictions were not supported in Experiment 1 across three measures of trust (single item, TIAS, and STS-AD) with only one main effect of dialogue style (higher trust in the informational dialogue style condition than no-speech condition for the single-item measure only). Free text comments from some participants indicated a perception that the IA at times communicated intentions (to find opportunities to overtake the bus when it was in motion) that were assertive or even aggressive, and this may have impacted predicted findings. However, no measure of perceived riskiness was included. In Experiment 2, the dialogue was adapted such that there were no communicated intentions to overtake the bus until it had stopped. It was found that perceived risk was lower in the physical agent embodiment conditions. Furthermore, and as predicted, the physical agent embodiment conditions resulted in higher trust than in the non-physically embodied speech-only condition (HTrust-1), and there was a marginally non-significant interaction indicating the highest trust in the conversation speech robot present condition compared to the conversational speech-only condition (HTrust-3). However, this was only the case for the single trust item with no such effects with TIAS or the STS-AD. An interaction was also found for blame: it was higher on the AV in the robot agent no dialogue condition than both robot agent dialogue conditions and higher than in the no-robot agent conditions. This interaction was not supportive of HBlame-3 and instead revealed that a 'silent' robot IA within an AV, results in higher blame in the event of an accident–a situation that could occur if, e.g., the IA malfunctions or is switched off.

Despite some findings in Experiment 2 fitting with the predictions, many did not. This could have been because participants had limited opportunity to experience the AV performing without an incident and given that all measures were taken after the accident had occurred. Indeed, as highlighted above, measures within previous studies were taken across scenarios without an accident outcome [40,42,43]. Thus, in Experiment 3, the paradigm was further developed such that participants experienced the AV driving behind multiple moving buses and negotiated four overtake attempts (including two fully committed to and without incident) before the fifth and critical accident event. As predicted, trust ratings (measured via TIAS questions) in the AV consistently and significantly increased across events before plummeting after the critical event, and there was an interaction with higher trust in the robot IA between some (albeit not all) of the earlier non-critical outcome events. However, contrary to predictions, there were no differences in trust ratings on any measure after the critical accident outcome Event 5. Furthermore, blame on the AV ratings was quite high, and again, there were no differences due to embodiment or dialogue style. Even in terms of social attribution measured using RoSAS, there were some promising findings prior to the critical accident event (i.e., significant increases in the perception of IA competence) but not between the robot and non-robot IA agents and with no differences due to IA embodiment or dialogue style after the critical event.

This is further confirmation–together with the findings from Experiment 1 and some from Experiment 2–that anthropomorphising a highly autonomous vehicle with an embodied humanoid robot IA does not consistently lead to higher trust in the AV than in conditions a non-embodied speech-only IA following a critical event, even when the accident outcome was ultimately caused by the actions of the other agent (in this case a pedestrian violating the Highway Code). Furthermore, dialogue style does not seem to have any consistent effect either. These findings stand in contrast to previous studies [40,42,43], and thus, we must warn of the many potential pitfalls of assuming that robot agent IAs will be universally beneficial to user trust in AV technology. They may, in fact, damage trust, in some circumstances at least. As suggested by [30], making autonomous systems more human-like may have a counterintuitive effect of increasing user expectations–with higher resultant performance expectancies. Even though the IAs manipulated and tested within the current experiments were not in any way linked to the vehicle controls (much like a satellite navigation system in most vehicles at least), participants may have expected an intervention to avoid the accident outcome, especially given the IAs awareness of actions and intentions during the journeys, communicated in real-time. For example, some participants mentioned being disconcerted about the robot turning to look at them to speak, despite our clear framing of the system in the car as only being an IA. One participant in Experiment 3 noted:

> *P88. 'I found it a bit disconcerting when the AV took its eyes off the road, even though I know logically its 'eyes' aren't actually doing the seeing/sensing.'*

Such performance expectations might have been further exacerbated in Experiment 3, given multiple experiences of 'the system' performing 'optimally' before the final critical collision outcome. Overall, what we have demonstrated provides further evidence of the AV Capability Hypothesis put forward by [30]. Linked to this is the general perception that AVs should perform the task of driving more optimally than humans (e.g., [67,68]). Linked to this are the findings regarding trust in and intention to use AV technology at the end of each experiment compared to the beginning and before having experienced the scenarios and negative events. While trust in and intention to use AV technology ratings were low before experiencing the scenarios, they reduced significantly after, most likely due to the final critical (collision) event in each case. Participants may have very high expectations of AVs and would not expect one to be involved in a collision even when the events leading up to it involve a pedestrian breaking the Highway Code/Rules of the Road and the AV not being able to stop in time. In fact, participants may have believed that the AV should have been able to avoid the collision, perhaps by stopping in time or taking another course of action–e.g., swerving and/or even mounting the kerb. This is an area that requires much future research attention–i.e., discretionary actions that could and perhaps should be performed by AVs to try and avoid accidents and how potential users perceive and judge these in terms of trust and possible blame if the discretionary/emergency action in itself results in a negative outcome (e.g., injuring the user of the AV or another road user(s) that AV may not have been able to detect when performing the action).

Limitations and Future Directions

Our experiments are not without limitations. There are further considerations that need to be addressed before we can, with a higher degree of confidence, recommend that embodying and humanising AVIAs will not be universally beneficial, including in the event of AV incidents and accidents. First, the experiments employed the SSGA method successfully used in past studies [17,30], with videos created within a driving simulator–including the IAs (robot and speech only)–and played to participants via experiments deployed online. Some previous research has provided evidence that video displays of a robot (telepresence) are not as effective as physical presence [69,70]. Future research should consider manipulations of agent embodiment and dialogue style under in-person testing conditions—i.e., with the IAs physically present, unlike the online experiments employed within the current experiments and related studies [40,42,43]. Second, only one type of humanoid robot was employed: a Nao with human-like features, and because it was used in similar previous studies [40,42,43]. However, people will often make judgements about a robot based on its appearance [71]. The Nao robot can be considered as having a childlike appearance, and the appearance of arms and legs may have caused participants to assume that it had more control [72] as related to the expectations and capabilities points discussed earlier. Additionally, the positioning and actions of the embodied IA need to be further considered. Within the current experiments, the Nao robot was positioned as if it was on the vehicle dashboard, facing forward most of the time and turning towards the participant when delivering dialogue, based on research such as that by [73]. In this case, however, the robot was acting as a driving assistant, and a person still had control of the car. It may be less appropriate when a user is not in control of the car. It would be valuable to test whether effects differ if the robot IA is positioned such that it is never viewing the road to reinforce its agency as an assistant and perhaps create a less ambiguous situation where it could be perceived as having at least some control over the functions of the AV. Alternatively, and to also address the potential issue of an IA robot having arms and legs, a multi-directional gaze-driving agent system such as NAMIDA might be beneficial [74]. However, such a device could further exacerbate user perception that the IA is in some way 'all seeing, all knowing' and should thus be able to intervene or trigger a system that can intervene in the event of a potential incident involving an AV. The use of weak robots (also known as human-dependent social robots) could also be considered [75]. The design of such devices moves away from the idea that the robot can perform many things and instead capitalises on the idea that its inabilities will encourage more human interaction and fewer expectations. A weak robot IA may also be considered more similar to a voice-only IA but could have higher physical embodiment.

There were also other aspects of an IA that could be manipulated that may affect trust and blame in the event of an AV being involved in an incident and/or accident. Ref. [76] categorise design variables of in-vehicle agents into four design spaces as part of a systematic literature review: Agent characteristics (variables such as agent type and appearance); Information presence (what type of information and how transparent it is); Verbal characteristics (characteristics of speech (e.g., male, female, genderless), style of speech); and Non-verbal characteristics (agent attitude, body gestures, expressions and customisation features such as matched personality). Another limitation to consider relates to cultural similarities and differences. The current experiments involved UK participants only. Previous studies have established differences in acceptance of Avs across cultures [77] and moral reasoning relating to AVs [78]. It is entirely plausible that trust and blame attribution towards AVs will differ when AVs are involved in incidents and/or accidents and that IAs and perceptions of them will also vary across countries and cultures–potentially impacting trust and blame.

Our measures of trust, blame and attitudes towards the IA also need to be considered. We employed three different measures of trust: a single measure based on and successfully used by [17,30], TiAS [42]; further developed and successfully used by [43] and STS-AD [58]–specifically developed and validated for autonomous driving research on trust). Within the field of human-robot interaction, it is often the case that two types of trust are important: performance and relational [79,80]. Performance trust relates to how well someone believes an agent can achieve the goals assigned to it. Relational trust refers to trust on a social level, in particular, a person's behaviour in relation to the agent. The TiAS and single measure items were aimed more at relational trust, and the STS-AD was used to measure performance and trust in the specific situation. As we did not manipulate the actual performance of the AV, the only effects on performance trust are likely to be residual secondary effects from a more positive perception of the competence of the AV. This can perhaps explain why no effects were found with the STS-AD. Second, our inclusion of a single blame question (focused on the AV and the pedestrian) is not unusual. For example, ref. [81] adopted a similar approach, albeit with more agents to which blame could be assigned (e.g., vehicle, pedestrian, manufacturer, Government). However, they also included an open-ended question probing motivation for participant blame ratings as well as to check for random responses. Such an approach could be considered in future research involving AV IAs. Third, we adopted the RoSAS [66] to measure attitudes towards a humanoid robot IA as well as a non-robot speech-only IA. Other researchers have adopted other methods to assess attitudes toward vehicle assistants, including the Kano questionnaire with dimensions on, e.g., product features (e.g., 'how would you feel if the automotive assistant could alert you to potential hazards?') and satisfaction improvement [82], although the RoSAS seemed to be the most appropriate within the current experiments with some promising findings in terms of competence although very low ratings for warmth and dis(comfort).

The IAs employed within the experiments used speech as the main form of communication, given that passengers within highly autonomous vehicles may not be expected to be visually engaged with them or, indeed fully with the driving conditions unfolding before them. However, visual information regarding manoeuvres, e.g., could also be provided as an additional source of informational assistance. IAs could relay information about the driving environment from real-time video footage to the user, which could enhance their situation awareness regarding elements of the journey, such as when manoeuvres that have higher levels of risk are being considered and executed. To achieve this, there is a requirement for real-time perception of streaming videos to react to manoeuvres made by an SDC. Video Object Detection (VOD) methods, which focus on detecting and tracking objects in video frames, have traditionally been used. An autopilot perception task called streaming perception has also been proposed by [83], which, unlike traditional VOD methods, allows for real-time perception. However, existing methods are still limited when it comes to managing complicated changes in motion, and subsequently, newer methodologies such as LongShortNet [84] and the DAMO-StreamNet framework [85], which offers a solution for real-time perception in autonomous driving have been proposed (see also [86]) and could be coupled with speech-based features of AV IA systems.

Finally, it is acknowledged that while well-powered, the current experiments involved UK participants only. Emerging research on public perceptions of AV safety [87], trust [88] and acceptance [89] has revealed differences in these measures across countries that can be attributed to factors such as cultural differences. Additionally, the appetite for AV technology in terms of acceptance will be higher in some countries, see [90] than in others, and this could be based on, e.g., how developed the traffic systems are as well as population factors such as age and when it expected or indeed enforced that citizens above a certain age will not be able to drive manual vehicles.

6. Conclusions

Caution should be exercised when attempting to anthropomorphise IAs within high AVs. Our findings reveal that an IA that was too persistent in its assertions (and being perceived as assertive) could further endanger trust following an accident involving an AV that was the fault of another agent. Ensuring that neither the IA nor the AV itself causes increased perceptions of risk must, therefore, be a consideration with some evidence of higher trust in a humanoid robot IA, markedly so when communicating using a first-person conversational rather than informational style. While providing opportunities for participants to learn to trust the AV via experience, its negotiating overtake attempts with an IA communicating intentions and actions was shown to increase trust; the type of IA (humanoid robot or speech-only system) did not seem to matter. However, providing such an experiential learning opportunity phase may have increased the perceived capabilities of the IA and the AV such that following an accident event; there were no differences in trust or blame due to IA embodiment or dialogue style. Future directions include testing the efficacy of our manipulations via in-person testing, considering robot agent type and other features such as placement and attributes (e.g., gender), and possibly considering other measures of trust and blame.

Author Contributions: Conceptualisation, C.D.W., Q.Z., P.L.M., V.E.K.M. and D.M.J.; Methodology, C.D.W., Q.Z., P.L.M., V.E.K.M. and D.M.J.; software, C.D.W.; validation, C.D.W., Q.Z., P.L.M., V.E.K.M., L.B. and T.K.; formal analysis, Q.Z., C.D.W. and P.L.M.; investigation, C.D.W., Q.Z., P.L.M., V.E.K.M. and D.M.J.; resources, C.D.W., Q.Z., P.L.M., V.E.K.M., L.B. and T.K.; data curation, C.D.W., Q.Z., V.E.K.M. and P.L.M.; writing—original draft preparation, P.L.M. (substantial), C.D.W., Q.Z. and V.E.K.M.; writing—review and editing, P.L.M. (substantial), Q.Z. and V.E.K.M.; visualisation, C.D.W., Q.Z., P.L.M. and V.E.K.M.; supervision, P.L.M. and D.M.J.; project administration, P.L.M. and D.M.J.; funding acquisition, P.L.M. and D.M.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the ESRC-JST project: Rule of Law in the Age of AI: Principles of Distributive Liability for Multi-Agent Societies. P.L.M. was the UK Lead Investigator, and Prof Tatsuhiko Inatani (Kyoto University) was the Japan Lead Investigator. ES/T007079/1-Rule of Law in the Age of AI: Principles of Distributive Liability for Multi-Agent Societies (ESRC and JST).

Institutional Review Board Statement: Full ethical approval was granted for the experiments reported in this paper from the Cardiff University School of Psychology Ethics Committee–Ref: EC.20.05.12.6041GA2. This was accompanied by a fully approved risk assessment–Ref: 1691755771_3780.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available via the ES/T007079/1-Rule of Law in the Age of AI: Principles of Distributive Liability for Multi-Agent Societies ESRC Data Archive. Please contact the corresponding author for details.

Acknowledgments: We thank our project collaborators for their invaluable discussions regarding all aspects of the research presented in this paper: Minoru Asada (Osaka University), Tatsuhiko Inatani (Kyoto University), Hirofumi Katsuno (Doshisha University), and Yoshiyuki Ueda (Kyoto University). We thank Arpit Patel for significantly contributing to the programming of the driving scenarios across all experiments. We also thank three anonymous reviewers of an earlier draft for their insightful comments and suggestions. Finally, we wish to dearly thank Dylan M. Jones, who we sadly lost too suddenly and too soon in 2022. Dylan was a Co-Investigator on the project (and so many others), an inspiration to us all, and a dear colleague and friend.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. World Health Organization. *Global Status Report on Road Safety 2023*; World Health Organization: Geneva, Switzerland, 2023; pp. 1–96.
2. National Highway Traffic Safety Administration. *Early Estimate of Motor Vehicle Traffic Fatalities in 2023*; Traffic Safety Facts—Report DOT HS 811 059; U.S. Department of Transportation: Washington, DC, USA, 2024; pp. 1–6.
3. Department for Transport. *Reported Road Casualties Great Britain, Annual Report 2023*; Department for Transport: Hastings, UK, 2024.
4. National Highway Traffic Safety Administration. *National Motor Vehicle Crash Causation Survey*; Report DOT HS 811 059; U.S. Department of Transportation: Washington, DC, USA, 2008; pp. 1–47.
5. SAE. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for on-Road Motor Vehicles*; SAE J3016_202104; SAE International: Warrendale, PA, USA, 2021.
6. Gao, W.; Jiang, Z.P.; Ozbay, K. Data-driven Adaptive Optimal Control of Connected Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 1122–1133. [CrossRef]
7. Zhao, J.; Wang, Z.; Lv, Y.; Na, J.; Liu, C.; Zhao, Z. Data-Driven Learning for H_∞ Control of Adaptive Cruise Control Systems. *IEEE Trans. Veh. Technol.* **2024**, 1–15. [CrossRef]
8. Fagnant, D.J.; Kockelman, K.M. Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations. *Transp. Res. Part A Policy Pract.* **2015**, *77*, 167–181. [CrossRef]
9. National Safety Council. *Estimating the Costs of Unintentional Injuries*; National Safety Council Report: Washington, DC, USA, 2012; pp. 1–2.
10. Trottenberg, P. *Treatment of the Value of Preventing Fatalities and Injuries in Preparing Economic Analysis—2011 Revision*; U.S. Department of Transportation: Washington, DC, USA, 2011.
11. Zhou, R.; Zhang, G.; Huang, H.; Wei, Z.; Zhou, H.; Jin, J.; Chang, F.; Chen, J. How would Autonomous Vehicles Behave in Real-world Crash Scenarios? *Accid. Anal. Prev.* **2024**, *202*, 107572. [CrossRef]
12. Susilawati, S.; Wong, W.J.; Pang, Z.J. Safety Effectiveness of Autonomous Vehicles and Connected Autonomous Vehicles in Reducing Pedestrian Crashes. *Transp. Res. Rec.* **2023**, *2677*, 1605–1618. [CrossRef]
13. 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]
14. Bansal, P.; Kockelman, K.M. Forecasting Americans' Long-term Adoption of Connected and Autonomous Vehicle Technologies. *Transp. Res. Part A Policy Pract.* **2017**, *95*, 49–63. [CrossRef]
15. Kaur, K.; Rampersad, G. Trust in Driverless Cars: Investigating Key Factors Influencing the Adoption of Driverless Cars. *J. Eng. Technol. Manag.* **2018**, *48*, 87–96. [CrossRef]
16. Kyriakidis, M.; Happee, R.; de Winter, J.C. 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]
17. Zhang, Q.; Wallbridge, C.D.; Jones, D.M.; Morgan, P.L. Public Perception of Autonomous Vehicle Capability Determines Judgment of Blame and Trust in Road Traffic Accidents. *Transp. Res. Part A Policy Pract.* **2024**, *179*, 103887. [CrossRef]
18. Penmetsa, P.; Sheinidashtegol, P.; Musaevr, A.; Adanu, E.K.; Hundall, M. Effects of the Autonomous Vehicle Crashes on Public Perception of the Technology. *IATSS Res.* **2021**, *45*, 485–492. [CrossRef]
19. Olaverri-Monreal, C. Promoting Trust in Self-driving Vehicles. *Nat. Electron.* **2020**, *3*, 292–294. [CrossRef]

20. 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]
21. Endsley, M.R. Toward a Theory of Situation Awareness in Dynamic Systems. *Hum. Factors* **1995**, *37*, 32–64. [CrossRef]
22. Endsley, M.R.; Kiris, E.O. The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Hum. Factors* **1995**, *37*, 381–394. [CrossRef]
23. Al-Saadi, Z.; Phan Van, D.; Moradi Amani, A.; Fayyazi, M.; Sadat Sajjadi, S.; Ba Pham, D.; Jazar, R.; Khayyam, H. Intelligent Driver Assistance and Energy Management Systems of Hybrid Electric Autonomous Vehicles. *Sustainability* **2022**, *14*, 9378. [CrossRef]
24. Aubeck, F.; Mertes, S.; Lenz, M.; Pischinger, S. A Stochastic Particle Filter Energy Optimization Approach for Power-split Tra-jectory Planning for Hybrid Electric Autonomous Vehicles. In Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 19 October–13 November 2020; pp. 1364–1369. [CrossRef]
25. Endsley, M.R. Situation Awareness in Future Autonomous Vehicles: Beware of the Unexpected. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*; Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y., Eds.; IEA 2018. Advances in Intelligent Systems and Computing; Springer: Cham, Switzerland, 2018; Volume 824. [CrossRef]
26. Zhao, J.; Lv, Y.; Zhao, Z. Adaptive learning based output-feedback optimal control of ct two-player zero-sum games. In *IEEE Transactions on Circuits and Systems II: Express Briefs*; IEEE: Piscataway, NJ, USA, 2022; Volume 69, pp. 1437–1441.
27. Zhao, J.; Jia, B.; Zhao, Z. Model-Free H_∞ Prescribed Performance Control of Adaptive Cruise Control Systems via Policy Learning. *IEEE Trans. Intell. Transp. Syst.* **2024**, 1–11. [CrossRef]
28. Liljamo, T.; Liimatainen, H.; Pöllänen, M. Attitudes and Concerns on Automated Vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, *59*, 24–44. [CrossRef]
29. Sweet, M.N.; Scott, D.M.; Hamiditehrani, S. Who will Adopt Private Automated Vehicles and Automated Shuttle Buses? Testing the Roles of Past Experience and Performance Expectancy. *Transp. Plan. Technol.* **2023**, *46*, 45–70. [CrossRef]
30. Zhang, Q.; Wallbridge, C.D.; Jones, D.M.; Morgan, P.L. The Blame Game: Double Standards Apply to Autonomous Vehicle Accidents. In Proceedings of the AHFE 2021 Conference on Human Aspects of Transportation, Virtual, 25–29 July 2021; Advances in Human Aspects of Transportation. Lecture Notes in Networks and Systems. Springer: Cham, Switzerland, 2021; pp. 308–314. [CrossRef]
31. Bainbridge, L. Ironies of Automation. *Automatica* **1983**, *19*, 775–779. [CrossRef]
32. Parasuraman, R.; Riley, V. Humans and Automation: Use, Misuse, Disuse, Abuse. *Hum. Factors* **1997**, *39*, 230–253. [CrossRef]
33. Wolmar, C. *Driverless Cars: On a Road to Nowhere?* London Publishing Partnership: London, UK, 2020.
34. Lee, J.D.; See, K.A. Trust in Automation: Designing for Appropriate Reliance. *Hum. Factors* **2004**, *46*, 50–80. [CrossRef] [PubMed]
35. Lee, J.; Kim, K.J.; Lee, S.; Shin, D.D. Can Autonomous Vehicles Be Safe and Trustworthy? Effects of Appearance and Autonomy of Unmanned Driving Systems. *Int. J. Hum.-Comput. Interact.* **2015**, *31*, 682–691. [CrossRef]
36. Yokoi, R. Trust in Autonomous Cars Does Not Largely Differ from Trust in Human Drivers when They Make Minor Errors. *Transp. Res. Rec.* **2024**. [CrossRef]
37. Colley, M.; Eder, B.; Rixen, J.I.; Rukzio, E. Effects of Semantic Segmentation Visualization on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. In Proceedings of the CHI '21: 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1–11. [CrossRef]
38. Hoff, K.A.; Bashir, M. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Hum. Factors* **2015**, *57*, 407–434. [CrossRef]
39. Ha, T.; Kim, S.; Seo, D.; Lee, S. Effects of Explanation Types and Perceived Risk on Trust in Autonomous Vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* **2020**, *73*, 271–280. [CrossRef]
40. Lee, S.C.; Sanghavi, H.; Ko, S.; Joen, M. Autonomous driving with an agent: Speech style and embodiment. In Proceedings of the Automotive UI '19: 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings, Utrecht, The Netherlands, 21–25 September 2019; pp. 209–214. [CrossRef]
41. Amirova, A.; Rakhymbayeva, N.; Yadollahi, E.; Sandygulova, A.; Johal, W. 10 Years of Human-NAO Interaction Research: A Scoping Review. *Front. Robot. AI* **2021**, *8*, 744526. [CrossRef]
42. Wang, M.; Lee, S.C.; Kamalesh Sanghavi, H.; Eskew, M.; Zhou, B.; Joen, M. In-vehicle Intelligent Agents in Fully Autonomous Driving: The Effects of Speech Style and Embodiment Together and Separately. In Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Leeds, UK, 9–14 September 2021; pp. 247–254.
43. Wang, M.; Lee, S.C.; Monavon, G.; Qin, J.; Jeon, M. Conversational Voice Agents are Preferred and Lead to Better Driving Performance in Conditionally Automated Vehicles. In Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seoul, Republic of Korea, 17–20 September 2022; pp. 86–95.
44. Onnasch, L.; Hildebrandt, C.L. Impact of Anthropomorphic Robot Design on Trust and Attention in Industrial Human-robot Interaction. In *ACM Transactions on Human-Robot Interaction (THRI)*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1–24.
45. Stanton, N.A.; Salmon, P.M. Human Error Taxonomies Applied to Driving: A Generic Driver Error Taxonomy and its Implications for Intelligent Transport Systems. *Saf. Sci.* **2009**, *47*, 227–237. [CrossRef]
46. Brown, I.D. Drivers' Margins of Safety Considered as a Focus for Research on Error. *Ergonomics* **1990**, *33*, 1307–1314. [CrossRef]

47. Faul, F.; Erdfelder, E.; Lang, A.-G.; Buchner, A. G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behav. Res. Methods* **2007**, *39*, 175–191. [CrossRef]
48. Jian, J.Y.; Bisantz, A.M.; Drury, C.G. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *Int. J. Cogn. Ergon.* **2000**, *4*, 53–71. [CrossRef]
49. Spain, R.D.; Bustamante, E.A.; Bliss, J. Towards an Empirically Developed Scale for System Trust: Take Two. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, New York, NY, USA, 22–26 September 2008. [CrossRef]
50. Gutzwiller, R.S.; Chiou, E.K.; Craig, S.D.; Lewis, C.M.; Lematta, G.J.; Hsiung, C.-P. Positive Bias in the 'Trust in Automated Systems Survey'? An Examination of the Jian et al. (2000) Scale. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; SAGE Publications: Los Angeles, CA, USA, 2019; Volume 63, pp. 217–221.
51. Korber, M. Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018), Florence, Italy, 26–30 August 2018. Advances in Intelligent Systems and Computing. [CrossRef]
52. Funke, F.; Reips, U.-D. Why Semantic Differentials in Web-Based Research Should Be Made from Visual Analogue Scales and Not from 5-Point Scales. *Field Methods* **2012**, *24*, 310–327. [CrossRef]
53. Grant, S.; Aitchison, T.; Henderson, E.; Christie, J.; Zare, S.; McMurray, J.; Dargie, H. A Comparison of the Reproducibility and the Sensitivity to Change of Visual Analogue Scales, Borg scales, and Likert Scales in Normal Subjects During Submaximal Exercise. *Chest* **1999**, *116*, 1208–1217. [CrossRef]
54. Kuhlmann, T.; Dantlgraber, M.; Reips, U.D. Investigating Measurement Equivalence of Visual Analogue Scales and Likert-type Scales in Internet-based Personality Questionnaires. *Behav. Res. Methods* **2017**, *49*, 2173–2181. [CrossRef] [PubMed]
55. Reips, U.D.; Funke, F. Interval-level Measurement with Visual Analogue Scales in Internet-based Research: VAS Generator. *Behav. Res. Methods* **2008**, *40*, 699–704. [CrossRef]
56. Voutilainen, A.; Pitkäaho, T.; Kvist, T.; Vehviläinen-Julkunen, K. How to ask about Patient Satisfaction? The Visual Analogue Scale is Less Vulnerable to Confounding Factors and Ceiling Effect than a Symmetric Likert Scale. *J. Adv. Nurs.* **2015**, *72*, 946–957. [CrossRef]
57. Yusof, N.A.D.M.; Jamil, P.A.S.M.; Hashim, N.M.; Karuppiah, K.; Rasdi, I.; Tamrin, S.B.M.; Sambasivam, S. Likert Scale vs. Visual Analogue Scale on Vehicle Seat Discomfort Questionnaire: A Review. *Malays. J. Med. Health Sci.* **2019**, *15*, 159–164.
58. Holthausen, B.E.; Wintersberger, P.; Walker, B.N.; Riener, A. Situational Trust Scale for Automated Driving (STS-AD): Development and Initial Validation. In Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Virtual, 21–22 September 2020; pp. 40–47.
59. Elliott, E.M.; Bell, R.; Gorin, S.; Robinson, N.; Marsh, J.E. Auditory Distraction can be Studied Online! A Direct Comparison Between In-person and Online Experimentation. *J. Cogn. Psychol.* **2022**, *34*, 307–324. [CrossRef]
60. Woods, K.J.; Siegel, M.H.; Traer, J.; McDermott, J.H. Headphone Screening to Facilitate Web-based Auditory Experiments. *Atten. Percept. Psychophys.* **2017**, *79*, 2064–2072. [CrossRef]
61. Bridgwater, T.; Giuliani, M.; van Maris, A.; Baker, G.; Winfield, A.; Pipe, T. Examining Profiles for Robotic Risk Assessment: Does a Robot's Approach to Risk Affect User Trust? In Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, UK, 23–26 March 2020; pp. 23–31.
62. Stuck, R.E.; Holthausen, B.E.; Walker, B.N. Chapter 8—The role of risk in human-robot trust. In *Trust in Human-Robot Interaction*; Nam, C.S., Lyons, J.B., Eds.; Academic Press: Cambridge, MA, USA, 2021; pp. 179–194. [CrossRef]
63. Hancock, P.A.; Billings, D.R.; Schaefer, K.E.; Chen, J.Y.C.; de Visser, E.J.; Parasuraman, R. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Hum. Factors* **2001**, *53*, 517–527. [CrossRef]
64. Seo, K.H.; Lee, J.H. The Emergence of Service Robots at Restaurants: Integrating Trust, Perceived Risk, and Satisfaction. *Sustainability* **2021**, *13*, 4431. [CrossRef]
65. Robinette, P.; Howard, A.M.; Wagner, A.R. Effect of Robot Performance on Human–Robot Trust in Time-Critical Situations. *IEEE Trans. Hum.-Mach. Syst.* **2017**, *47*, 425–436. [CrossRef]
66. Carpinella, C.M.; Wyman, A.B.; Perez, M.A.; Stroessner, S.J. The Robotic Social Attributes Scale (RoSAS) Development and Validation. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 254–262.
67. Penmetsa, P.; Adanu, E.K.; Wood, D.M.; Wang, T.; Jones, S.L. Perceptions and Expectations of Autonomous Vehicles—A Snapshot of Vulnerable Road User Opinion. *Technol. Forecast. Soc. Chang.* **2019**, *133*, 9–14. [CrossRef]
68. Schoettle, B.; Sivak, M. *A Survey of Public Opinion About Autonomous and Self-Driving Vehicles in the US, the UK, and Australia*; University of Michigan Transportation Research Institute: Ann Arbor, MI, USA, 2014.
69. Li, J. The Benefit of being Physically Present: A Survey of Experimental Works Comparing Copresent Robots, Telepresent Robots and Virtual Agents. *Int. J. Hum.-Comput. Stud.* **2015**, *77*, 23–37. [CrossRef]
70. Bainbridge, W.A.; Hart, J.W.; Kim, E.S.; Scassellati, B. The Benefits of Interactions with Physically Present Robots over Video-displayed Agents. *Int. J. Soc. Robot.* **2015**, *3*, 41–52. [CrossRef]
71. Powers, A.; Kiesler, S. The Advisor Robot: Tracing People's Mental Model from a Robot's Physical Attributes. In Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, Salt Lake City, UT, USA, 2–3 March 2006; pp. 218–225.

72. Cheng, P.; Meng, F.; Yao, J.; Wang, Y. Driving with Agents: Investigating the Influences of Anthropomorphism Level and Physicality of Agents on Drivers' Perceived Control, Trust, and Driving Performance. *Front. Psychol. Hum. Media-Interact.* **2022**, *13*, 883417. [CrossRef]
73. Tanaka, T.; Fujikake, K.; Yonekawa, T.; Yamagish, M.; Inagami, M.; Kinoshita, F.; Aoki, H.; Kanamori, H. Driver Agent for Encouraging Safe Driving Behavior for the Elderly. In Proceedings of the 5th International Conference on Human Agent Interaction, Bielefeld, Germany, 17–20 October 2017; pp. 71–79.
74. Tamura, S.; Ohshima, N.; Hasegawa, K.; Okada, M. Design and Evaluation of Attention Guidance Through Eye Gazing of "NAMIDA" Driving Agent. *J. Robot. Mechatron.* **2021**, *33*, 24–32. [CrossRef]
75. Okada, M. Weak Robots. *JASP Rev.* **2022**, *2022*, 220409. [CrossRef]
76. Lee, S.C.; Jeon, M.A. Systematic Review of Functions and Design Features of In-vehicle Agents. *Int. J. Hum.-Comput. Stud.* **2022**, *165*, 102864. [CrossRef]
77. Yun, Y.; Oh, H.; Myung, R. Statistical Modeling of Cultural Differences in Adopting Autonomous Vehicles. *Appl. Sci.* **2022**, *11*, 9030. [CrossRef]
78. Rhim, J.; Lee, G.-B.; Lee, J.-H. Human Moral Reasoning Types in Autonomous Vehicle Moral Dilemma: A Cross-cultural Comparison of Korea and Canada. *Comput. Hum. Behav.* **2020**, *102*, 39–56. [CrossRef]
79. Law, T. Measuring Relational Trust in Human-robot Interactions. In Proceedings of the Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, UK, 23–26 March 2020; pp. 579–581.
80. Law, T.; Scheutz, M. Chapter 2—Trust: Recent concepts and evaluations in human-robot interaction. In *Trust in Human-Robot Interaction*; Nam, C.S., Lyons, J.B., Eds.; Academic Press: Cambridge, MA, USA, 2021; pp. 27–57. [CrossRef]
81. Pöllänen, E.; Read, G.J.M.; Lane, B.R.; Thompson, J.; Salmon, P.M. Who is to Blame for Crashes Involving Autonomous Vehicles? Exploring Blame Attribution Across the Road Transport System. *Ergonomics* **2020**, *63*, 525–537. [CrossRef] [PubMed]
82. Wan, F.; Teng, J.; Feng, L. Exploring User Attitudes and Behavioral Intentions towards Augmented Reality Automotive Assistants: A Mixed-Methods Approach. *World Electr. Veh. J.* **2024**, *15*, 258. [CrossRef]
83. Li, M.; Wang, Y.-X.; Ramanan, D. Towards Streaming Perception. *arXiv* **2020**, arXiv:2005.10420. [CrossRef]
84. Li, C.; Cheng, Z.-Q.; He, J.-Y.; Li, P.; Luo, B.; Chen, H.; Geng, Y.; Lan, J.-P.; Xie, X. Longshortnet: Exploring temporal and semantic features fusion in streaming perception. *arXiv* **2023**, arXiv:2210.15518. [CrossRef]
85. He, J.-Y.; Cheng, Z.-Q.; Li, C.; Xiang, W.; Chen, B.; Luo, B.; Geng, Y.; Xie, X. DAMO-StreamNet: Optimizing Streaming Perception in Autonomous Driving. In Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI 2023), Macao, China, 19–25 August 2023. [CrossRef]
86. Hauptmann, A.; Lijun, Y.; Wenhe, L.; Yijun, Q.; Zhiqi, C.; Liangke, G. *Robust Automatic Detection of Traffic Activity*; Carnegie Mellon University Mobility 21 Final Report; Mobility21; Carnegie Mellon University: Pittsburgh, PA, USA, 2023; pp. 1–72. Available online: https://rosap.ntl.bts.gov/view/dot/68085 (accessed on 13 November 2024).
87. Moody, J.; Bailey, N.; Zhao, J. Public Perceptions of Autonomous Vehicle Safety: An International Comparison. *Saf. Sci.* **2020**, *121*, 554–561. [CrossRef]
88. Du, N.; Robert, L.; Yang, 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]
89. Etzioni, S.; Hamadneh, J.; Elvarsson, A.B.; Esztergár-Kiss, D.; Djukanovic, M.; Neophytou, S.N.; Sodnik, J.; Polydoropoulou, A.; Tsouros, I.; Pronello, C.; et al. Modeling Cross-National Differences in Automated Vehicle Acceptance. *Sustainability* **2020**, *12*, 9765. [CrossRef]
90. Dudziak, A.; Stoma, M.; Kuranc, A.; Caban, J. Assessment of Social Acceptance for Autonomous Vehicles in Southeastern Poland. *Energies* **2021**, *14*, 5778. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.