Transportation Research Part F: Psychology and Behaviour 119 (2026) 103589

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

journal homepage: www.elsevier.com/locate/trf

Trust and blame in autonomous vehicles: examining the effects of cyber readiness and response in the UK and Japan☆

Victoria Marcinkiewicz [a,b,c], Qiyuan Zhang [a,b,c], Minoru Asada [f,g], Yoshiyuki Ueda [h], Hirofumi Katsuno [i], Tatsuhiko Inatani [j], Phillip L. Morgan [a,b,c,d,e,k,*]

[a] *School of Psychology, Cardiff University, 70 Park Place, Cardiff CF10 3AT, United Kingdom*
[b] *Cardiff University Centre for AI, Robotics, and Human-Machine Systems (IROHMS), United Kingdom*
[c] *Cardiff University Human Factors Excellence Research Group (HuFEx), United Kingdom*
[d] *Visiting Professor at Luleå University of Technology, Psychology, Division of Health, Medicine and Rehabilitation, Sweden*
[e] *Distinguished Visiting Professor at the Faculty of Education, Science, Technology and Mathematics, University of Canberra, Australia*
[f] *The University of Osaka, Yamadaoka 1-1, Suita City, Osaka Pref. 565-0871, Japan*
[g] *International Professional University of Technology in Osaka, Japan*
[h] *Institute for the Future of Human Society, Kyoto University, 46 Yoshidashimoadachi-chocho Sakyo-ku Kyoto, 606-8501, Japan*
[i] *Doshisha University, 601 Gembucho, Higashi-iru, Imadegawa Karasuma-dori, Kamigyo-ku, Kyoto 602-8580, Japan*
[j] *Kyoto University, Graduate School of Law, Yoshidahonmachi Sakyo-ku Kyoto, 606-8501, Japan*
[k] *Cardiff University Digital Transformation Innovation Institute (DTII), United Kingdom*

ARTICLE INFO

Keywords:
Autonomous vehicles
Self-driving cars
Cyber security
Cyber attack
Trust
Blame

ABSTRACT

An adverse cyber attack, incident or event either on an autonomous vehicle (AV) itself and/or its connected infrastructure could lead to various consequences including but not limited to disruption, reputational damage (e.g. at the company and broader levels), legal implications and financial penalties. Cyber attacks are also likely to undermine trust which is a key factor in the uptake and use of new (transport) technologies and is linked to acceptance, adoption and continued use. Preparing for and responding appropriately to a cyber attack targeting AV technologies is likely to be critical in minimising its impact and for maintaining public confidence and trust. This paper presents data from experiments conducted in the UK and Japan exploring the effects of cyber readiness and cyber response on trust in AVs before and after a cyber attack. Using Simulation Software Generated Animations, findings indicate that after a cyber attack, trust in an AV and the company responsible for the AV sharply declines. The level of cyber readiness has the potential to impact trust in an AV before a cyber attack occurs – when more mature practices were demonstrated, trust in an AV and the company responsible was higher. Following a cyber attack, the extent to which trust declined depended on the type of cyber response. Positive responses led to a smaller reduction in trust than negative responses. Defining the term cyber affected its desirability. The pattern of results was largely similar between the UK and Japan.

☆ This article is part of a Special issue entitled: 'Cybersecurity & Transport' published in Transportation Research Part F: Psychology and Behaviour.

* Corresponding author at: School of Psychology, Cardiff University, 70 Park Place, Cardiff CF10 3AT, United Kingdom
E-mail address: morganphil@cardiff.ac.uk (P.L. Morgan).

https://doi.org/10.1016/j.trf.2026.103589
Received 29 August 2025; Received in revised form 16 February 2026; Accepted 9 March 2026
Available online 19 March 2026

1. Introduction & literature review

The ability for road vehicles via Artificial Intelligence (AI) and autonomous systems to drive themselves under most and possibly all conditions – without the need for human-in-control transitions – continues to become a likely reality. Designers and vehicle manufacturers globally are heavily investing in further advancing autonomous vehicle (AV) technologies with countries such as the UK and Japan amongst those at the forefront of knowledge and regulation development. According to the Society of Automotive Engineers (SAE), which defines 6 levels of automation, the highest level a vehicle can achieve is Level 5. SAE Level 5 is defined as a vehicle that can drive itself under all conditions with no human intervention required (SAE International, 2021) – an AV. Currently, the UK and Japan (like e.g. US) permit vehicles with lower levels of automation (e.g. Level 2 – more than one autonomous system) to be used legally on their roads. Testing for higher levels of automation (Level 3+ – vehicle can drive itself some of the time with human intervention required at times) is also advancing with notable test-sites in the US (e.g., San Francisco) as well as in both the UK (e.g., London: Lancefield, 2025) and Japan (e.g., Yokohama: Kageyama, 2025).

Such continuously evolving developments in technology as well as policies, standards and regulations are paving the way for a future that includes AVs into national and global transport system networks. AVs are set to bring many benefits and amongst some of the most discussed advantages are: reducing road traffic accidents; improving traffic flow and reducing congestion; lowering emissions; and potentially allowing humans to undertake other activities besides driving, such as reading and/or working. There are, however, still many concerns and challenges including ethical dilemmas in the event of something going wrong, technological inaccuracies and infrastructure issues, which can differ between environments. Such concerns and challenges have inhibited not only aspects of AV development but also the opportunity to test and deploy them even on a small scale.

In more recent years, cyber security of AVs and/or their connected infrastructure, including the safety and privacy of passengers and their information, has become another key concern. With the number of potential entry points into systems and networks growing, it is feared that AVs, due to their highly connected and autonomous nature, could be susceptible to more frequent and sophisticated cyber attacks than most current vehicles. A cyber attack on one or more AV and/or its connected infrastructure is inevitable and could have serious (and potentially catastrophic) implications (e.g. disrupting the safety of a road network/area, impacting critical national infrastructure). Attack types could be physical and/or remote, such as malicious onboard diagnostic port (OBD) device attacks and GPS spoofing attacks (Pham & Xiong, 2021). Technical solutions have been proposed including advanced and hybrid intrusion detection systems, as well as regulations, guidelines and standards such as UN R155 (UNECE[a], 2021) and UN R156 (UNECE[b], 2021), ISO/SAE 21434:2021 (ISO[a], 2021) and ISO 24089:2023 (ISO[b], 2023) that are becoming mandated to ensure best practice cyber security across an industry, which currently lacks established protocols. Nevertheless, cyber security is widely recognised as a socio-technical challenge (Proctor & Chen, 2017; Linkov et al., 2019), and human interaction with AV systems remains a potential point of vulnerability.

No matter the type and sophistication of procedural and technical solutions to defend against various cyber attacks, threat actors will also strive to compromise an AV system(s) and/or the infrastructure through the user. Humans will still interact with the system e.g. by entering navigation information, updating settings and preferences, installing updates, and more. It is likely that threat actors will try and exploit these and other human interactions with the AV by e.g. preying on human cyber risk vulnerabilities (including cognitive biases) to gain entry to the system(s). To date, there has been a dearth of research on the psychological and human factors aspects associated with cyber attacks on AVs. One such factor is trust. Trust in automation more broadly hinges on antecedents such as perceived safety and security as well as perceived reliability and predictability (Hancock et al., 2011; Lee & Moray, 1992; Lee & See, 2004; Schaefer et al., 2016) and therefore understanding the impact an act of cyber can have on trust and/or its antecedents is paramount.

Trust takes time to establish, usually through experience with a system (Olaverri-Monreal, 2020); it builds slowly under normal operation but erodes rapidly after failures (Körber et al., 2016; Lee & See, 2004). This suggests that designing for trustworthy everyday behaviour is necessary but not necessarily sufficient when an AV is under duress. When failures, errors, malfunctions or unavoidable crashes occur, users often look beyond system competence and usability to questions of fairness, accountability, and ethical transparency (Bonnefon et al., 2016). However, unlike malfunctions, errors or failures, acts of cyber can be accidental or intentional, not just mechanical, technical or perceptual. With users set to engage with a system like never before – from inside – the parameters of risk, safety and security have the potential to be largely different in an AV compared to other technologies.

Whereas many risks are well characterised, cyber attacks on AVs pose uncertain risks, including direct safety threats to vehicle occupants, other road users, and pedestrians, as well as indirect legal, financial, and reputational consequences for manufacturers, regulators, and governments. A cyber attack could also influence the perceived safety and security of AVs. These perceptions could reduce acceptance and hinder adoption and continued use across individuals, groups, societies, countries and cultures. Early efforts to understand public attitudes toward AV-related risks have relied predominantly on survey methodologies (Choi & Ji, 2015; Kyriakidis et al., 2015; Schoettle & Sivak, 2014). Although these studies revealed some concerns about hacking, misuse, and data privacy, this alongside cyber security was not their primary focus, and the implications of these concerns for trust and technology acceptance were not examined in depth.

More recently and directly focused toward understanding cyber, Khan et al. (2023) conducted an international survey (Australia, New Zealand, US and the UK) with 2062 respondents. The survey examined six perceived cyber barriers to AV deployment namely, data privacy; AV connectivity; ITS infrastructure; AV cyber security regulation; AV cyber security comprehension, and AV cyber insurance. The findings showed that as education levels increase, the significance of a cyber barrier to AV deployment decreases. However, as AV comprehension and cyber security knowledge increase, the perception of a cyber barrier becomes significantly more important. In addition, the study demonstrated differences in perceptions of cyber barriers and AV deployments based on gender, age, income, and geographic location.

In addition, growing empirical literature has begun to examine the consequences of cyber attacks for trust in AVs, behavioural responses, and intentions to use the technology. Simulator-based experiments consistently show that exposure to (simulated) cyber incidents (e.g., ransomware messages or system malfunctions) significantly reduces trust in automation and increases manual takeover behaviour (Lim et al., 2024; Payre et al., 2022; Payre et al., 2023). Importantly, trust sharply declined after a cyber attack and remained degraded with only partial recovery, even when the system returned to error-free operation (Lim et al., 2024). This suggests that cyber incidents could have enduring psychological effects. Individual differences and contextual factors further moderate behavioural responses to cyber threats: sensation seeking, prior cyber security training, and scenario characteristics, influence take-over and safety behaviours (Chen et al., 2021; Wang et al., 2024; Zhang et al., 2023). Despite this, there has been relatively little research examining the psychological and human factors dimensions of cyber attacks on AVs.

Closely related to trust is the concept of blame assignment. Blame is an important indicator of attitudes toward AVs (e.g., Awad et al., 2020; Bennett et al., 2020; Hong, 2020; Hong et al., 2021; Liu & Du, 2021; Pöllänen et al., 2020; Wallbridge et al., 2024; Zhang et al., 2024) and often serves as a proxy for accountability. When people perceive that a responsible agent (e.g. human, organisational or other - e.g. non-human agent) is to blame, this can affect their willingness to trust that agent in the future (Kim et al., 2004; Mayer et al., 1995). Research shows that how blame is attributed - fairly or unfairly - also plays a role in trust repair. For example, Tomlinson and Mayer (2009) argue that trust repair hinges on attributional judgements. If individuals perceive blame as unfairly assigned, they experience greater mistrust. However, trust recovery can occur (to a degree at least) when an organisation or agent acknowledges appropriate responsibility, provides transparent explanations, and implements corrective measures (Tomlinson & Mayer, 2009). Furthermore, blame attribution is often complicated by the ambiguity of responsibility. Research by Malle et al. (2014) and de Visser et al. (2018) revealed that people are more likely to blame human agents (who are perceived to have intent and accountability) than automated systems, even if both share responsibility. Blame is not just about causation but also neglect or lack of diligence. In high-risk safety-critical contexts, people are more sensitive to failures - and therefore more likely to assign blame when something goes wrong (Malle et al., 2014; Shaver, 1985). If users perceive that an organisation failed to prepare or ignored (known) risks, they are more likely to assign blame and reduce their trust in that organisation (e.g. Coombs, 2007; Mayer et al., 1995). Although these studies are not specific to cyber security or AVs, they provide a basis for understanding how organisational unpreparedness - particularly for known risks - can erode stakeholder trust and invite blame - consistent with wider human factors literature.

When users perceive a system as well-prepared, visible safeguards and transparent communication provide reassurance and enhance perceived reliability (Hoff & Bashir, 2015; Lee & See, 2004). *Readiness (or preparedness)*, by extension, communicates competence and reliability, which remain core antecedents of trust in automation (Mayer et al., 1995). Within the AV cyber security domain, technical studies emphasise the importance of secure communication, authentication, and intrusion detection in ensuring subsystem resilience, with each component (e.g., hardware, communication protocols, certificate authorities, PKI infrastructure) needing to be "trusted" in an engineering (technical) sense (Sun et al., 2022). Comparatively, little is known about how the communication of such (cyber) readiness - potentially through reviews or formalised ratings - shapes *user* trust, acceptance and adoption intentions. Existing evidence suggests that individuals are more willing to adopt AVs when they are exposed to positive information (Anania et al., 2018), implying that signalling an AV's level of cyber readiness could meaningfully increase both trust and likelihood of adoption.

In addition to cyber readiness, appropriately *responding* to cyber threats or incidents has also been recognised as a factor that may significantly affect acceptance and trust in AVs (Gorine & Khan, 2024; Marcinkiewicz et al., 2023). Broader cyber security research exploring data breaches underpins this notion. Data breaches can lead to a significant loss of trust amongst customers, and a delayed disclosure is understood to lead to a greater immediate decline in trust as well as diminishing opportunities for subsequent trust repair (Muzatko & Bansal, 2018). However, trust is not always irreparably damaged as often there is a willingness to continue using the company, providing they demonstrate transparency and proactive communication in response to the breach (Strzelecki & Rizun, 2022). Experimental evidence from Bentley and Ma (2020) further underscores the importance of response quality. They tested consumer reactions to various apology components across data breach scenarios and concluded that trust recovery does not just depend on acknowledging fault, but also on expressions of emotional sincerity and future commitments such as promises of forbearance and reparative action. That is, certain apology features are more impactful than others in high-blame environments. These findings align closely with the Mayer et al. (1995) model of organisational trust, which identifies ability (including perceived competence in incident response), benevolence, and integrity as foundational dimensions of trustworthiness. Therefore, cyber responses including timely disclosure, transparent communication, and sincere reparative efforts each have the potential to influence trust in AVs especially in the aftermath of factors that cause violations.

Despite cyber security readiness and response having been identified as potentially important components for the increase in and potentially maintenance of trust in AV ecosystems (Marcinkiewicz et al., 2023), research on AV cyber security remains predominantly technical, with limited integration of human-centred experimental methods, cross-cultural comparisons, and combined behavioural and attitudinal outcomes. The main aim of this current empirical paper consisting of three human-centred designed experiments, is to address such concerns. A second aim is to provide a cross-country and cultural comparison of trust in AVs in the event of a cyber attack.

2. Experiment 1

Whilst it is impossible for AVs to be completely cyber secure, an effective strategy is to not try to eliminate all cyber attacks against them (there will always be entry points), but to accept that they will occur – even if infrequently - and assess the potential consequences (Lin et al., 2016). It is also crucial to determine interventions that include ensuring users are better prepared in the event of a cyber attack. The main aim of Experiment 1 is to determine whether level of cyber readiness and type of cyber response impacts trust and blame assignment in the event of a cyber attack on an AV, using a human-centric methodology. Based on the evidence reviewed and noting the novelty of this area and paper, there are numerous hypotheses proposed:

- H1 - Trust in the AV will be higher amongst those informed that it has a high cyber readiness rating compared to a low cyber readiness rating.
- H2 - Trust in the AV will be higher for participants receiving a positive cyber response than negative cyber response.
- H3 - Cyber readiness and cyber response will interact, such that the effect of cyber response on trust in the AV will depend on the level of cyber readiness. Specifically, trust will be highest when cyber readiness is high and the response is positive, and lowest when cyber readiness is low and the response is negative.
- H4 - Trust in the company responsible for the AV will be higher for those informed that the AV has a high cyber readiness rating than low cyber readiness rating.
- H5 - Trust in the company responsible for the AV will be higher for participants receiving a positive cyber response than negative cyber response.
- H6 - Cyber readiness and cyber response will interact, such that the effect of cyber response on trust in the company responsible for the AV will depend on the level of cyber readiness. Specifically, trust will be highest when cyber readiness is high and the response is positive, and lowest when cyber readiness is low and the response is negative.
- H7 – The AV company will have less blame attributed toward them for a cyber attack on their AV when they have demonstrated more mature cyber practices - i.e. a higher level of cyber readiness and a positive, responsible and proactive cyber response.

2.1. Method

2.1.1. Participants

An a priori power analysis was conducted in G*Power 3.1 (Cohen, 1988; Faul et al., 2009) for a 3 × 2 between-subjects factorial ANOVA, powered to detect the cyber readiness × cyber response interaction ($df = 2$). Assuming a medium effect size (Cohen's $f = 0.25$), $\alpha = 0.05$, and power $(1 - \beta) = 0.80$, the required sample size was $N = 163$. This power analysis was intended for the primary between-subjects interaction test. To ensure an adequate sample after exclusions and to approach balanced cell sizes, 178 participants were recruited online from the UK via *Prolific©* and randomly assigned to conditions until saturation. Fifteen datasets were not usable either due to being incomplete across multiple measures (e.g. participants did not answer the question(s)) and/or because of incorrect answers to attention check questions.

One-hundred participants identified as male; sixty female; one non-binary; one as other; and one selected ‘prefer not to say’. Ages ranged from 18 to 76 years (*M* 39.0, *SD* 13.29). One-hundred-and-twenty-seven held a full UK driving licence; twenty-five a provisional UK driving licence; two were in the process of obtaining a provisional UK driving licence and nine did not hold a UK driving licence or were in the process of obtaining one. On average, qualified drivers had their licence for 18 years and drove ~6262 miles (10,077 km) per year. Participants were required to have normal/normal-corrected vision; be fluent in English as a first or second language and be ≥18-years. The experiment took 20–30 min to complete, and participants were renumerated GBP £3.75. The experiment had to be completed on a desktop or laptop computer.

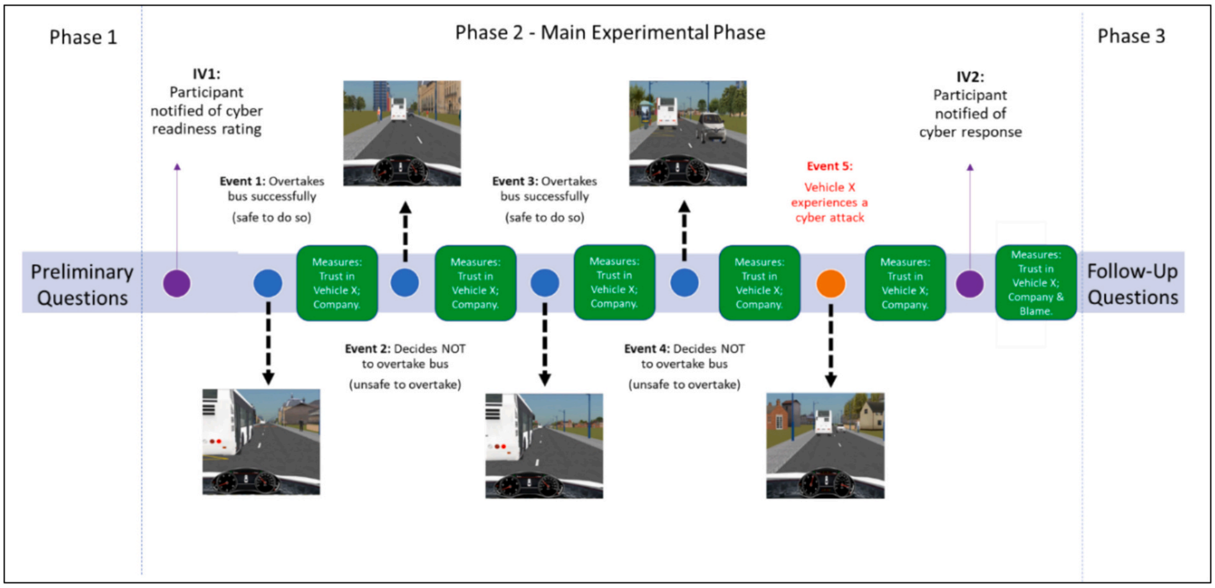
Phase 1
Phase 2 - Main Experimental Phase
Phase 3
IV1: Participant notified of cyber readiness rating
Event 1: Overtakes bus successfully (safe to do so)
Event 3: Overtakes bus successfully (safe to do so)
Event 5: Vehicle X experiences a cyber attack
IV2: Participant notified of cyber response
Preliminary Questions
Measures: Trust in Vehicle X; Company.
Measures: Trust in Vehicle X; Company.
Measures: Trust in Vehicle X; Company.
Measures: Trust in Vehicle X; Company.
Measures: Trust in Vehicle X; Company.
Measures: Trust in Vehicle X; Company & Blame.
Follow-Up Questions
Event 2: Decides NOT to overtake bus (unsafe to overtake)
Event 4: Decides NOT to overtake bus (unsafe to overtake)

Fig. 1. Experimental phases.

2.1.2. Materials

Simulation Software Generated Animations (SSGAs) depicting a hypothetical driving scenario were developed using *SCANeR© Studio* (*AV Simulation©*). The scenario was created and recorded in the driving simulator in what would be regarded as full Level 5 autonomous driving mode and embedded into an online experiment hosted on *Qualtrics©*. Noting this methodology has been adopted in related research (Wallbridge et al., 2024; Zhang et al., 2022). The driving scenario depicted an AV labelled Vehicle X executing a journey before experiencing an unspecified cyber attack (the critical event). During the journey, Vehicle X had to navigate five events - E1 to E5 - and culminated in the critical event (E5) (Fig. 1). Each event involved Vehicle X negotiating a bus resulting in an overtake (E1, E3, E5 - deemed safe to do so due to conditions - e.g. traffic flow, no pedestrians in sight) or non-overtake action (E2, E4 - deemed unsafe to overtake due to conditions).

At all times, participants had a full uninterrupted view of the road ahead visible through the windscreen, and road behind visible via simulated side mirrors. This allowed participants to witness Vehicle X responding to the road conditions. An animated dashboard that displaying e.g. a speedometer, rev counter, and other features (using icons) such as a fuel gauge and battery life (Fig. 2). The dashboard reacted to events accordingly e.g. the speedometer/rev counter reduced when Vehicle X slowed or sped up. During E5, the dashboard malfunctioned (noting this was the subject of an unspecified cyber attack) depicting that something was wrong with the AV (Fig. 2). There was also an auditory warning (a beeping sound) which sounded multiple times alerting the participant to the unusual activity. Immediately after E5, participants were asked a knowledge check question (“What happened at the end of the video?”) to assess their interpretation. Participants were given four choices to select from plus an option to select ‘prefer not to say’ (Appendix A). After responding, participants were informed that the event was the result of a cyber attack. This framing was provided prior to completing any trust or blame measures. Participants who did not correctly identify the event were excluded from the analysis. Throughout the journey, Vehicle X acted in accordance with UK road laws (Highway Code). Vehicle X would overtake with its indicator light on, maintained a safe distance behind the bus, obeyed the speed limit (30mph/48.28kph) and slowed down accordingly when the bus approached a bus stop.

Cyber readiness and response varied between participants. Cyber readiness was presented to participants as part of a ten-feature star-rating review of Vehicle X. The review included nine popular other features of cars today (according to e.g. the UK Parkers Guide, date): costs, comfort, design, environmental friendliness, practicality, performance, reliability, road worthiness and safetyA tenth feature - cyber (security) readiness - was added for the purposes of the present experiment to represent the AVs capability to prevent, detect, and respond to a potential act of cyber. The cyber rating was manipulated as 0.8 (low), 3.0 (medium), and 4.8 (high) out of 5 stars while the ratings for the other features remained constant. These values reflected the approximate lower bound, mid-point, and upper bound of ratings typically reported by Parkers© (2025) across a variety of features. This distributional pattern indicates that the three anchors fall within realistic and commonly encountered values. The anchors were intended primarily as contextual reference points rather than as a strict psychometric calibration. No pretest was conducted. The empirically observed rating ranges are however, based on a nationally recognised automotive review platform which increases the likelihood that participants would perceive the three levels as meaningful and interpretable differences in readiness.

Cyber response was manipulated to either consist of four positive statements or four negative statements containing details on how the company responded (Appendix B). These were provided after participants were informed that E5 was a cyber attack. Participants were also informed in all cases that minimal legal requirements were met in terms of both cyber readiness and cyber response i.e. Vehicle X was legally allowed to operate s even if it had a low star-rating. Star-ratings were operationalised to indicate how well the company/vehicle was performing *above and beyond* minimum requirements with e.g. a 5-star-rating indicating that the company could not be rated any higher and a 0-rating indicating that the company could not be rated any lower - but still would be meeting minimum (legal) requirements.

Trust was measured via self-reported measures. During Phase 1, there was a question on existing levels of trust and intentions to use *any* AV (not Vehicle X): ‘Imagine that self-driving cars will be deployed on a large scale on UK roads. Using a Visual Analogue Scale (VAS), participants were asked to rate how much they currently trust and would be likely to use AV technology’ – with VAS anchors of ‘Do not Trust at all’ to ‘Completely Trust’ / ‘Would not use’ to ‘Would use all the time’. VAS offer greater granularity (Wu & Leung, 2017) with the ‘number’ selected not visible to the participant. This prevents certain uncontrollable bias such as participants only

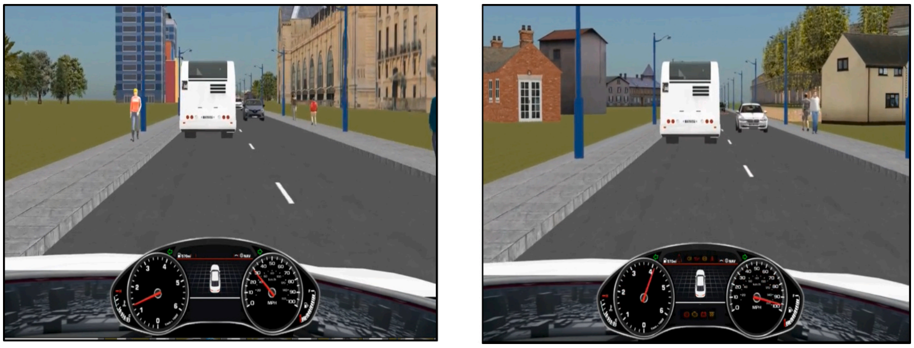

Fig. 2. Dashboard of Vehicle X before (left)/during (right) the occurrence of a cyber attack.

picking even/odd numbers, rounding to the nearest e.g. whole number or always choosing the same number (Sung & Wu, 2018). During Phase 2, initial trust in Vehicle X and the company responsible for Vehicle X was measured also using VAS questions after participants had read the review of the AV. At this stage participants were also asked ten questions in relation to general expectations of AVs (Appendix C).

Trust was also measured after each event E1 – E5 using the Situational Trust Scale for Automated Driving (STS-AD) (Holthausen et al., 2020); adapted to collect data using a VAS as opposed to a 7-point Likert scale. Although trust was measured after each event, event was not treated as a focal predictor in the primary hypotheses. Accordingly, the primary analyses examined the between-subjects effects of cyber readiness and cyber response. Effect of event were examined, where appropriate, in supplementary repeated-measures analyses.

Blame assignment relating to the cyber attack was also measured, at the end of the main experimental Phase 2 (after E5 – critical event). A series of statements (e.g. 'the company was most to blame for the cyber attack' - see Appendix D) were provided with a VAS and left and right end anchors - 'Fully Agree' to 'Fully Disagree'. By adopting this approach, blame and trust data could be more easily compared. Prefer not to say options were available. Finally, trust was once again measured at the end the experiment during Phase 3, also using VAS to ascertain post-trust levels in Vehicle X, the company responsible for Vehicle X as well as trust in other AVs in general (i.e. not specifically Vehicle X).

2.1.3. Design

In Experiment 1 (as in all three experiments), a 3 (cyber readiness: high, medium, low) × 2 (cyber response: positive, negative) between-subjects factorial design was employed. Trust was also measured repeatedly across five events, with event treated as a within-subjects factor. However, event was not a manipulated factor of primary theoretical interest and was analysed separately where relevant. All participants were exposed to the same five events depicting Vehicle X navigating an environment on a continuous journey. There were two independent variables (IVs) operationalised to determine what information a participant received before and after the journey – and thus six treatment combinations (Table 1). IV1 related to the information provided about cyber readiness rating (beginning of Phase 2) - low, medium or high (operationalised by star-ratings: 0.8; 3.0 or 4.8 / 5 stars). IV2 related to the information provided about the cyber response (end of Phase 2) post E5 - four positive or four negative statements detailing the company's response.

2.1.4. Procedure

At the outset of Phase 1 (Fig. 1), participants read an information sheet and provided consent to take part. They were not informed about the cyber element of the experiment to minimise expectation effects. They then completed a preliminary demographics questionnaire. Then followed questions on trust in and likelihood of using AVs with VAS responses, followed by ten (Experiment 1) / four (Experiment 2a and 2b) questions on general expectations of AVs. During Phase 2 (Fig. 1), participants were asked to familiarise themselves with the star-rating criteria and had to first rate (out of five stars) each feature of an AV for how desirable it was to them. Then there was a requirement to read, feature by feature, the star-ratings that had been pre-determined for Vehicle X, followed by two questions on the trustworthiness and likelihood of using Vehicle X based on the review ratings provided.

Next, participants watched the scenario with the five events involving Vehicle X. Between events, participants rated the extent to which they agreed with seven short statements based on what they had just witnessed. Six questions were about Vehicle X from the STS-AD and the seventh was trust in the company. During Phase 3 and after E5 (the critical event – a cyber attack), participants were informed of the company's response and additional questions on blame assignment were also asked. There was also a free-text comment option before debriefing.

2.2. Results

Data was screened and checked for skewness and kurtosis with outliers detected using the interquartile range (IQR) method: IQR*1.5 deducted from the lower limit and added to the upper limit. Data points that fell outside of the limits were reduced to the upper or increased to the lower limit. The findings are based on 163 usable datasets that were included within all analyses, based on sample sizes of 28 (High-Positive, Low-Negative), 27 (High-Negative, Medium-Positive, Medium Negative), and 26 (Low-Positive).

Phase 1 - General Attitudes Toward AVs.

First, initial trust in and intentions to use AVs was considered. Participants used the full range of the 0-100 scale for both measures. Overall, initial trust in AVs was relatively low ($M = 36.64$, $SD = 28.10$), whereas intentions to use AVs were slightly higher ($M = 49.77$, $SD = 32.85$). The wide ranges and relatively large standard deviations indicate substantial variability in participants' baseline attitudes toward AV technology.

Table 1
Experiment 1 treatment combinations.

		IV1		
		High Readiness	Medium Readiness	Low Readiness
IV2	Positive Response	High-Positive	Medium-Positive	Low-Positive
	Negative Response	High-Negative	Medium-Negative	Low-Negative

Next, on a 101-point VAS where '0' was 'Fully Disagree' and '100' was 'Fully Agree', participants were asked to what extent they agreed with 10 statements (Appendix C). Participants expressed generally positive attitudes toward rating information and product reviews when considering using AVs. Descriptive statistics for each statement can be viewed in Table 2. Because responses were recorded on a 0–100 visual analogue scale, both the mean and median are reported to provide a more complete description of central tendency. As is common with VAS data, responses were distributed across the scale, with some clustering toward the upper end for items reflecting favourable attitudes.

Phase 1 – Ranking of Desirable Features.

Desirable features of an AV were rated by participants using star-ratings between 0 (not that desired) and 5 (highly desired). Safety and reliability ranked as the most desirable features whilst cyber and design least desirable (Fig. 3).

Phase 2 – Trust.

Prior to experiencing E1-E5, initial trust (based only on the information received) was measured. A one-way between-subjects analysis of variance (ANOVA) was conducted to examine the effect of cyber readiness (high, medium, low) on initial trust in Vehicle X. There was no significant main effect of cyber readiness on initial trust in Vehicle X, $F(2, 160) = 1.29$, $p = .28$. The same ANOVA was also conducted to compare the effect of cyber readiness on initial trust in the company responsible for Vehicle X. There was a significant main effect of cyber readiness, $F(2, 160) = 5.94$, $p = .003$. Post-hoc tests using a Tukey HSD method revealed that trust was significantly higher in the high compared to the low cyber readiness condition ($p = .043$, 95% C.I. = [0.23, 17.38]) and in the medium compared to the low condition ($p = .003$, 95% C.I. = [−3.53, 20.77]). There was no significant difference between the medium and high conditions ($p = .62$).

Trust Across Events (E1 – E4).

After participants experienced each event, trust in Vehicle X and in the company was measured using the overall trust score from the STS-AD scale. A 3 (cyber readiness: high, medium, low) × 4 (event: E1–E4) mixed-design ANOVA was conducted to examine trust in Vehicle X with cyber readiness as a between-subjects factor and repeated measurements across events as a within-subjects factor. There was no significant main effect of cyber readiness on trust in Vehicle X, $F(2, 160) = 0.22$, $p = .80$. The same ANOVA was conducted for trust in the company responsible for Vehicle X, which also showed no significant main effect of cyber readiness, $F(2, 160) = 1.47$, $p = .24$.

Trust After E5.

Trust following E5 were examined separately using one-way between-subjects ANOVAs to assess the effect of cyber readiness (high, medium, low). After E5, when a cyber incident occurred, trust declined markedly for both Vehicle X (Fig. 4) and the company responsible for it (Fig. 5). At this stage, participants were aware of the cyber attack but had not been told how the company responded - i.e. participants had only received information about the AVs cyber readiness. There was no significant main effect in trust in Vehicle X depending on the level of cyber readiness, $F(2, 160) = 0.25$, $p = .77$ or on trust in the company responsible for Vehicle X, $F(2, 160) = 0.78$, $p = .462$.

Finally, a repeated-measures ANOVA across all five events (E1–E5) revealed significant differences in trust in Vehicle X and the company responsible for Vehicle X, all $ps < 0.001$. Trust in Vehicle X and the company was significantly lower during the overtake events (E1 and E3) compared to the non-overtake (E2 and E4) events and reached its lowest after E5 (the cyber incident) compared to all other events, all $ps < 0.001$.

A 3 (cyber readiness: high, medium, low) × 2 (cyber response: positive, negative) two-way between-subjects ANOVA was conducted on post-trust in Vehicle X after all events once participants were informed about the company's response (Table 3). There was no significant interaction between cyber readiness and cyber response, $F(2, 157) = 0.50$, $p = .61$ and no significant main effect for cyber readiness $F(2, 157) = 0.73$, $p = .48$. There was however a significant main effect for cyber response, $F(1, 157) = 20.454$, $p < .001$, $\eta_p^2 = 0.12$, indicating post-trust in Vehicle X was higher when the company's response was positive (Fig. 6).

Post-event trust in the company (Table 4) was analysed using the same ANOVA design as for Vehicle X. There was no significant interaction between cyber readiness and cyber response, $F(2, 157) = 1.04$, $p = .36$, and no significant main effect for cyber readiness $F(2, 157) = 0.79$, $p = .45$. There was however a significant main effect for cyber response, $F(1, 157) = 62.53$, $p < .001$, $\eta_p^2 = 0.29$, revealing that trust in the company responsible for Vehicle X were significantly higher when the company's response was positive (Fig. 7).

Next, a paired samples *t*-test revealed a significant difference in initial ($M = 79.74$, $SD = 15.72$) and post-trust ratings ($M = 34.20$,

Table 2
Participants agreeableness with 10 prescribed statements about AVs.

Statement	Mean	Median	*SD*	Min	Max
I am more likely to use a self-driving car with higher star-ratings.	87.59	92	14.93	24	100
Before using a self-driving car, I would find out as much information as I could about its specification*.	89.93	99	16.69	0	100
Whether I would use a self-driving car would be influenced by what other people are saying about them.	60.83	68	27.24	0	100
A self-driving car must achieve a 5-star-rating across all its specifications.	72.07	79	28.57	0	100
Whether I would use a self-driving car would be influenced by reviews that I read.	71.09	76	24.90	0	100
Some specifications of a self-driving car are more important to me than others.	78.33	81	20.79	0	100
Whether I would use a self-driving car would be influenced by my own experience with them.	75.71	79	22.43	0	100
I believe that products and services with higher ratings are better quality than those with lower ratings*.	80.90	85	18.41	12	100
I believe that products and services with higher ratings can be trusted more than those with lower ratings*.	79.30	84	18.19	17	100
I would be willing to pay more for a product or service with higher ratings*.	77.88	81	18.90	0	100

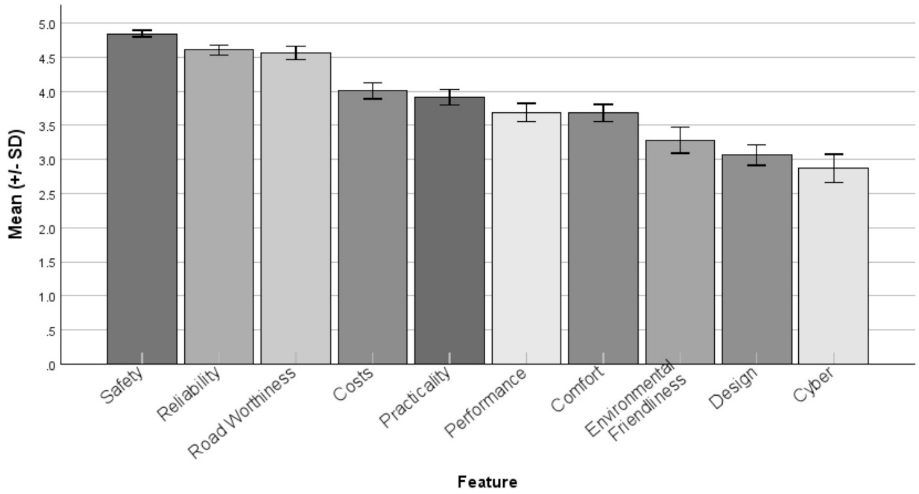

Fig. 3. Rankings of AV desirable features (error bars +/-SD).

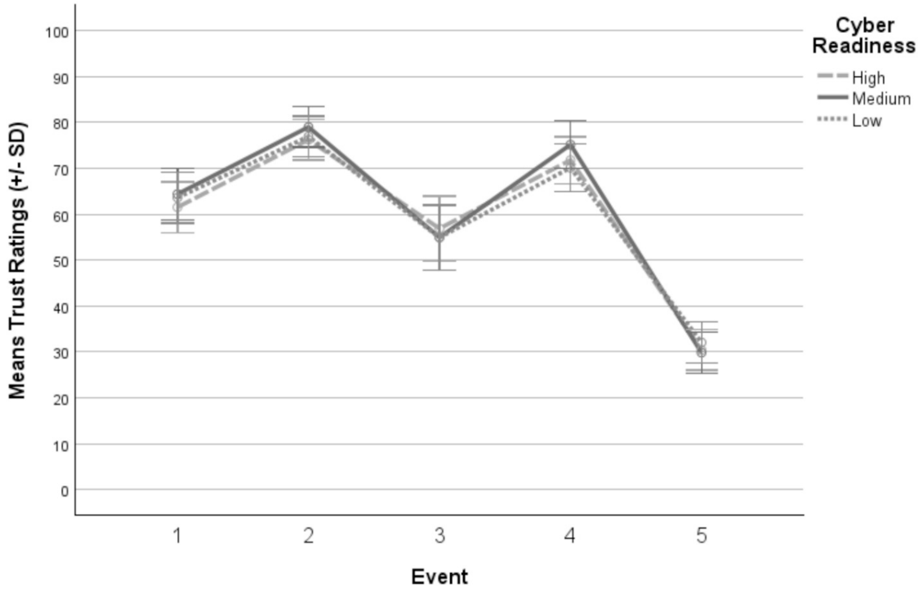

Fig. 4. Mean trust rating in Vehicle X after E1 - E5 (error bars +/-SD).

SD = 28.11) in Vehicle X, $t(162) = 22.38$, $p < .001$, $d = 1.99$. Trust in Vehicle X was significantly lower at the end of the experiment than at the outset. Another paired samples *t*-test compared initial and post-trust ratings in the company responsible for Vehicle X. There was a significant reduction in trust in the company responsible for Vehicle X (initial: $M = 69.64$, $SD = 19.50$; oost: $M = 33.23$ $SD = 27.97$); $t(162) = 17.29$, $p < .001$, $d = 1.51$.

Post-trust in *any* AV (Table 5) was analysed using a 3 (cyber readiness: high, medium, low) × 2 (cyber response: positive, negative) two-way between-subjects ANOVA.

There was no significant interaction between cyber readiness and cyber response, $F(2, 157) = 0.73, p = .48$, and no significant main effect for cyber readiness $F(2, 157) = 0.71, p = .49$. There was a significant main effect for cyber response, $F(1, 157) = 7.49, p < .007$, $\eta_p^2 = 0.05$, indicating that post-trust in *any* AV (not specifically Vehicle X) was significantly lower for participants who received a negative compared to a positive response (Fig. 8).

The same ANOVA was used to evaluate post-use of *any* AV (not specifically Vehicle X) (see Table 6 for descriptives). There was no significant interaction between cyber readiness and response, $F(2, 157) = 2.53$, $p = .08$ and a no significant main effect for cyber readiness $F(2, 157) = 1.01$, $p = .37$. There was however a significant main effect for cyber response, $F(1, 157) = 5.86$, $p = .02$, $\eta_p^2 = 0.04$. Participants who received a negative cyber response were less likely to use *any* other AV compared to those who received a positive response (Fig. 9).

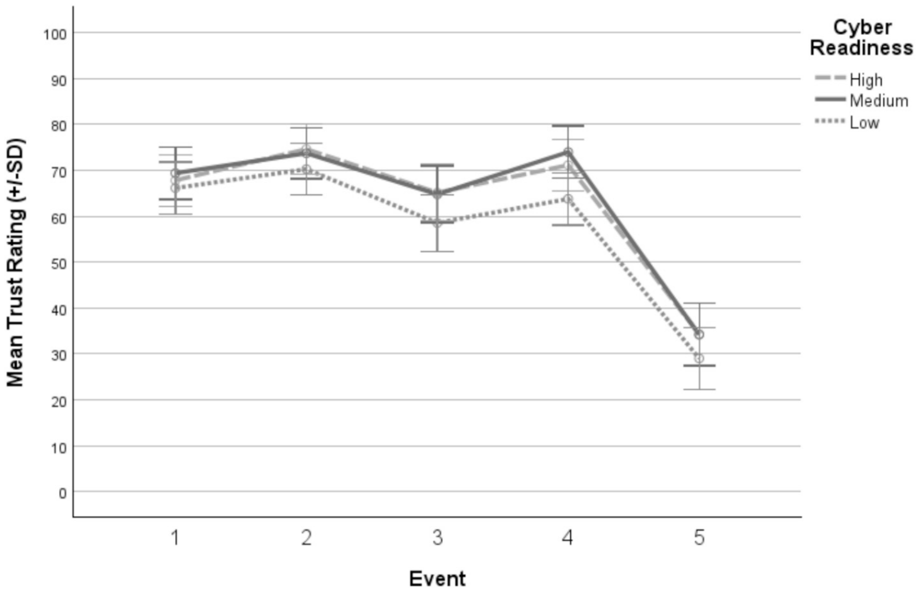

Fig. 5. Mean trust rating in the company responsible for Vehicle X after E1 - E5 (error bars +/-SD).

Table 3
Descriptive findings: post-trust in Vehicle X.

		IV1		
		High readiness	Medium READINESS	Low readiness
IV2	Positive response	*M* 49.36 *SD* 31.73	*M* 38.52 *SD* 30.33	*M* 42.54 *SD* 26.27
	Negative response	*M* 24.56 *SD* 23.64	*M* 23.07 *SD* 21.42	*M* 26.00 *SD* 25.13

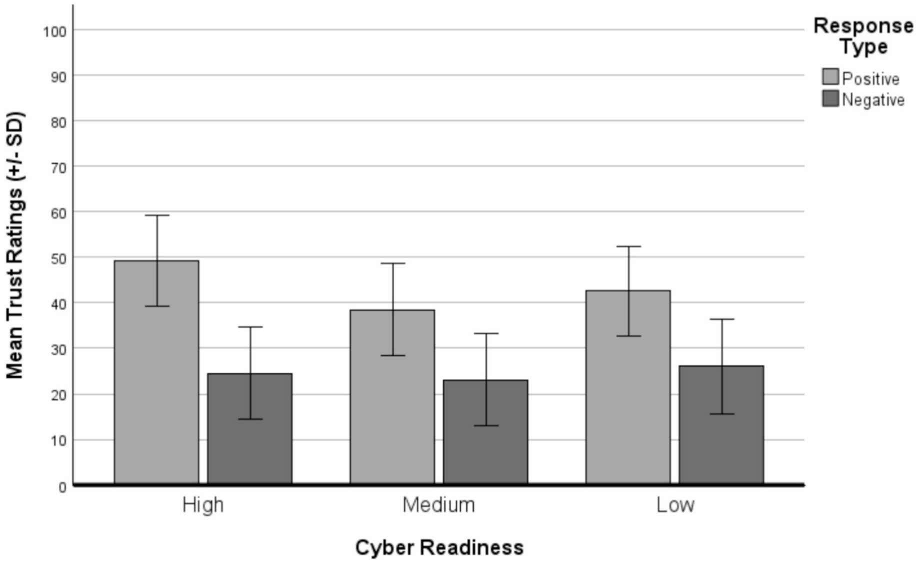

Fig. 6. Mean post-trust in Vehicle X (error bars +/-SD).

Table 4
Descriptive findings: post-trust in the company responsible for Vehicle X.

		IV1		
		High readiness	Medium readiness	Low readiness
IV2	Positive Response	*M* 54.00 *SD* 27.71	*M* 45.30 *SD* 26.63	*M* 43.79 *SD* 27.28
	Negative Response	*M* 17.22 *SD* 19.07	*M* 21.30 *SD* 21.96	*M* 16.00 *SD* 17.73

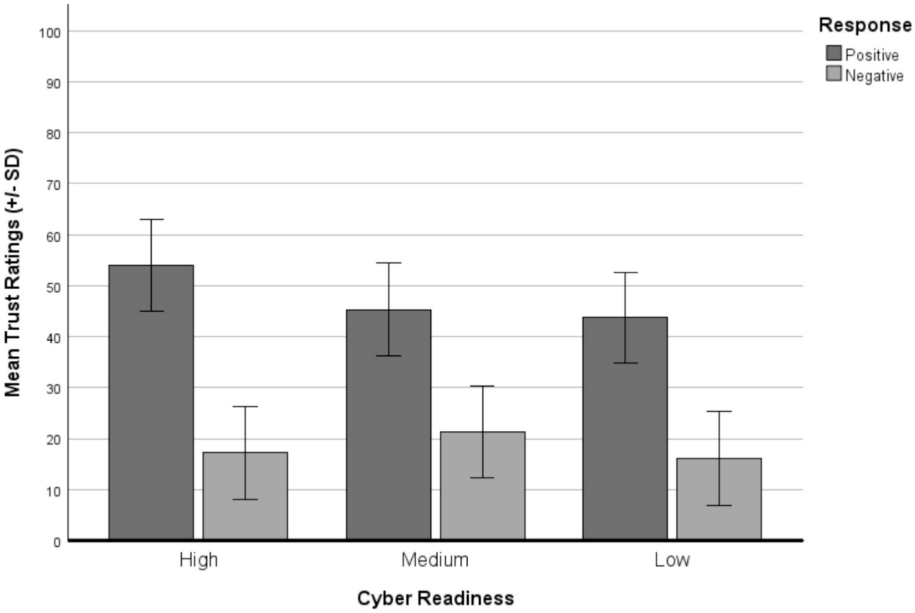

Fig. 7. Mean post-trust in the company responsible for Vehicle X (error bars +/-SD).

Table 5
Post trust ratings in any AV (not specifically Vehicle X).

		IV1		
		High readiness	Medium readiness	Low readiness
IV2	Positive Response	*M* 50.00 *SD* 29.92	*M* 43.15 *SD* 26.85	*M* 48.14 *SD* 29.83
	Negative Response	*M* 30.37 *SD* 28.19	*M* 33.30 *SD* 27.96	*M* 41.04 *SD* 27.60

Phase 2 - Blame.

Responses to the following four blame attribution statements were analysed using the same statistical approach. A 3 (cyber readiness: high, medium, low) × 2 (cyber response: positive, negative) two-way between-subjects ANOVA was conducted to examine a participants' level of agreement with each statement. Higher scores indicated greater agreement.

For the statement ‘The self-driving car company is most to blame for the occurrence of the cyber attack’ (see Table 7 for descriptives) there was no significant interaction between cyber readiness and cyber response, $F(2, 157) = 0.13$, $p = .88$ and no significant main effect for cyber readiness $F(2, 157) = 2.27$, $p = .11$. There was however a significant main effect for cyber response, $F(1, 157) = 9.66$, $p < .002$, $\eta_p^2 = 0.06$. Participants who received a negative cyber response assigned significantly more like to agree that the AV company was most to blame than those who received a positive response (Fig. 10).

For the statement ‘Based on the star-ratings, the self-driving car company could have been better prepared for the cyber attack on Vehicle X' (descriptives are in Table 8) there was a significant interaction between cyber readiness and response, $F(2, 157) = 4.36$, $p = .01$, $\eta_p^2 = 0.05$. Simple main effects analyses indicated that participants encountering a negative cyber response were significantly more likely to agree that the company could have been better prepared when in the high and medium readiness conditions, $p = .001$ and $p = .004$, respectively. There were no differences between cyber readiness and response in the low readiness conditions, $p = .58$ (Fig. 11). Regardless of the cyber response, participants agreed the company could have been better prepared.

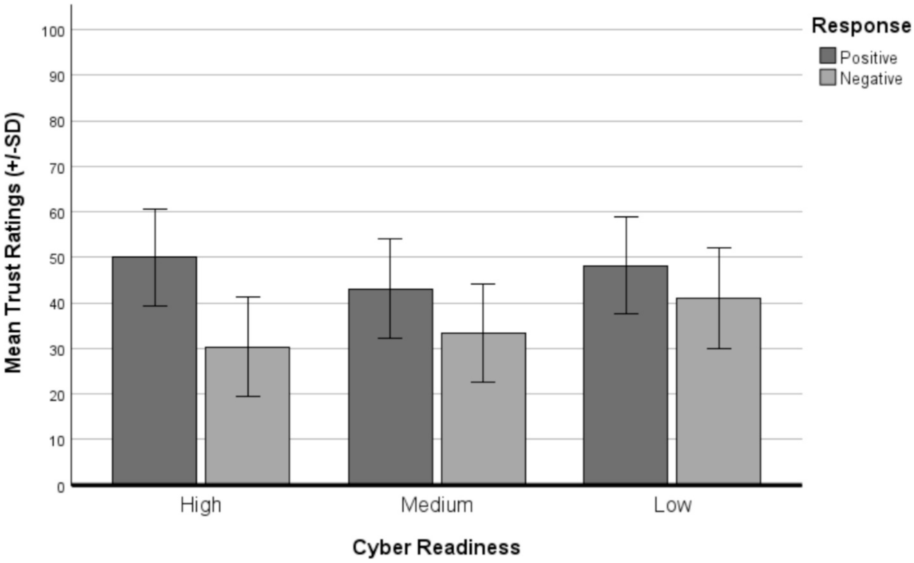

Fig. 8. Mean trust ratings in any AV (not specifically Vehicle X) (error bars +/-SD).

Table 6
Descriptive findings: post-use ratings for any AV (not specifically Vehicle X).

		IV1		
		High readiness	Medium readiness	Low readiness
IV2	Positive Response	*M* 53.68 *SD* 32.16	*M* 40.89 *SD* 31.40	*M* 50.89 *SD* 33.21
	Negative Response	*M* 26.15 *SD* 30.16	*M* 39.15 *SD* 32.756	*M* 44.04 *SD* 30.59

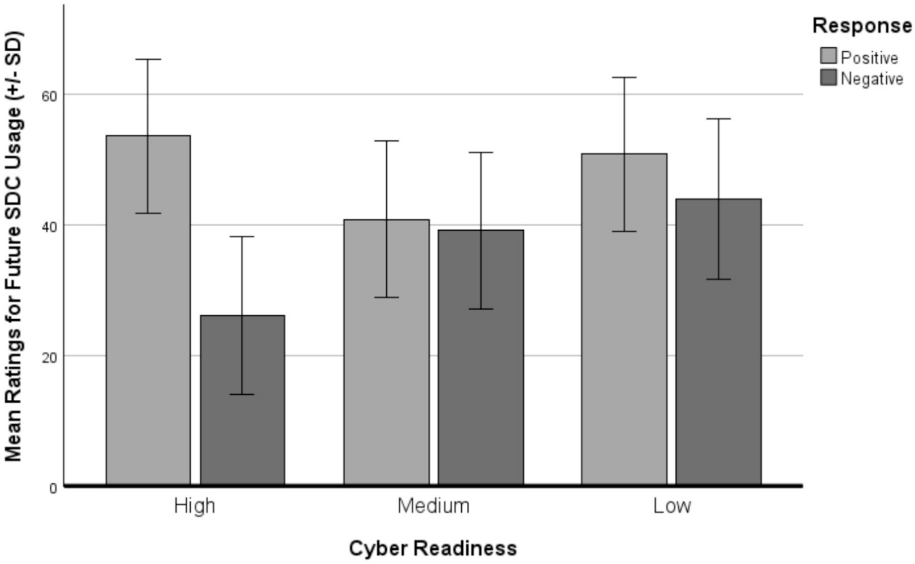

Fig. 9. Mean usage ratings for any AV (not specifically Vehicle X) (error bars +/-SD).

For the statement ‘The self-driving car company responded appropriately to the cyber attack’ (descriptives are in Table 9) there was no significant interaction between cyber readiness and response, $F(2, 157) = 0.08$, $p = .92$, and no significant main effect for cyber readiness $F(2, 157) = 1.17$, $p = .31$. There was however a significant main effect for cyber response, $F(1, 157) = 426.56$, $p < .001$, $\eta_p^2 =$

Table 7
Descriptive findings for the statement ‘The company was most to blame’.

		IV1		
		High readiness	Medium readiness	Low readiness
IV2	Positive Response	*M* 51.96 *SD* 32.82	*M* 52.48 *SD* 33.70	*M* 62.61 *SD* 29.40
	Negative Response	*M* 69.07 *SD* 25.36	*M* 64.04 *SD* 27.45	*M* 76.50 *SD* 24.44

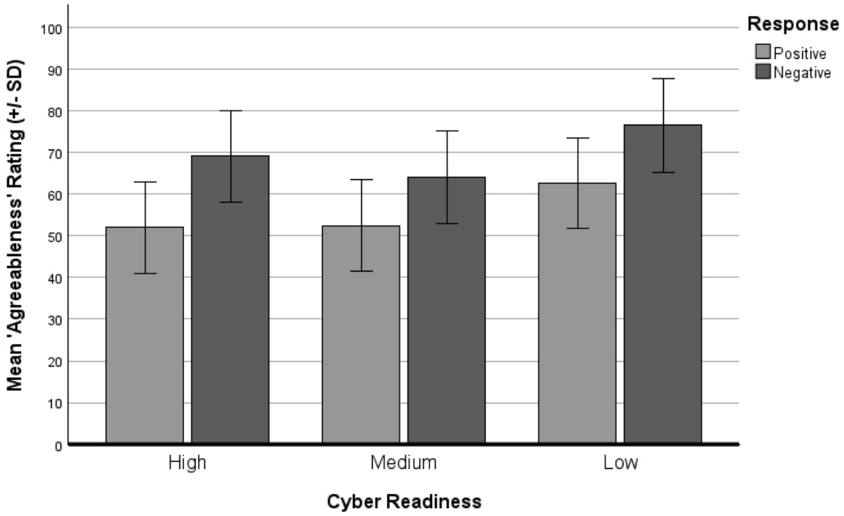

Fig. 10. Mean agreement rating with ‘The company was most to blame’ (error bars +/-SD).

Table 8
Descriptive findings for the statement ‘The company could have been better prepared’.

		IV1		
		High readiness	Medium readiness	Low readiness
IV2	Positive Response	*M* 77.11 *SD* 27.45	*M* 72.59 *SD* 25.91	*M* 85.07 *SD* 15.34
	Negative Response	*M* 88.93 *SD* 12.06	*M* 88.67 *SD* 13.28	*M* 82.00 *SD* 22.72

0.73, indicating that participants were significantly more likely to agree the response was appropriate in the positive cyber response condition compared to the negative condition (Fig. 12).

For the statement ‘The cyber attack on Vehicle X was inevitable’ (descriptives in Table 10) was analysed there was no significant interaction between cyber readiness and response, $F(2, 157) = 0.45$, $p = .64$ and no significant main effect for cyber response, $F(1, 157) = 1.24$, $p = .27$. There was however a significant main effect for cyber readiness $F(2, 157) = 8.64$, $p < .001$, $\eta_p^2 = 0.10$. Post hoc tests using Tukey's HSD indicated that agreement with idea the cyber attack was inevitable was significantly higher in the low than high cyber readiness condition ($p = .003$) and low than medium, $p < .001$ condition (Fig. 13).

To address concerns about potential construct coupling between the manipulation and the company being *most to blame*, the effect of response type on *most to blame* was re-estimated while adjusting for participants' evaluation of how appropriate the company's response was – ‘The self-driving car company responded appropriately to the cyber attack’. In the initial model, response type had a moderate effect ($b = -14.07$, $\beta = -0.24$), with negative responses producing higher blame. However, when participants' evaluation of the appropriateness of the response was included as a covariate, the effect of response type was substantially attenuated ($b = -0.55$) and the standardised coefficient reversed direction ($\beta = 0.65$), indicating a strong suppression effect. This pattern suggests that the effect of cyber response on blame is largely explained by a participant's perception of how appropriate the response is, rather than through response type alone.

Phase 3 – Follow-Up Questions.

A paired samples *t*-test compared whether attitudes about trusting and using AVs (not specifically Vehicle X) had changed during the experiment. There was a significant difference in the reported likelihood of trusting an AV before ($M = 36.64$, $SD = 28.10$) and after

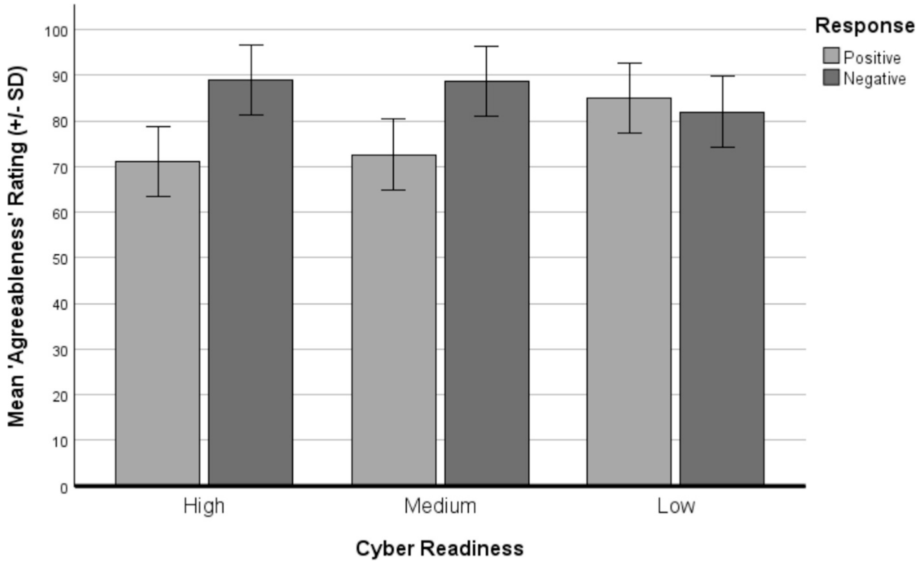

Fig. 11. Mean agreement rating with 'The company could have been better prepared' (error bars +/-SD).

Table 9
Descriptive findings for the statement 'The self-driving car company responded appropriately to the cyber attack.'

		IV1		
		High readiness	Medium readiness	Low readiness
IV2	Positive Response	*M* 70.93 *SD* 25.08	*M* 70.30 *SD* 20.77	*M* 74.93 *SD* 19.55
	Negative Response	*M* 8.19 *SD* 10.44	*M* 10.30 *SD* 14.17	*M* 14.46 *SD* 19.33

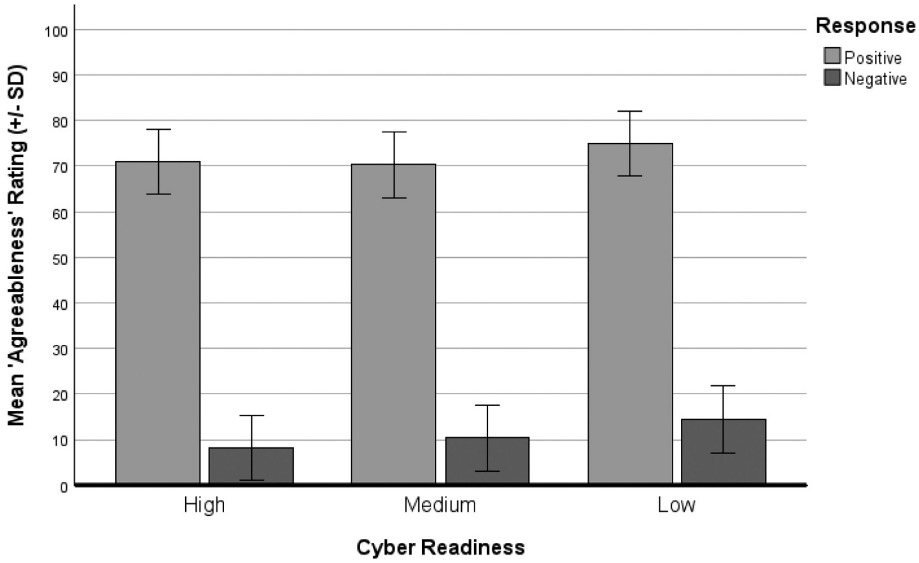

Fig. 12. Mean agreement rating with 'The self-driving car company responded appropriately to the cyber attack' (error bars +/-SD).

($M = 41.10$, $SD = 28.91$) the experiment, $t(162) = 2.83$, $p = .005$, $d = 0.16$. There was also a significant difference in the reported likelihood of using an AV before ($M = 49.77$, $SD = 32.85$) and after ($M = 42.57$, $SD = 32.53$) the experiment, $t(162) = 4.214$, $p < .001$, $d = 0.22$.

Finally free-text comments provided indications that despite being an "interesting study", it was potentially "unclear what 'cyber'

Table 10
Descriptive findings for the statement ‘The Cyber Attack was Inevitable’.

		IV1		
		High readiness	Medium readiness	Low readiness
IV2	Positive Response	*M* 39.32 *SD* 30.54	*M* 39.15 *SD* 30.49	*M* 60.39 *SD* 30.61
	Negative Response	*M* 50.52 *SD* 31.90	*M* 39.26 *SD* 32.13	*M* 65.23 *SD* 28.98

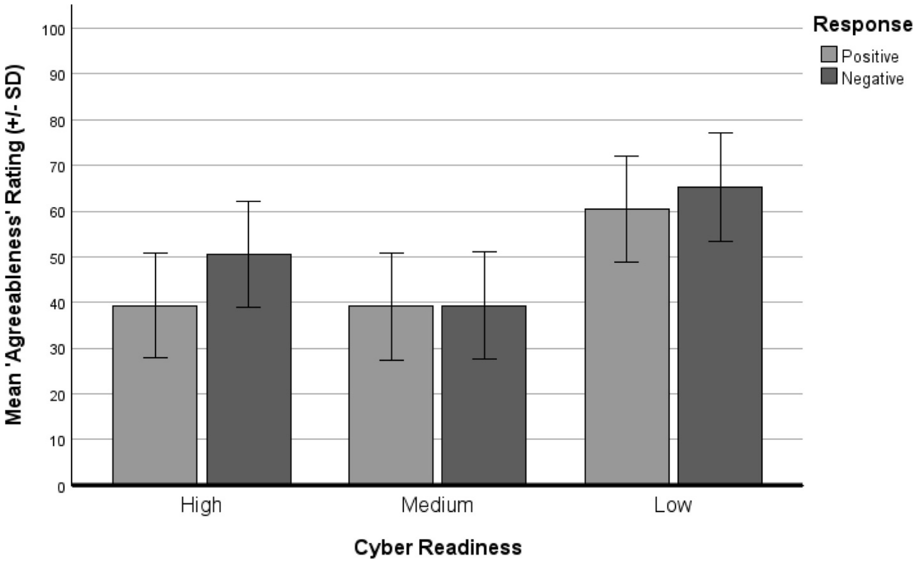

Fig. 13. Mean agreement rating with ‘The Cyber Attack was Inevitable’ (error bars +/-SD).

represented” to many, and some participants “didn't understand what [cyber] meant until the cyber attack occurred”. One pertinent comment highlighted that “the lack of cyber security measures in place changed everything. It is such a serious threat and cannot be understated. This changed my opinion of the company from ‘competent’ to ‘negligent’ immediately” demonstrating a lack of understanding which could be reflective of a wide mix of cyber literacy and understanding amongst the sample.

2.3. Experiment 1 discussion

Experiment 1 was designed to investigate the effects of cyber readiness and cyber response on trust rating and blame assignment in AVs following a cyber attack. Contrary to H1, cyber readiness did not affect initial trust in Vehicle X. One explanation for this could be that participant judgement required experience of being driven by Vehicle X before adjusting their trust levels. However, cyber readiness also did not affect trust (i.e. they did not differ between readiness conditions) after any of the events, including critical E5, and it also did not affect overall post-trust in Vehicle X. It therefore could be that participants did not fully understand or value cyber readiness in their judgements of the AV. This is emphasised in free-text comments provided by participants which indicated that the term ‘cyber’ was not well understood suggesting low cyber literacy amongst the sample. The lack of understanding of the term ‘cyber’ concurs with findings from a variety of different research areas which have found that cyber security was not top of mind for many potential adopters of AVs (Bansal & Kockelman, 2017). The lack of understanding of cyber - for many of the participants at least - could also explain why in this Experiment, out of ten desirable features of an AV, ‘cyber’ was perceived as a less desired feature than others listed e.g. design and was ranked last, when it should arguably be regarded as a more important feature. This concurs with findings from Pettigrew et al. (2018) whose qualitative study highlighted that while safety and privacy were recurring concerns, there was low articulation of cyber specific risks, and participants generally lacked understanding of how external parties could compromise vehicles (Pettigrew et al., 2018) thus undermining the perceived importance of cyber.

Cyber readiness did however significantly have an effect on initial trust in the company responsible for Vehicle X, as predicted in H4. The level of cyber readiness influenced how much initial trust participants placed in the company responsible for Vehicle X before experiencing an AV journey. The main difference was observed between ‘low’ and ‘high’ and ‘low’ and ‘medium’. Trust seemingly plateaus once the level of cyber readiness achieved is perceived to be at least ‘satisfactory’ (medium) as increasing cyber readiness from medium to high did not significantly increase initial trust in the company responsible for Vehicle X any further. This plateau effect could suggest that cyber readiness influences initial trust only up to a threshold, beyond which additional improvements yield diminishing returns (McKnight et al., 2002). However, this difference observed in initial trust of the company was not later observed throughout each of the events (including the critical event).

Cyber readiness also did not affect overall post-trust in the company responsible for Vehicle X. One explanation for this could be attributed to the dynamic nature of trust. Trust in automated systems, particularly in high-stakes environments such as AVs, is not static but develops and evolves as users accumulate experience (Lee & See, 2004). It is possible that a participant's initial trust was shaped by surface-level cues - the cyber readiness rating (Hancock et al., 2011; Muir, 1994). This preliminary rating, sometimes referred to as upfront trust (de Visser et al., 2016), may exert a strong influence early on, but are susceptible to revision once users observe system behaviour in real-time. As participants experienced Vehicle X during the experimental events, including the critical incident, their trust may have shifted from being cognitively driven (based on the star-ratings) to being more experience-based (based on observed reliability and performance). This aligns with research suggesting that trust is recalibrated as users gather direct evidence about a system's competence, predictability, and integrity (Hoff & Bashir, 2015; Madhavan & Wiegmann, 2007).

Furthermore, there were significant differences in trust in Vehicle X and the company responsible for Vehicle X between each of the events. One explanation for this could be the perceived riskiness of the events for example, trust was significantly higher in non-overtake events than overtake events, the latter which could be perceived as a riskier manoeuvre due to the complexity and potential for negative outcomes (Gold et al., 2015; Körber et al., 2019). Notably after the cyber attack event (E5) there was a marked drop in trust in both Vehicle X and the company responsible for Vehicle X. This decline is likely attributable to the perceived heightened risk of the situation as users tend to dynamically calibrate their trust based on situational cues related to safety and perceived system competence (Lee & See, 2004).

The experiment also explored the effect of cyber response on post-trust. Participants who experienced a positive cyber response reported significantly higher post-trust in Vehicle X compared to those who experienced a negative response as predicted in H2. This effect was consistent across all levels of cyber readiness. Regardless of the level of cyber readiness, the company's response after E5 to the cyber attack had a consistent effect on participants' post-trust with positive responses leading to higher post-trust in Vehicle X. These findings were also true for post-trust in the company responsible for Vehicle X as predicted in H5. These results underscore the critical role of an organisations behaviour during crisis events in shaping trust outcomes. Prior research has highlighted the importance of transparent, timely, and effective responses in maintaining or repairing trust following failures or adverse events (Lee & See, 2004; Mayer et al., 1995). How the company responds to a cyber event seems therefore to have a greater precedence on trust than cyber readiness. This suggests that participants based their trust on the actual outcomes and experience as opposed to consumed knowledge. A high cyber readiness rating did not amplify the benefits of a positive response or mitigate the effects of a negative one as predicted in H3 and H6 – post-trust levels were primarily influenced by the type of response alone.

Participants were also asked about likelihood of *trusting* AVs in the future, to ascertain whether cyber readiness and response toward a specific car such as Vehicle X could have the potential to affect trust in other AVs. Interestingly, participants in the high readiness condition who experienced a negative cyber response reported lower trust in other AVs afterwards. This might reflect a violation of trust as participants may have expected a ‘better’ cyber response but instead the negative cyber response could have triggered a sharper drop in trust - not just in Vehicle X, but in AVs overall. A negative cyber response may therefore be especially damaging to trust when expectations are high i.e. when a vehicle is described to have a high cyber readiness rating and those expectations are not met - possibly because it contradicts expectations of competence and readiness. It is also suggestive that under certain conditions (i.e. high cyber readiness; negative response) a single incident that is negatively responded to could influence trust in AVs beyond the vehicle that ascertains a cyber attack. It therefore seems that cyber response matters more than cyber readiness in shaping trust in AVs.

Participants were also asked about their likelihood of *using* AVs in the future to ascertain whether cyber readiness and response toward a specific car such as Vehicle X could have the potential to affect using other AVs. Overall, participants in the high readiness condition who experienced a negative cyber response reported a lower likelihood of using other AVs in the future compared to those with a positive response and same readiness level. Again, high cyber readiness could raise expectations and if an AV is seemingly mature but fails to respond effectively to incidents, it could be that disappointment is stronger and future usage intent drops more. A poor cyber response reduces participants' willingness to adopt the technology - even beyond the specific vehicle in question. Trust and intention to use are seemingly influenced by the type of cyber response to cyber incidents and an incident could affect not just trust in Vehicle X, but also perceptions of and attitudes toward other AVs.

In addition to examining trust in AVs, the experiment also investigated whether cyber readiness and cyber response influenced blame assignment. At the descriptive level, participants reported higher blame in the negative response condition than in the positive response condition, consistent with H7. Although response type initially appeared to influence blame, a sensitivity analysis revealed that this effect was largely accounted for by participants' perceptions of how appropriate the company's response was. This indicates that the observed blame differences were not driven directly by the response manipulation itself, but rather by participants' subjective judgement of how appropriately the company handled the incident. This suggests that a participants' subjective evaluation of the appropriateness of the company's cyber response plays a central role in blame attribution, over and above any direct effect of readiness level or response type. Perceived appropriateness therefore appears to operate as the central psychological mechanism linking post-incident responses to blame.

From a practical perspective, this implies that certain response types may be insufficient to limit impact and reputational damage following cyber incidents. Designing responses that are perceived as timely, transparent, and appropriate is likely to be critical for reducing blame and maintaining public trust in AVs following a cyber attack.

Overall, the findings from Experiment 1 suggest that cyber readiness and cyber response may influence trust and blame assignment in AVs. However, further research is needed to examine whether participants' cyber literacy and/or indeed the way in which cyber is communicated to participants early on in the Experiment – affects their initial trust, and whether these effects have the potential to persist after a cyber attack, once experience with the AV has been gained. Furthermore, this Experiment used participants from a single UK sample and therefore caution should be taken as the findings may not generalise to the diverse future user population. Further investigation should therefore consider whether comparable differences exist in populations both within as well as across different cultures.

3. Experiment 2a (UK) and experiment 2b (Japan)

In Experiment 1, the effects of cyber readiness and cyber response on trust in AVs and the company responsible for the AV following a cyber attack were investigated. There was perhaps an underestimation that participants would understand the meaning of the term cyber (security). However, cyber was ranked the lowest amongst nine other desirable features of an AV raising concerns about the participants level of cyber literacy which is known to mediate trust in technology-based systems and can also vary between countries (Kritzinger & Von Solms, 2010; Yang & Kim, 2025). In fact, countries such as the UK and Japan could demonstrate divergence across many dimensions including legal, infrastructural, industrial, sociocultural, and sociotechnical, as well as having similarities. These differences (and similarities) reflect broader national strategies that influence the pace and manner of AV integration into current transport systems.

For example, from a legislative perspective, the UK has arguably adopted a more proactive and flexible regulatory framework than Japan. The UKs Automated and Electric Vehicles Act (2018) laid the groundwork for AV liability and insurance, and the more recent Automated Vehicles Act (2024) has expanded this framework to formally recognise a 'user' and permit the deployment of AVs under defined conditions (CCAV, 2024). In contrast, Japan has pursued a more incremental and standards-driven regulatory trajectory regularly adapting their existing Road Traffic Act and Vehicle Act. The use of increasingly connected and autonomous vehicles in Japan is also closely aligned with UNECE regulatory standards, promoting international interoperability and safety compliance (UNECEa, 2023). Japan has also established a governmental working group - the Sub Working Group on Social Rules Concerning Automated Vehicles - under the Digital Agency to review and clarify, using a flexible and adaptable approach similar to the UK, how existing laws apply to AV use.

The UK and Japan also exhibit divergent yet complementary approaches to cyber security governance structures and managing cyber security risks. The UK has adopted a broad national cyber security framework, including its National Cyber Security Strategy, which emphasises resilience, public-private cooperation, and general cyber security preparedness across critical national infrastructure sectors (Cabinet Office, 2022). For connected autonomous vehicles (CAVs) specifically, the UK government, through the Department for Transport (DfT), the Centre for the Protection of National Infrastructure (CPNI) and the Centre for Connected and Autonomous Vehicles (CCAV), has issued high-level cyber security principles. These include recommendations on threat modelling, secure software development, and incident response (DfT, CPNI & CCAV, 2017). However, the UK has yet to fully implement legally binding cyber security mandates specific to AV systems, relying instead on voluntary guidance and industry-led implementation. By contrast, Japan's regulatory and operational strategy is characterised by early adoption of binding international standards, particularly the UNECE WP.29 forum and R155 and R156 - governing cyber security management systems and software update protocols (UNECE[a,] [b], 2021). This move has placed Japan amongst the global leaders in formal regulatory alignment for AV cyber security.

The Japanese approach is further bolstered by close collaboration between government agencies such as the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) and the National Institute of Information and Communications Technology (NICT), alongside some of Japan's major automotive manufacturers like Toyota, Honda, and Nissan. Another key differentiator between the UK and Japan is the integration of cyber risk assessment. In Japan, organisations have begun developing actuarial models that evaluate the economic implications of cyber attacks on AVs (VicOne Corporation, 2025). This level of foresight is largely absent in the UK context, where cyber risks in AVs have not yet been fully translated into structured insurance models or publicly available economic assessments. Unlike in the UK, a failure to meet certain standards could even prevent access to markets for automakers aiming to sell in countries such as Japan. Differences in legislative and cyber security approaches along with cultural distinctions could therefore offer a valuable insight into trust in AVs. Japan has long demonstrated high acceptance of automation and advanced technologies and its collectivist society in comparison to the UK individualistic society could ultimately enhance the understanding of trust in AV ecosystems.

These contextual and cultural differences provide a basis for comparing perceptions of AV cyber security risks across countries as perceptions of cyber security readiness and response may vary across national contexts. Experiment 2a (UK) and Experiment 2b (Japan) were therefore designed - largely based on Experiment 1 – to investigate whether the provision of definitions for the desired AV features, including cyber security, would shift perceptions of their importance and affect trust and blame assignment in Vehicle X and the company responsible for Vehicle X. Consequently, comparing responses across UK and Japanese samples provides an opportunity to examine how a cultural context may shape trust and blame attribution in AVs. In addition to the hypotheses outlined in Experiment 1 and noting the cultural differences in approaches to AV implementation and cyber security, it was also hypothesised that:

- H8 – The desirability of cyber security in relation to AVs will be ranked higher than in Experiment 1, when the term is defined.
- H9 – Trust in the AV itself will be lower in the UK than in Japan.
- H10 - Trust in the company responsible for the AV will be lower in the UK than Japan.
- H11 - An AV company will have less blame attributed in Japan than the UK.

3.1. Method

3.1.1. Participants

An a priori power analysis was conducted in G*Power 3.1 (Cohen, 1988; Faul et al., 2009) for a 3 × 2 between-subjects factorial ANOVA, powered to detect the cyber readiness × cyber response interaction ($df = 2$). Assuming a medium effect size (Cohen's $f = 0.25$), $\alpha = 0.05$, and power $(1 - \beta) = 0.80$, the required sample size was $N = 300$. This power analysis was intended for the primary between-subjects interaction test.

For Experiment 2a, 163 UK participants were recruited via *Prolific©* and randomly assigned to conditions until equal numbers were achieved. Eight participant datasets were not usable either due to being incomplete across multiple measures and/or due to failing attention checks. This resulted in 155 usable datasets. The sample consisted of 78 participants identifying as male; and seventy-seven identifying as female. Ages in the UK sample ranged from 20 to 79 ($M = 42$, $SD = 12.85$). One hundred and twenty-three held a full UK driving licence; twenty a provisional UK driving licence; one was in the process of obtaining a provisional UK driving licence; ten did not hold a UK driving licence and were not in the process of obtaining one; and one preferred not to say. On average, qualified drivers had held a licence for 19-years and drove ~6120 miles (9849-km) per year.

For Experiment 2b, 545 participants were recruited in Japan via the online crowdsourcing platform *Lancers* and were also randomly assigned to conditions until equal numbers were achieved in each. Of these, one-hundred-and-fifty-five Japanese participants were matched to the UK participants based on four factors: age; gender; likelihood of using an AV, and propensity to trust an AV. Table 11 presents the variable and associated descriptives. Although matched factors across the UK and Japan are well-balanced, the samples differ slightly on baseline trust $t(308) = 2.23$, $p = .03$, Cohen's $d = -0.25$, 95% CI [−0.48, −0.03]. However, the effect size is small. Analyses account for this variable where relevant. Inclusion criteria were the same as in Experiment 1 – and with Japanese participants having to be fluent in Japanese as a first or second language. Participants were renumerated GBP £3.75 / 700 JPY¥.

3.1.2. Materials

Unlike Experiment 1, in Experiment 2a (UK) and Experiment 2b (Japan) each desirable feature which formed the review of Vehicle X had a short ~31-word description (Appendix E). Six statements about potential factors that could influence beliefs, trust and likelihood of using AVs accompanied by the corresponding averages were also omitted from the preliminary phase. Everything else was as in Experiment 1. For Experiment 2b, the materials were translated from English into Japanese using a reputable agency and were checked by professors and research assistants familiar with the experimental design. Japanese participants were told Vehicle X acted in accordance with Japan's Rules of the Road which are largely akin to the UK Highway Code – which UK participants had been informed Vehicle X was acting in accordance with.

3.1.3. Design

Experiment 2a and Experiment 2b followed the same design as Experiment 1 (see Section 2.1.3).

3.1.4. Procedure

Experiment 2a and Experiment 2b followed the same procedure as Experiment 1 (see Section 2.1.4). However, when participants received the ten-feature star-rating review of Vehicle X, they were also asked to read the accompanying definition for each feature which detailed the meaning of: costs, comfort, cyber, design, environmental friendliness, practicality, performance, reliability, road worthiness and safety (Appendix E).

3.2. Results

Data was screened and checked as in Experiment 1. Data points that fell outside of the limits were reduced to the upper or increased to the lower limit. Following this process, the results are based on 310 usable datasets which were matched based on participant demographics.

Phase 1 - General Attitudes Toward AVs.

An independent samples *t*-test was performed to compare intentions to trust AVs in the UK and in Japan. There were significant differences between participants in the UK ($M = 41.21$, $SD = 28.09$) and Japan ($M = 47.86$, $SD = 24.29$); $t(308) = -2.23$, $p = .03$, $d = 0.25$. Japanese participants trusted AVs more than participants in the UK (Fig. 14). An independent samples *t*-test compared intentions

Table 11
Balance table of matched UK and Japan samples.

Variable	UK Mean / SD	Japan Mean / SD	Test	*p*-value	Cohen's *d*
Gender (% Male)	50.3%	50.3%	χ^2	1.00	N/A
Age	41.63 / 12.89	41.48 / 13.02	*t*-test	0.92	0.01
Trust in AV Technology	41.21 / 28.09	47.86 / 24.29	*t*-test	0.03	0.25
Likelihood of Using AV Technology	52.07 / 33.30	48.57 / 29.88	*t*-test	0.33	0.11

Japanese participants were aged between 20 and 77 ($M = 41$, $SD = 13.00$). One-hundred-and-forty-six held a full Japanese driving licence, and nine participants did not hold a Japanese driving licence and were not in the process of obtaining one. On average, qualified drivers had held a licence for 21-years and drove ~2790 miles (4490 km) per year.

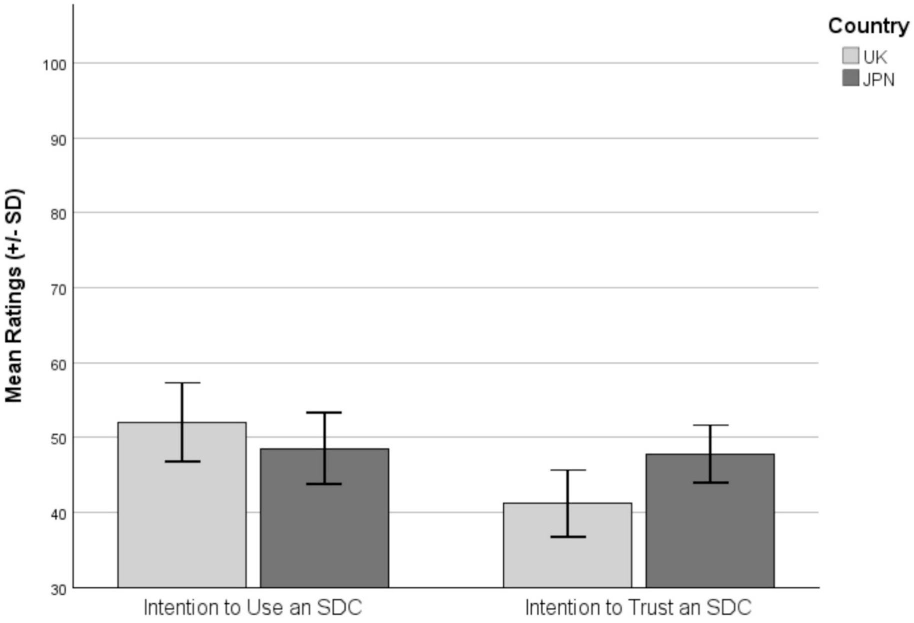

Fig. 14. Intentions of participants in the UK and Japan to use and trust any AV (error bars +/-SD).

to use AVs amongst UK and Japanese participants. There were no significant differences (UK: $M = 52.07$, $SD = 33.30$; Japan: $M = 48.57$, $SD = 29.88$); $t(308) = 0.98$, $p = .33$. Note: the descriptives (means) for the UK are largely consistent with findings in Experiment 1 (Table 3).

On a 101-point VAS where '0' was 'Fully Disagree' and '100' was 'Fully Agree', participants were then asked to what extent they agreed with 4 statements. To compare general expectations of AVs across the UK and Japan, independent samples *t*-tests were performed. For three of the statements, there were significant differences found in general expectations of AVs between participants in the UK and Japan (Table 12). Japanese participants were significantly less likely to agree that higher ratings are associated with quality or trust and were significantly less agreeable to the prospect of paying more for a product or service with higher ratings.

Phase 1 – Ranking of Desirable Features.

Next, desirable features of an AV were rated by participants using star-ratings between 0 (not that desired) and 5 (highly desired). Safety and road worthiness ranked as the most desirable features in the UK whilst in Japan, safety and reliability were top two ranking features. Design and environmental friendliness were ranked as the least desirable features in both the UK and Japan. In comparison to Experiment 1, cyber was no longer ranked the lowest and instead was ranked as the fifth most important feature in both countries. All rankings are presented in Fig. 15. Independent samples *t*-tests revealed significant differences for five of the desired feature-specific expectations between the UK and Japan. Participants in the UK had significantly higher expectations for Environmental Friendliness, $t(303) = 2.13$, $p = .03$, $d = 0.22$; Cyber, $t(301) = 2.86$, $p = .005$, $d = 0.24$; Reliability, $t(303) = 3.37$, $p = .001$, $d = 0.38$; Safety, $t(303) = 7.94$, $p < .001$, $d = 0.91$ and Road Worthiness, $t(303) = 8.38$, $p < .001$, $d = 0.97$. There were no significant differences in the other features between the UK and Japan.

Phase 2 – Trust.

Table 12
Statements and corresponding averages for the UK and Japan.

Statement	UK	Japan	Independent Samples t-test
Before using a self-driving car, I would find out as much information as I could about its specification	*M* 77.47 *SD* 20.27	*M* 79.74 *SD* 20.28	$t(308) = -0.99$, $p = .33$.
I believe that products and services with higher ratings are better quality than those with lower ratings	*M* 80.35 *SD*16.98	*M* 72.05 *SD* 17.25	$t(308) = 4.27$, $p < .001$*, $d = 0.49$.
I believe that products and services with higher ratings can be trusted more than those with lower ratings	*M* 81.21 *SD* 16.11	*M* 72.55 *SD* 17.49	$t(308) = 4.53$, $p < .001$*, $d = 0.52$.
I would be willing to pay more for a product or service with higher ratings	*M* 76.05 *SD* 18.87	*M* 58.36 *SD* 15.85	$t(308) = 8.93$, $p < .001$*, $d = 1.05$.

* Significant differences.

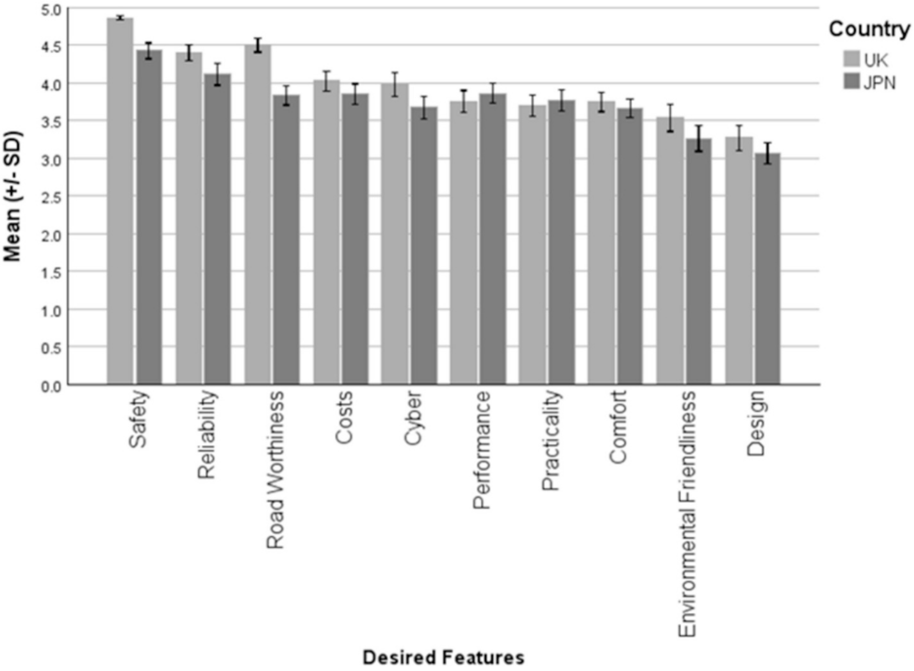

Fig. 15. UK and Japan Rankings of Desired Features in an AV (error bars +/-SD).

Prior to experiencing the events, initial trust in Vehicle X (based solely on the information provided) was assessed. A 3 (cyber readiness: high, medium, low) × 2 (country: UK, Japan) two-way between-subjects ANOVA examined initial trust. There was no significant interaction between cyber readiness and country, $F(2,304) = 0.02$, $p = .98$. There was however a significant main effect in initial trust in Vehicle X based on the level of cyber readiness $F(2, 304) = 13.93$, $p < .001$, $\eta_p^2 = 0.08$. Post hoc Tukey HSD tests revealed significant differences in initial trust in Vehicle X across both countries for high and low cyber readiness: UK, $p < .001$; JPN, $p < .001$ and medium and low cyber readiness: UK, $p = .01$; JPN $p = .01$. There was also a significant main effect for country, $F(1,304) = 26.21$, $p < .001$, $\eta_p^2 = 0.08$, indicating Japanese participants being initially less trusting of Vehicle X (Fig. 16). All other post hoc tests were non-significant.

Initial trust in the company responsible for Vehicle X was analysed using the same 3 × 2 ANOVA. There was no significant interaction between cyber readiness and country, $F(2, 304) = 0.70$, $p = .50$ and no significant main effect for country, $F(1, 304) = 0.00$,

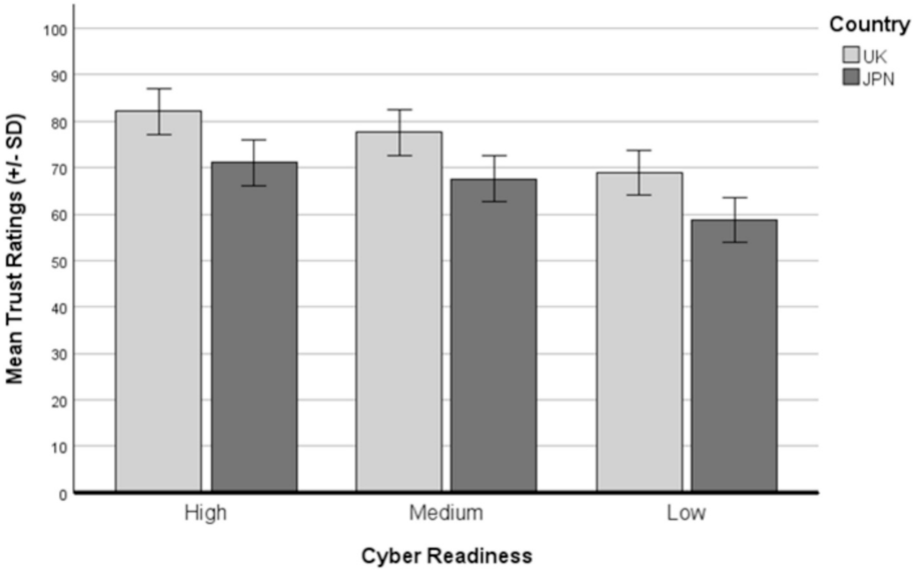

Fig. 16. Initial trust ratings UK/Japan in Vehicle X depending on cyber readiness (error bars +/-SD).

$p < .99$. There was however a significant main effect for cyber readiness $F(2, 304) = 19.18$, $p < .001$, $\eta_p^2 = 0.11$. Post hoc Tukey HSD tests revealed significant differences in intial trust in the company responsible for Vehicle X between high and low cyber readiness: UK, $p < .001$; JPN $p < .001$ and medium and low cyber readiness: UK, $p < .001$; JPN $p = .02$. All other post hoc tests were non-significant (Fig. 17).

Trust Across Events (E1–E4).

A 3 (cyber readiness: high, medium, low) × 2 (country: UK, Japan) × 4 (event: E1–E4) mixed-design ANOVA was conducted to examine trust across the events in Vehicle X and in the company responsible for Vehicle X. Cyber readiness and country were treated as between-subjects factors, and event was treated as a within-subjects factor.

Vehicle X.

The three-way interaction between event, country, and cyber readiness was not significant, $F(6, 912) = 0.15$, $p = .99$, $\eta_p^2 = 0.00$. Likewise, the event × cyber readiness interaction was not significant, $F(6, 912) = 0.74$, $p = .62$, $\eta_p^2 = 0.01$, nor was the cyber readiness × country interaction, $F(2, 304) = 0.08$, $p = .93$, $\eta_p^2 = 0.00$. However, the event × country interaction was significant, $F(3, 912) = 7.14$, $p < .001$, $\eta_p^2 = 0.02$, suggesting that the pattern of trust across events differed between UK and Japanese participants. Pairwise comparisons showed that UK participants reported significantly higher trust than Japanese participants at E2 (mean difference = 7.39, $p < .001$) and E4 (mean difference = 9.63, $p < .001$). No significant differences were observed at E1 ($p = .06$) or E3 ($p = .84$). There was a significant main effect of event, $F(3, 912) = 64.22$, $p < .001$, $\eta_p^2 = 0.17$, and a significant main effect of country, $F(1, 304) = 11.23$, $p < .001$, $\eta_p^2 = 0.04$, with UK participants ($M = 67.44$, $SD = 1.12$) reporting higher overall trust than Japanese participants ($M = 62.13$, $SD = 1.12$). The main effect of cyber readiness was not significant, $F(2, 304) = 2.21$, $p = .11$, $\eta_p^2 = 0.01$.

Trust in the Company.

The same 3 × 2 × 4 mixed-design ANOVA was conducted for trust in the company responsible for Vehicle X. The three-way interaction between event, country, and cyber readiness was not significant, $F(6, 912) = 0.51$, $p = .80$, $\eta_p^2 = 0.00$. The two-way interactions between country × cyber readiness, $F(2, 304) = 1.01$, $p = .36$, $\eta_p^2 = 0.01$, event × country, $F(3, 912) = 2.32$, $p = .07$, $\eta_p^2 = 0.01$, and event × cyber readiness were not significant, $F(6, 912) = 1.47$, $p = .18$, $\eta_p^2 = 0.01$. There was however a significant main effect of event, $F(3, 912) = 29.54$, $p < .001$, $\eta_p^2 = 0.09$, indicating that trust in the company varied across E1-E4. There was also a significant main effect of cyber readiness, $F(2, 304) = 4.28$, $p = .02$, $\eta_p^2 = 0.03$. Tukey HSD post hoc tests indicated that trust was significantly higher in the high compared to low readiness condition ($p = .004$). All other post hoc tests were non-significant. For the between-subjects effects, there was a significant main effect of country, $F(1, 304) = 8.33$, $p = .004$, $\eta_p^2 = 0.03$, with UK participants ($M = 69.\ 44$, $SD = 1.26$) reporting higher overall trust than Japanese participants ($M = 64.31$, $SD = 1.26$).

Trust After E5.

Vehicle X.

After the cyber incident (E5), there was a marked decline in trust in Vehicle X (Figs. 18 and 19) and the company responsible for Vehicle X (Figs. 20 and 21). At this stage, participants were aware of the cyber attack but had not been informed as to how the company responded. A 3 (cyber readiness: high, medium, low) × 2 (country: UK, Japan) two-way between-subjects ANOVA examined trust in Vehicle X after E5. There was no significant interaction in trust in Vehicle X between cyber readiness and country, $F(2,304) = 1.09$, $p = .34$ and no significant main effect in trust in Vehicle X between countries $F(1,304) = 1.42$, $p = .23$. There was however a significant main effect in trust in Vehicle X depending on the level of cyber readiness (high, medium, low) $F(2, 304) = 4.52$, $p = .01$, $\eta_p^2 = 0.03$. Pairwise comparisons indicated that trust was significantly higher in the high compared to low readiness condition (mean difference = 5.88, $p = .01$) and in the high compared to medium readiness condition (mean difference = 6.50, $p = .007$). There was no significant difference between the medium and low readiness conditions ($p = .79$).

Trust in the Company.

The same 3 × 2 ANOVA was conducted for trust in the company responsible for Vehicle X. There was a significant interaction between cyber readiness and country, $F(2, 304) = 3.62$, $p = .03$, $\eta_p^2 = 0.02$, indicating that the effect of cyber readiness on trust differed between the UK and Japan. Simple effects analyses revealed that, in the UK sample, trust was significantly lower in the low readiness condition compared to both the high ($p < .001$) and medium ($p = .006$) readiness conditions. No significant differences between readiness conditions were observed in the Japanese sample (all $ps > 0.47$). There were no significant country differences within any readiness condition (all $ps > 0.09$).

A repeated-measures ANOVA examining trust in Vehicle X across the five events (E1–E5) and revealed a significant main effect of event, $F(3.38, 1027.30) = 334.15$, $p < .001$, $\eta_p^2 = 0.52$, indicating that trust varied across the driving scenarios. Like Experiment 1, pairwise comparisons showed that trust was significantly lower during the overtake events (E1 and E3) than during the non-overtake events (E2 and E4) and was lowest following the cyber incident (E5), which differed significantly from all preceding events (all $ps < 0.001$). There was also a significant event × country interaction, $F(3.38, 1027.30) = 8.55$, $p < .001$, $\eta_p^2 = 0.03$, indicating that the pattern of trust change across events differed between UK and Japanese participants. All other interactions were not significant.

A 3 (cyber readiness: high, medium, low) × 2 (cyber response: positive, negative) × 2 (country: UK, Japan) between-subjects ANOVA was conducted on post-event trust in Vehicle X, measured following disclosure of the company's cyber response (Figs. 22 and 23). There was no significant three-way interaction between country, cyber readiness and cyber response, $F(2, 298) = 0.53$, $p = .59$. The cyber readiness × cyber response and cyber response × country interactions, as well as the main effects of cyber readiness and country, were not significant (all $ps > 0.05$). However, there was a significant two-way interaction between cyber readiness and country, $F(2, 298) = 6.51$, $p = .002$, $\eta_p^2 = 0.04$. Simple main effects analysis indicated that UK participants ($M = 20.05$, $SD = 3.27$) were less trusting of Vehicle X in the low cyber readiness condition compared to Japanese participants ($M = 37.26$, $SD = 3.27$). No other simple main effects were significant. There was also a main effect for cyber response $F(1, 298) = 25.01$, $p < .001$, $\eta_p^2 = 0.08$. In both countries, participants who received a positive (UK: $M = 37.63$, $SD = 2.70$; JPN: $M = 43.63$, $SD = 2.70$) cyber response compared to a

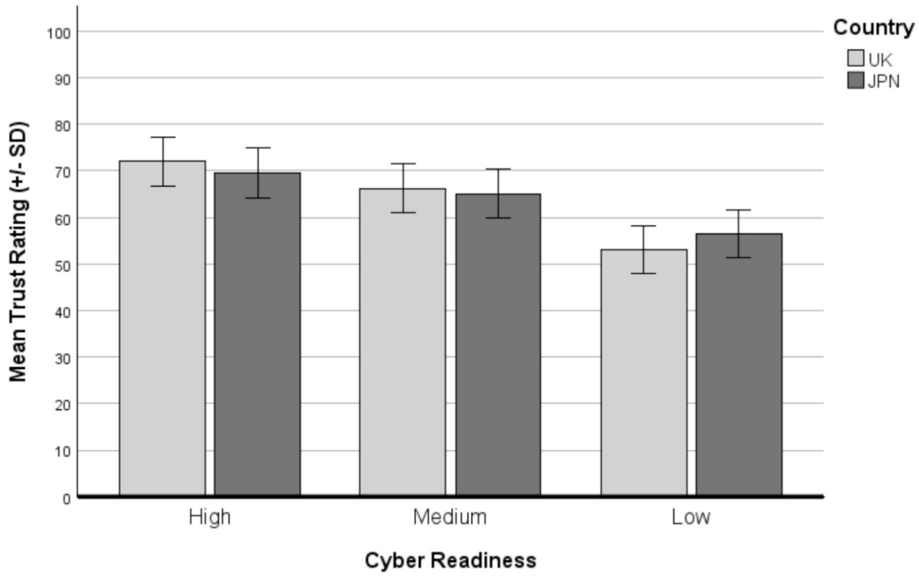

Fig. 17. Initial trust ratings UK/Japan in company responsible for Vehicle X depending on cyber readiness (error bars +/-SD).

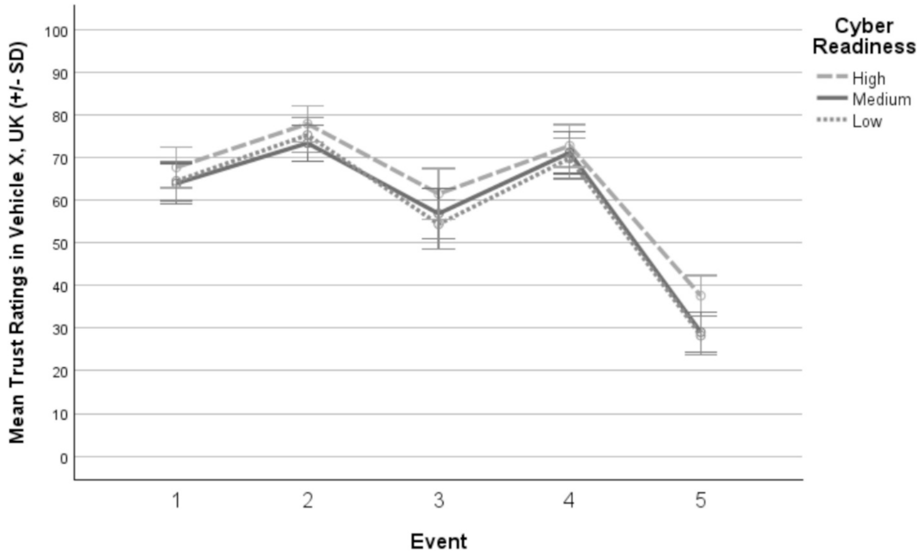

Fig. 18. Mean Trust Ratings in in the UK for Vehicle X following each event E1 – E5 (error bars +/-SD).

negative cyber response (UK: $M = 26.44$, $SD = 2.75$; JPN: $M = 27.51$, $SD = 2.75$) were more trusting of Vehicle X after all events.

The same 3 (cyber readiness: high, medium, low) × 2 (cyber response: positive, negative) × 2 (country: UK, Japan) between-subjects ANOVA was conducted on post-event trust in the company responsible for Vehicle X (Figs. 24 and 25). There was no significant three-way interaction between country, cyber readiness and cyber response, $F(2, 298) = 0.09$, $p = .92$. The cyber readiness × cyber response and cyber response × country interactions as well as the main effects of cyber readiness and country, were not significant (all $ps > 0.05$). There was however a significant two-way interaction between cyber readiness and country, $F(2, 298) = 5.88$, $p = .003$, $\eta_p^2 = 0.04$. Simple main effects analysis indicated that UK participants ($M = 16.95$, $SD = 2.99$) were less trusting of the company responsible for Vehicle X in the low cyber readiness condition compared to Japanese participants ($M = 37.42$, $SD = 2.99$). No other simple effects were significant. There was also a significant main effect of cyber response $F(1, 298) = 60.49$, $p < .001$, $\eta_p^2 = 0.17$. In both countries, participants who received a positive cyber response (UK: $M = 37.74$, $SD = 2.47$; JPN: $M = 48.20$, $SD = 2.47$) compared to negative (UK: $M = 20.11$, $SD = 2.52$; JPN: $M = 26.96$, $SD = 2.52$) were more trusting of the company responsible for Vehicle X after each event.

To examine whether overall trust changed following exposure to the experimental scenario, a 2 (exposure: initial trust, post trust) × 2 (country: UK, Japan) mixed-design ANOVA was conducted. There was a significant main effect of exposure, $F(1, 308) = 384.34$, p

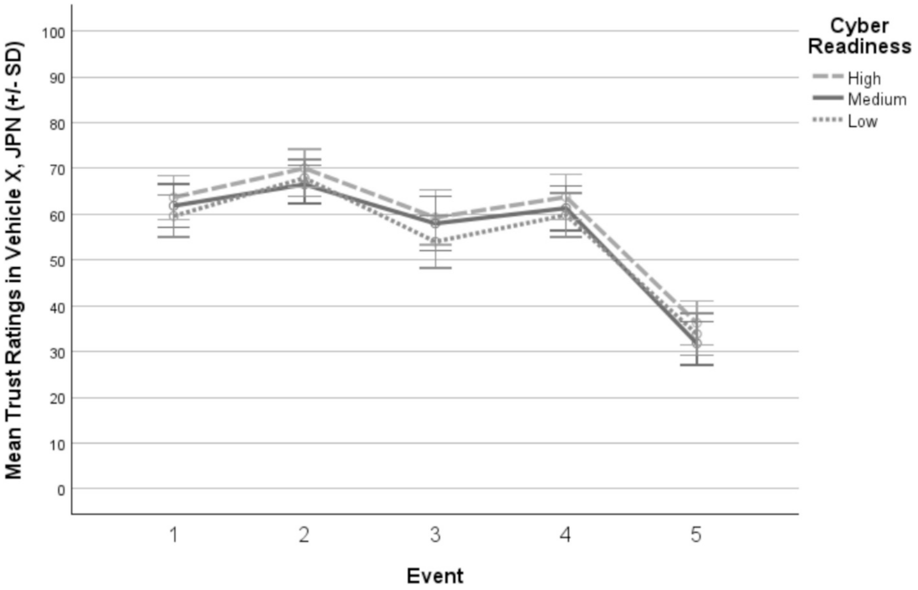

Fig. 19. Mean Trust Ratings in Japan for Vehicle X following each event E1 – E5 (error bars +/-SD).

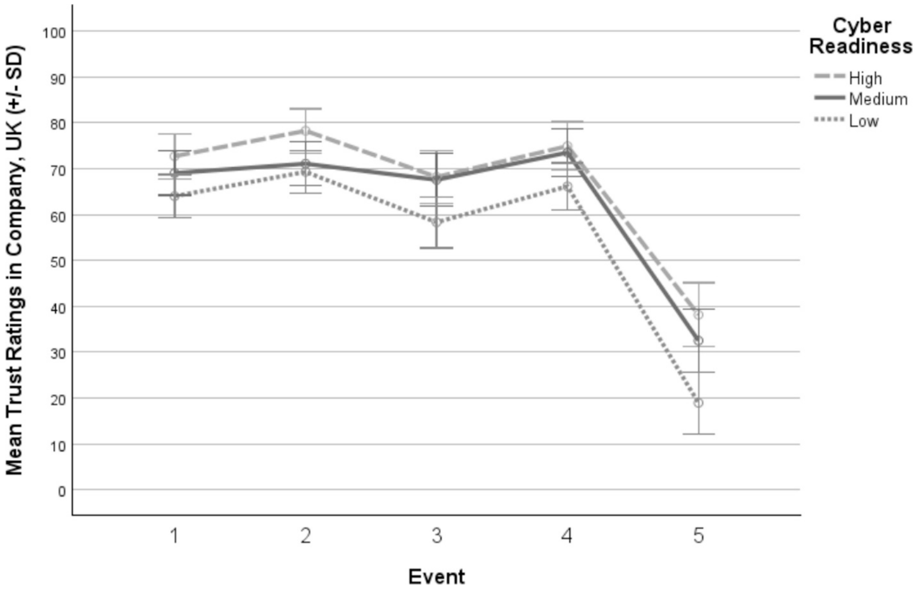

Fig. 20. Mean Trust Ratings in the UK for the company responsible for Vehicle X following each event E1 – E5 (error bars +/-SD).

< .001, $\eta_p^2 = 0.56$, indicating that trust in Vehicle X was substantially lower after the experiment than at the outset. This effect was qualified by a significant trust × country interaction, $F(1, 308) = 17.16$, $p < .001$, $\eta_p^2 = 0.05$. Pairwise comparisons showed that UK participants reported higher trust than Japanese participants prior to experiencing the events (mean difference = 10.45, $p < .001$), whereas trust did not differ significantly between countries following the events (mean difference = 2.05, $p = .48$). Descriptive statistics indicated initial trust of 76.24 ($SD = 1.44$) for the UK and 65.79 ($SD = 1.44$) for Japan, and initial trust of 32.22 ($SD = 1.99$) and 35.68 ($SD = 1.99$), respectively.

Phase 2 - Blame.

Responses to the following four blame attribution statements were analysed using the same statistical approach. For each statement, a 3 (cyber readiness: high, medium, low) × 2 (cyber response: positive, negative) × 2 (country: UK, Japan) between-subjects ANOVA was conducted to examine agreement with the statement.

For the statement ‘The self-driving car company is most to blame for the occurrence of the cyber attack’, the three-way interaction between cyber readiness, cyber response, and country was not significant, $F(2, 298) = 0.83$, $p = .44$, nor were the cyber readiness ×

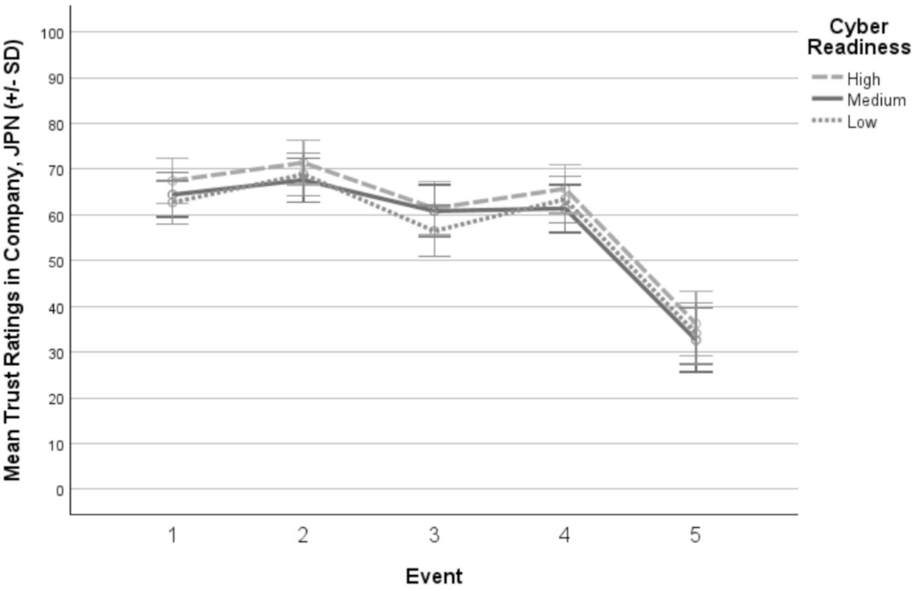

Fig. 21. Mean Trust Ratings in Japan for the company responsible for Vehicle X following each event E1 – E5 (error bars +/-SD).

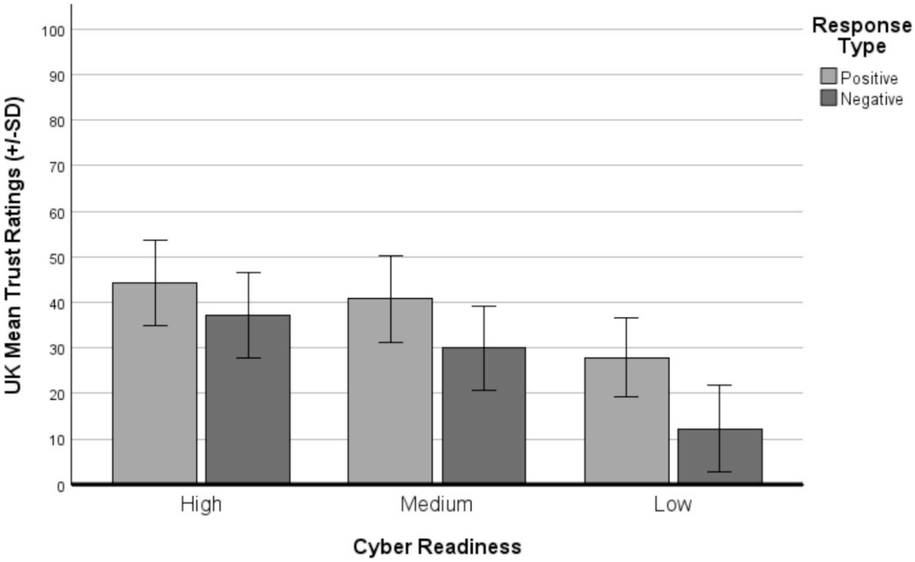

Fig. 22. Post-Trust Ratings in Vehicle X in the UK (error bars +/-SD).

cyber response or country × cyber response interactions (all *ps* $\geq$ 0.40). However, the cyber readiness × country interaction was significant, $F(2, 298) = 3.17$, $p = .044$, $\eta p^2 = 0.02$. Pairwise comparisons indicated that UK participants reported greater agreement than Japanese participants in the low readiness condition under both negative ($p = .019$) and positive ($p = .045$) response conditions, and in the medium readiness condition when the response was negative ($p < .001$). No country differences were observed in the high readiness condition or in the medium readiness condition when the response was positive. There was also a significant main effect for cyber response, $F(1, 298) = 14.62, p < .001$, $\eta_p^2 = 0.05$. Participants who received a negative response (UK: $M = 70.63$, $SD = 2.86$; JPN $M = 57.86$, $SD = 2.86$) compared to positive cyber response (UK: $M = 57.47$, $SD = 2.811$; JPN $M = 49.34$, $SD = 2.81$) were more likely to agree that the AV company was most to blame. All other tests were non-significant.

For the statement ‘Based on the star-ratings, the self-driving car company could have been better prepared for the cyber attack on Vehicle X'. The three-way interaction between country, cyber readiness, and cyber response was not significant, nor were any two-way interactions (all *ps* > 0.05). There was a significant main effect of country, $F(1, 298) = 14.61, p < .001$, $\eta_p^2 = 0.047$, indicating that UK participants reported greater agreement than Japanese participants overall. There was also a significant main effect for cyber response $F(1,298) = 16.73, p < .001$, $\eta_p^2 = 0.05$ such that participants who received a negative cyber response reported higher agreement that

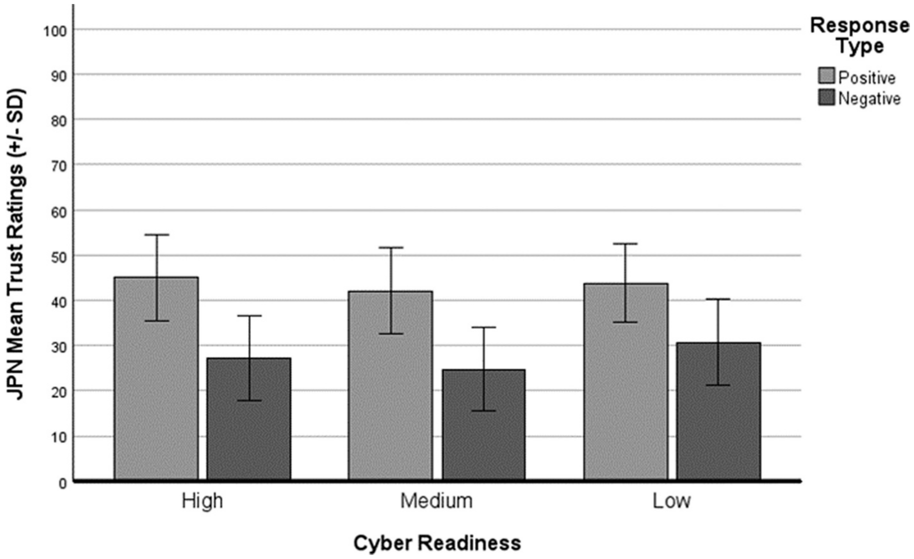

Fig. 23. Post-Trust Ratings in Vehicle X in Japan (error bars +/-SD).

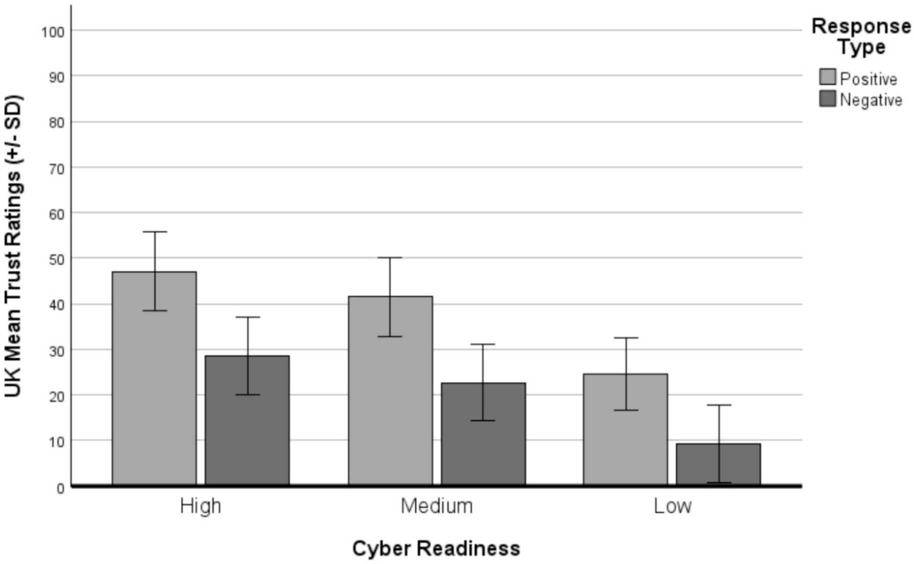

Fig. 24. Post-Trust Ratings in the company responsible for Vehicle X in the UK (error bars +/-SD).

the company could have been better prepared (UK: $M = 84.13$, $SD = 2.50$; JPN: $M = 74.00$, $SD = 2.50$) compared to those with a positive response (UK: $M = 73.23$, $SD = 2.46$; JPN: $M = 64.50$, $SD = 2.46$). All other tests were not significant.

For the statement 'The self-driving car company responded appropriately to the cyber attack', the three-way interaction between country, cyber readiness, and cyber response was not significant, $F(2, 298) = 1.08$, $p = .34$, $\eta p^2 = 0.01$. The cyber readiness × cyber response interaction was also not significant, $F(2, 298) = 0.55$, $p = .58$, $\eta p^2 = 0.00$. There was however a significant two-way interaction between country × cyber response, $F(1, 298) = 12.66$, $p < .001$, $\eta p^2 = 0.04$. Simple main effects analyses for the country × cyber response interaction showed that participants in both countries perceived the company's response as more appropriate following a positive response than a negative response, but this difference was larger in the UK. In the positive response condition, UK participants reported higher agreement ($M = 71.03$, $SD = 2.48$) than Japanese participants ($M = 57.72$, $SD = 2.48$), $p < .001$. There was also a significant two-way interaction between country × cyber readiness, $F(2, 298) = 7.47$, $p < .001$, $\eta_p^2 = 0.05$, indicating that the effects of response and readiness differed between the UK and Japan. Simple effects indicated that in the UK, agreement that the company responded appropriately was higher in the high readiness condition ($M = 56.02$, $SD = 3.10$) than in the medium ($M = 42.26$, $SD = 3.07$) and low readiness conditions ($M = 38.81$, $SD = 2.99$), whereas no significant differences between readiness conditions were observed in the Japanese sample. In addition, there was a significant main effect of cyber response, $F(1, 298) = 307.83$, $p < .001$, $\eta_p^2 =$

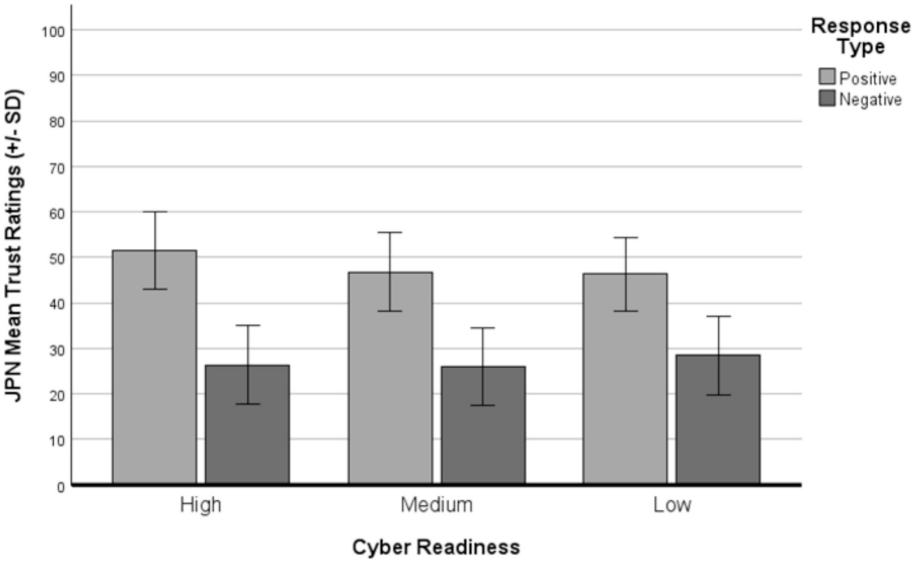

Fig. 25. Post-Trust Ratings in the company responsible for Vehicle X in Japan (error bars +/-SD).

0.51, such that participants who received a positive response ($M = 64.38$, $SD = 1.75$) reported greater agreement that the company responded appropriately than those who received a negative response ($M = 20.58$, $SD = 1.78$). All other tests were non-significant.

For the statement 'The cyber attack on Vehicle X was inevitable'. The three-way between-subjects ANOVA revealed no significant interactions between country, cyber readiness, and cyber response (all $ps \geq 0.051$). There was a significant main effect of country, $F(1, 297) = 8.68$, $p = .003$, $\eta_p^2 = 0.03$, indicating that UK participants reported greater agreement that the cyber attack was inevitable than Japanese participants (UK: $M = 51.91$, $SD = 2.17$; JPN: $M = 42.97$, $SD = 2.16$). The main effects of cyber readiness, $F(2, 297) = 1.19$, $p = .31$, and cyber response, $F(1, 297) = 1.20$, $p = .28$, were not significant.

To address concerns about potential construct coupling between the manipulation and the company being *most to blame*, the effect of response type on *most to blame* was re-estimated while adjusting for participants' evaluation of how appropriate the company's response was - The self-driving car company responded appropriately to the cyber attack. In the initial model, the manipulation had a moderate effect on blame ($b = -10.79$, $\beta = -0.21$). However, when participants' evaluation of the appropriateness of the response was included as a covariate, the effect of the manipulation was greatly reduced and became small ($b = 3.70$, $\beta = 0.07$). This pattern indicates that the influence of response type on blame operates largely through participants' subjective evaluation of the appropriateness of the company's response.

Phase 3 – Follow-Up Questions.

A 3 (cyber readiness: high, medium, low) × 2 (cyber response: positive, negative) × 2 (country: UK, Japan) three-way between-subjects ANOVA was conducted to examine the likelihood of trusting *any* other AV. The three-way interaction between cyber readiness, cyber response, and country was not significant, $F(2, 298) = 1.52$, $p = .22$, $\eta p^2 = 0.01$. All two-way interactions were also non-significant (all $ps \geq 0.18$). There were no significant main effects of country, $F(1, 298) = 0.37$, $p = .55$, $\eta_p^2 = 0.001$, or cyber readiness, $F(2, 298) = 1.57$, $p = .21$, $\eta_p^2 = 0.01$. The main effect of cyber response was also not significant, $F(1, 298) = 3.65$, $p = .057$, $\eta p^2 = 0.01$, although this effect approached significance, indicating that participants who received a negative cyber response reported lower trust than those who received a positive response. This trend provides tentative evidence that organisational responses may shape trust at the level of the wider AV ecosystem, rather than solely affecting perceptions of the specific vehicle involved.

The same ANOVA was used to examine the effects of country, cyber readiness, and cyber response on post intentions to use *any* AV. The three-way interaction was not significant, $F(2, 298) = 1.18$, $p = .31$, $\eta p^2 = 0.01$, and all two-way interactions were also non-significant (all $ps \geq 0.19$). There was a significant main effect of cyber response, $F(1, 298) = 6.18$, $p = .013$, $\eta p^2 = 0.02$, such that participants who received a negative cyber response reported lower intentions to use *any* AV in the future than those who received a positive response. The main effects of country, $F(1, 298) = 0.01$, $p = .91$, and cyber readiness, $F(2, 298) = 0.27$, $p = .77$, were not significant.

3.3. Experiment 2a and 2b discussion

Experiments 2a and 2b built on Experiment 1 by adding definitions of desirable AV features (including cyber) to investigate the effects of cyber readiness and cyber response on trust and blame assignment in AVs following a cyber attack. Some of the non-significant findings relating to key hypotheses in Experiment 1 could have been due to some participants not adequately understanding the meaning of cyber when presented with the term during Phase 1. Experiments 2a and 2b were also designed to explore similarities and differences between citizens from two countries and unique cultures - UK and Japan. In Experiment 1, "cyber" was ranked as the least desirable feature. After providing participants with a clear definition in Experiment 2, its desirability increased in line with predictions of H8, rising to fifth place out of ten features in both the UK and Japan samples, narrowly behind cost in fourth position. This change indicates that clarifying the meaning of the cyber feature led participants to view cyber more favourably and to assign it greater relative desirability. In contrast, the rankings of most other features showed little change between experiments, suggesting that these features were already well understood.

In line with Experiment 1 and predictions H1 and H4, cyber readiness affected initial trust in Vehicle X and the company responsible for Vehicle X, similarly in both the UK and Japan., Participants who were told that Vehicle X and the company responsible for Vehicle X had a higher cyber readiness rating, demonstrated higher levels of trust. This effect could be attributed to the provision of definitions added to the cyber readiness rating which likely served as a surface-level cue that shaped a participant's initial trust rating (Hancock et al., 2011; Muir, 1994;). Participants may have associated the readiness rating with the vehicles perceived security or reliability of automated systems, two factors which significantly affect initial trust (e.g., Hoff & Bashir, 2015; Lee & See, 2004).

Despite differences in initial trust in Vehicle X, differences in trust during E1-E4 in both Vehicle X and the company responsible for Vehicle X largely did not exist in either the UK or Japan. These findings were consistent with Experiment 1 and are likely attributable to the dynamic nature of trust and its recalibration as users gather direct evidence about a system's competence, predictability, and integrity (Hoff & Bashir, 2015; Madhavan & Wiegmann, 2007). After the critical event (E5), like in Experiment 1, there was a marked drop in trust in Vehicle X and the company in both countries. Interestingly, amongst the UK sample the cyber readiness rating impacted trust in Vehicle X itself as predicted in H1. This suggests that the participants initial trust was likely based on the information provided about cyber readiness rather than other factors such as cultural predispositions. It appears that this cue was retained in memory in line with predictions of H3 and H6, and while it did not influence trust when the system operated effectively, it did affect trust when the system experienced a cyber attack, which is consistent with findings that initial trust can shape subsequent responses to violations or failures (Lee & See, 2004; Muir, 1987).

The level of cyber readiness also impacted trust after E5 in the company responsible for Vehicle X. There was a significant interaction between cyber readiness and country which suggests that participants in the UK and Japan reacted differently to trusting the company depending on how well prepared it was. When the company had a low cyber readiness rating, in line with predictions in H10, participants in the UK were less trusting than participants in Japan despite receiving the same information. The fact that country did not have a significant main effect on trust yet significantly interacted with cyber readiness could indicate that trust is evaluated and adjusted in response to violations or risks differently across cultures. It may also reflect cultural differences in perceived organisational responsibility or tolerance for failure. For example, UK participants' exhibiting lower trust in the low cyber readiness condition may reflect the countries individualistic culture that place a higher premium on proactive risk management and institutional accountability (Hofstede, 2001). In contrast, Japanese participants appeared more forgiving, maintaining higher trust levels of trust in the company even after the AV experienced a cyber incident. This might be because participants interpreted organisational failure within a broader situational context, consistent with research suggesting that collectivist cultures are more tolerant of organisational shortcomings when efforts (even if insufficient) are perceived to have been made (Hofstede, 2001; Yamagishi, 2001).

The Experiments were also designed to explore possible effects of cyber readiness and cyber response on post-trust in both Vehicle X and the company responsible for Vehicle X. A two-way interaction between cyber readiness and country again showed that UK participants were more sensitive to the company's lack of preparedness, potentially reflecting cultural differences in how organisational competence and risk are evaluated. It also suggests that participants based their post-trust on the actual outcomes, experience as well as the information they had received about the AV. There was also a significant main effect of cyber response as predicted in H2 and H5. Regardless of country, participants who were presented with a positive cyber response following the critical event were more trusting of Vehicle X and the company responsible for Vehicle X than those who received a negative response. This reinforces the idea that how a company responds to a cyber incident – for example, whether it communicates transparently, takes responsibility, and resolves the issue effectively - can play a critical role in restoring or maintaining user trust when something goes wrong. This is consistent with Mayer et al.'s (1995) trust model, where perceptions of integrity and benevolence - in addition to ability - contribute to trust, especially in recovery situations.

During the Phase 3, participants were asked about trust in and likelihood of using *any* other AV. As in Experiment 1, this broadened the focus to ascertain whether cyber readiness and cyber response toward a specific car such as Vehicle X could have the potential to affect trust in other AVs. The absence of significant effects for cyber readiness and cyber response on trust in other AV suggests that trust in Vehicle X did not generalise to *any* AV as a broader category. This indicates that participants' trust judgements were system-specific rather than system-wide, supporting the view that trust in automation is calibrated to the perceived characteristics and experiences of a particular system rather than to similar technologies. It also suggests that participants did not interpret the cyber attack as an issue of safety or reliability of AV driving technology more generally. This finding is theoretically important because it implies that negative cyber security events may damage trust in a specific manufacturer or system without necessarily undermining acceptance of the wider technology, thereby suggesting a degree of compartmentalisation in the formation of trust and user attitudes toward AVs. This finding is however, cautionary due to the results only approaching significance.

Participants were also asked questions about how they would assign blame in the given situation. Both readiness and response had different effects for participants in the UK and Japan in line with predictions in H11. This suggests that cultural backgrounds likely shaped how people interpreted cyber readiness and how they judged a company's reaction and assigned blame following a cyber attack.

4. General discussion and conclusion

Three experiments were conducted online in both the UK (Experiment 1 and 2a) and Japan (Experiment 2b) participants to investigate the effects of cyber readiness and cyber response on trust and blame assignment in AVs following a cyber attack. With the gradual impending implementation of AVs into transport networks within many parts of the world – including the UK and Japan, these findings underscore the critical role that not only cyber readiness and cyber response could play in shaping human trust in AVs, but also the potential role of user cyber literacy - as determined qualitatively in Experiment 1 and examined in Experiments 2a and 2b (via design modifications).

Cyber response seemingly plays a greater role in – at least to some extent – preventing the diminishment of trust. However, cyber readiness (especially when defined – as in Experiments 2a and 2b) can influence the development of initial trust both in an AV itself as well as the company responsible for the AV. The interactions which did not reach significance indicated cross-cultural agreement regarding the importance of cyber readiness. Both UK and Japanese participants similarly interpret cyber readiness information as a determining factor in initial trust in an AV. This aligns with broader literature suggesting universal or cross-cultural consensus in the foundational criteria influencing trust in automated systems (e.g., Hoff & Bashir, 2015; Lee & See, 2004). However, the extent to which initial trust has the potential to continue to affect trust when something goes wrong - even as the surface-level cues (cyber readiness rating) are replaced with actual experiences where nothing goes wrong – can be questioned due to the mixed findings across the three experiments. In practice, this could mean that cyber readiness has the potential to influence early-stage trust in AVs for some populations in some countries like the UK and Japan, but whether such initial trust later impact trust when something goes wrong e.g. a cyber attack may not be universal. These societal and cultural dimensions – based on the findings of the current paper - could have tangible implications for how AV cyber security strategies are designed, communicated, and received across different populations both within and between countries.

Furthermore, it appears that trust in AVs is not only diminished by a single cyber attack and inadequate preparation, but also by poor organisational responses to such incidents. In all experiments, companies reported to have responded negatively were trusted less, while positive responses helped to prevent further loss of trust. This finding aligns with prior research (although here in an AV context) emphasising that the content and perceived sincerity of an organisation's response – particularly transparent explanations and genuine apologies - play a critical role in trust repair (Gillespie & Dietz, 2009; Kim et al., 2009). In Japan, the AV company was blamed mainly based on their response whereas in the UK, the level of readiness also impacted blame. An inadequate response to a cyber attack could therefore lead to an erosion of trust in a particular AV, the company responsible for it and greater blame assignment could arise. Debatably, trust in the wider AV ecosystem could also be affected as mixed findings from this study suggest that a cyber attack on one AV might have far-reaching implications on trust in and likelihood of using other AVs.

Preparing for and responding appropriately to the eventuality of a cyber attack on AV and/or its connected infrastructure in a way that resonates with a vast array of future users has therefore been identified as important contributing factors that could mitigate the negative impacts such a cyber attack could have on trust. There are many reasons why automotive companies should be ensuring that mature cyber security practices (including readiness and response activities) are built into AV technologies - to meet regulatory compliance activities, mitigate against financial penalties (fines/downtime costs), reduce reputational damage, prevent the loss of information, as well as mitigating against other consequences that could arise from an cyber attack. However, it appears that there is another important reason for companies to exhibit mature cyber security practices – it could impact end-user trust.

Findings indicate that after a cyber attack, trust in an AV and the company responsible for an AV sharply declines. The level of cyber readiness has the potential to impact initial trust in an AV before a cyber attack occurs - when more mature practices were demonstrated trust in an AV and the company responsible for the AV was higher. Following a cyber attack, the extent to which trust declined depended on the type of cyber response: positive responses led to a smaller reduction in trust than negative responses. Defining the term cyber increased its desirability, while the rankings of other well-understood features remained largely unchanged. The pattern of results was largely similar between the UK and Japan. A failure therefore to exhibit mature cyber security practices could potentially jeopardize trust in AVs which could in turn prevent AV technologies from gaining traction and might even impact the uptake, usage and wide-spread adoption of AVs.

5. Limitations and future directions

Despite offering many novel experimental insights into trust and blame on in AVs and manufactures following a cyber attack, there are limitations concerning inherent challenges of examining trust and cyber security in the context of emerging technologies such as AVs and the constraints of the chosen justified methodologies.

Uncertainty of Autonomous Vehicles.

One limitation relates to the hypothetical nature of AVs that do not require a human to takeover controls and drive them at any point (i.e. SAE 2021, Level 4+) which whilst becoming more likely, may never become fully operational or may only exist in limited operating domains (as is the case in some parts of some countries now). It is highly likely that the mass consumer market would not be the first to adopt the technology and currently this results in collecting data from participants who have never directly experienced an AV. This poses a challenge as the basis of trust may shift with real-world exposure as experience and familiarity – two key factors in trusting automated systems - is gained (Hoff & Bashir, 2015). In the absence of experiencing an AV(s) prior to taking part in an experiment (as is likely the case here), participants will lack a reference point for what constitutes as normal AV behaviour and must instead rely on judgements when interpreting the AVs actions. Prior research indicates that hypothetical judgements of behaviour frequently diverge from actual responses, particularly when emotions, time pressure, or social influences are involved (Ajzen, 1991).

The challenge is amplified in the context of cyber security as participants may struggle to conceptualise what a cyber attack on an AV might entail, how it unfolds, and/or how the vehicle should appropriately respond. As a result, trust may be based more on abstract notions of AVs and cyber. Nevertheless, the findings are important – e.g. in terms of potential early adopters of AVs (those with little to no experience of them) and also for companies designing, building and deploying them – who must act to ensure that they are robust as they can possibly be with human users in mind, including in the event of something going wrong - such as a cyber attack, a collision, and so on.

Methodological.

All experiments were conducted online and whilst this increased sample reach (including across countries) and efficiency, it may have inherently limited the degree of experimental control. Specifically, it was not feasible to verify participants' level of engagement, detect possible environmental distractions, or assess the reliability of self-reported trust metrics which could all contribute confounds. Replicating the experiment in-person and within a driving simulator would be beneficial. Participants would have experience of being driven autonomously (albeit in a simulator) and it would also be possible to gather physiological measures that arguably relate to trust (such as eye-tracking data - fixations, saccades, and pupil dilation) (Körber, 2019; Zhang et al., 2022). In-person testing would also provide tighter control over potential confounds such as background distractions (noting specific instructions to take part in a distraction free environment). Eye tracking data for example, could be useful to e.g. examine where participants focus (e.g. at perceptual, attentional or deeper processing levels), and the time spent looking or dwelling on/at various elements e.g. the dashboard both before and during the cyber attack (indicated by the malfunctioning features), and the road / other scenery. Other metrics such as fixation duration, saccadic movements, and pupil dilation can offer nuanced insights into attentional focus and cognitive load, which are closely associated with trust and situational awareness (Fridman et al., 2019), the latter not measured in this study.

It should also be noted that some terminology used to manipulate ‘Cyber response’ (e.g., ‘isolated the incident’ or ‘business continuity plan’) may have been more technical/ sector specific than intended at the outset, potentially limiting participants' comprehension of the response. This is particularly salient given that the term ‘cyber’ itself was not well understood until it was defined, underscoring the importance of accounting for cyber awareness differences as well as the use of accessible language in future research to ensure consistent interpretation of the company's actions.

Finally, differences in trust was noted between one type of event (overtake/non-overtake of a bus). Extending the scenario to a wider range of situations (events) is of importance to examine whether and to what extent the findings are generalisable to different driving contexts (Gold et al., 2015). Nevertheless, this study provides a human-centred perspective on trust in AVs following a cyber attack, offering findings that may help inform the design and focus of highly anticipated future on-road research studies.

Ethical Statement

All experiments presented in this paper have been reviewed and approved by The School of Psychology Research Ethics Committee (SREC), Cardiff University and have also undergone and passed a risk assessment.

Funding

This research was funded by an ESRC-JST project: Rule of Law in the Age of AI: Principles of Distributive Liability for Multi-Agent Societies - ES/T007079/1 (the last author was UK PI; the second to last author was the Japan PI).

CRediT authorship contribution statement

Victoria Marcinkiewicz: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Qiyuan Zhang:** Resources, Methodology, Data curation, Conceptualization. **Minoru Asada:** Writing – review & editing, Funding acquisition, Conceptualization. **Yoshiyuki Ueda:** Writing – review & editing, Data curation, Conceptualization. **Hirofumi Katsuno:** Writing – review & editing, Conceptualization. **Tatsuhiko Inatani:** Writing – review & editing, Supervision, Funding acquisition, Data curation. **Phillip L. Morgan:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

A warm thank you goes to the wider project team for their valuable contributions in translating materials and liaising with data collection agencies.

Appendix A. Knowledge check question.

What happened at the end of the video?

- The dashboard features of Vehicle X disappeared - there was no warning sound
- A warning sounded and the dashboard of Vehicle X began flashing and moving unusually
- The dashboard of Vehicle X began flashing and moving unusually - there was no warning sound
- A warning sounded and the dashboard features of Vehicle X disappeared
- Prefer not to say

Appendix B. Information provided to a participant about the cyber response

Positive cyber response	Negative cyber response
Having learnt about the cyber attack, the *SD*C company behind Vehicle X:	Having learnt about the cyber attack, the *SD*C company behind Vehicle X:
- Enacted their business continuity plan which included detailed cyber incident response and disaster recovery processes; - Isolated the incident to minimise downtime and disruption; - Reported the incident to the relevant authorities in a timely manner; - Promptly informed all affected customers and offered compensation for the incident.	-Were unable to enact their business continuity plan as, along with incident response and disaster recovery processes, it did not exist; -Failed to isolate the incident which would have minimised downtime and disruption; -Failed to report the incident to the relevant authorities in a timely manner; -Did not promptly inform all affected customers and offered no compensation for the incident.

Appendix C. Ratings of potential factors that could influence beliefs, trust and likelihood of using AVs

Statement	Mean	Mode	*SD*	Range
I am more likely to use a self-driving car with higher star-ratings.	87.59	100	14.927	0–76
Before using a self-driving car, I would find out as much information as I could about its specification*.	89.93	100	16.690	0–100
Whether I would use a self-driving car would be influenced by what other people are saying about them.	60.83	100	27.238	0–100
A self-driving car must achieve a 5-star-rating across all its specifications.	72.07	100	28.567	0–100
Whether I would use a self-driving car would be influenced by reviews that I read.	71.09	100	24.895	0–100
Some specifications of a self-driving car are more important to me than others.	78.33	100	20.786	0–100
Whether I would use a self-driving car would be influenced by my own experience with them.	75.71	100	22.432	0–100
I believe that products and services with higher ratings are better quality than those with lower ratings*.	80.90	100	18.410	0–88
I believe that products and services with higher ratings can be trusted more than those with lower ratings*.	79.30	100	18.186	0–83
I would be willing to pay more for a product or service with higher ratings*.	77.88	100	18.902	0–100

* Statements used in Experiment 2a (UK) and Experiment 2b (Japan). All other statements were disregarded.

Appendix D. Four Statements Relating to Blame Assignment.

Statement 1	The self-driving car company is most to blame for the occurrence of the cyber attack.
Statement 2	Based on the star-ratings, the self-driving car company could have been better prepared for the cyber attack on Vehicle X.
Statement 3	The self-driving car company responded appropriately to the cyber attack.
Statement 4	The cyber attack on Vehicle X was inevitable.

Appendix E. Alphabetical list of ten desirable AV features and their meaning

Comfort

This rating is about the self-driving cars level of comfort in terms of ride quality, suspension, level of noise when motoring, adjustability of seating positions, seat comfiness, leg room and visibility.

Cyber

This rating is about the self-driving cars capability to identify and protect itself against cyber related incidents which if undetected could lead to user information being stolen/disclosed, and/or could cause disruption/harm.

Design

This rating is about the external shape/build of the self-driving car, its internal aesthetics as well as whether the design of each feature is fit for purpose and/or easy to use.

Environmental Friendliness

This rating is about the self-driving cars environmental friendliness and the impact of its emissions released into the atmosphere (if not electric/fuel cell), including Carbon Dioxide, Nitrogen Oxide and particulates output.

Practicality

This rating is about how practical the self-driving car is in terms of the amount and versatility of internal space and compartments, and it also considers its external dimensions and features.

Performance

This rating is about the self-driving cars performance in terms of its engine and gearbox, speed, braking and cornering capabilities, smoothness, acceleration as well as its adaptability to different weather conditions.

Reliability

This rating is about any known reliability issues, recalls or frequently reported faults associated with the self-driving car based on a variety of sources including trade contacts, owner reviews and reports.

Road Worthiness

This rating is about the condition of the self-driving car in terms of its components working correctly, wear and tear and/or involvement in prior accidents which could impact its road worthiness.

Running Costs

This rating is about how much the self-driving car is expected to cost per year in terms of road tax, warranty and servicing, and it also takes into account fuel economy.

Safety

This rating is about the self-driving cars ability to avoid accidents and/or handle a crash based on an evaluation of various systems and features which could impact safety should they malfunction.

Data availability

The data and materials supporting the findings of this study are available on the Open Science Framework at: https://osf.io/5984s/overview?view_only=c326c06f77f74a3d98fc90c1d564a0a2.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Anania, E. C., Rice, S., Walters, N. W., Pierce, M., Winter, S. R., & Milner, M. N. (2018). The effects of positive and negative information on consumers' willingness to ride in a driverless vehicle. *Transport Policy, 72*, 129–137. https://doi.org/10.1016/j.tranpol.2018.04.002

Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., … Rahwan, I. (2020). Drivers are blamed more than their automated cars when both make mistakes. *Nature Human Behaviour, 4*(2), 134–143. https://doi.org/10.1038/s41562-019-0762-8

Bansal, P., & Kockelman, K. M. (2017). Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies. *Transp. Res. Part A: Policy and Practice, 95*, 49–63. https://doi.org/10.1016/j.tra.2016.10.013

Bennett, J. M., Challinor, K. L., Modesto, O., & Prabhakharan, P. (2020). Attribution of blame of crash causation across varying levels of vehicle automation. *Safety Science, 132*, Article 104968. https://doi.org/10.1016/J.SSCI.2020.104968

Bentley, J., & Ma, L. (2020). Testing perceptions of organizational apologies after a data breach crisis. Public Relations Review 46(5). doi:https://doi.org/10.1016/j.pubrev.2020.101975.

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science, 352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Cabinet Office. (2022). Government cyber security strategy: 2022 to 2030 – Building a cyber-resilient public sector. https://www.gov.uk/government/publications/government-cyber-security-strategy-2022-to-2030.

Centre for Connected and Autonomous Vehicles [CCAV]. (2024). *Automated Vehicles Act summary. GOV.UK. Available from: Centre for Connected and Autonomous Vehicles [CCAV]. (2024). Automated Vehicles Act summary. GOV.UK. [Accessed 14th July 2025]*.

Chen, Q., Romanowich, P., Castillo, J., Roy, K. C., Chavez, G., & Xu, S. (2021). ExHPD: exploiting human, physical, and driving behaviors to detect vehicle cyber attacks. *IEEE Internet of Things Journal, 8*(18), 14355–14371. https://doi.org/10.1109/JIOT.2021.3069951

Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human Computer Interaction, 31*(10), 692–702. https://doi.org/10.1080/10447318.2015.1070549

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Coombs, W. T. (2007). Protecting organization reputations during a crisis: the development and application of situational crisis communication theory. *Corporate Reputation Review, 10*(3), 163–176. https://doi.org/10.1057/palgrave.crr.1550049

VicOne Corporation. (2025). *VicOne awarded cyber risk analysis project for the autonomous driving era by the general insurance rating Organization of Japan. Business wire*. Retrieved from https://vicone.com/company/press-releases/vicone-awarded-cyber-risk-analysis-project-by-the-general-insurance-rating-organization-of-japan (accessed 3rd April 2025)..

DfT, CPNI and CCAV. (2017). *The key principles of vehicle cyber security for connected and automated vehicles. UK Government*. Retrieved from https://www.gov.uk/government/publications/principles-of-cyber-security-for-connected-and-automated-vehicles/the-key-principles-of-vehicle-cyber-security-for-connected-and-automated-vehicles (Accessed 27th June 2025).

Fridman, L., Reimer, B., Mehler, B., & Freeman, W. T. (2019). Cognitive load estimation in the wild. *Proce. National Academy of Sci., 116*(28), 13774–13779. https://doi.org/10.1145/3173574.3174226

Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review, 34*(1), 127–145. https://doi.org/10.5465/AMR.2009.35713319

Gold, C., Körber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in automation - before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing, 3*, 3025–3032. https://doi.org/10.1016/j.promfg.2015.07.847

Gorine, A., & Khan, O. (2024). Trust after a simulated cyber attack in autonomous vehicles. *Human Factors in Cybersecurity, 12*(1), 33–51. https://doi.org/10.11648/j.ajcst.20240704.11

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors, 53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Hoff, K. A., & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Sage Publications.

Holthausen, B., Wintersberger, P., Walker, B., & Riener, A. (2020). A situational trust scale for automated driving (STS-AD): development and initial validation. *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 1–10. https://doi.org/10.1145/3409120.3410637

Hong, J. W. (2020). Why is artificial intelligence blamed more? Analysis of faulting artificial intelligence for self-driving car accidents in experimental settings. *International Journal of Human Computer Interaction, 36*(18), 1768–1774. https://doi.org/10.1080/10447318.2020.1785693

Hong, J. W., Cruz, I., & Williams, D. (2021a). AI, you can drive my car: how we evaluate human drivers vs. self-driving cars. *Computers in Human Behavior, 125*, Article 106944. https://doi.org/10.1016/J.CHB.2021.106944

International Organization for Standardization (ISOa)/SAE International. (2021). ISO/SAE 21434:2021 – Road vehicles – Cybersecurity engineering. https://www.iso.org/standard/70918.html.

International Organization for Standardization (ISOb). (2023). ISO 24089:2023 – Road vehicles – Software update engineering. https://www.iso.org/standard/77796.html.

Kageyama, Y. (2025). Nissan shows driverless vehicle in Japan, plans autonomous service by 2027, eyes level 4 autonomy by 2030. *AP News*. March 10 https://apnews.com/article/5c12444c3931d1c7a0280789d2b0cba9.

Khan, S. K., Shiwakoti, N., Stasinopoulos, P., et al. (2023). A multinational empirical study of perceived cyber barriers to automated vehicles deployment. *Scientific Reports, 13*, 1842. https://doi.org/10.1038/s41598-023-29018-9

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2009). Trust and satisfaction, two stepping stones for successful E-commerce relationships: a longitudinal exploration. *Information Systems Research, 20*, 237–257. https://doi.org/10.1287/isre.1080.0188

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence-versus integrity-based trust violations. *The Journal of Applied Psychology, 89*, 104–118. https://doi.org/10.1037/0021-9010.89.1.104

Körber, M. (2019). Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.)*, 823. Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). IEA 2018. Advances in Intelligent Systems and Computing*. Springer, Cham. https://doi.org/10.1007/978-3-319-96074-6_2.

Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. *Transp. Res. Part F: Traffic Psychology and Behaviour, 39*, 19–32. https://doi.org/10.1016/j.trf.2016.03.002

Kritzinger, E., & Von Solms, S. H. (2010). Cyber security for home users: a new way of protection through awareness enforcement. *Computers & Security, 29*(8), 840–847. https://doi.org/10.1016/j.cose.2010.08.001

Kyriakidis, M., Happee, R., & de Winter, J. C. F. (2015). Public opinion on automated driving: results of an international questionnaire among 5,000 respondents. *Transp. Res. Part F: Traffic Psychology and Behaviour, 32*, 127–140. https://doi.org/10.1016/j.trf.2015.04.014

Lancefield, N. (2025). *Tesla successfully tests fully self-driving car on UK roads. The Independent*. Retrieved from. https://www.independent.co.uk/tech/tesla-tests-selfdriving-car-london-swindon-musk-b2796060.html.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35*(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lim, C., Prendez, D., Boyle, L. N., & Rajivan, P. (2024). The impact of cybersecurity attacks on human Trust in Autonomous Vehicle Operations. *Human Factors: J. Human Factors and Ergonomics Soc., 67*(5), 485–502. https://doi.org/10.1177/00187208241283321

Lin, P. S., Wang, Z., & Guo, R. (2016). *Impact of connected vehicles and autonomous vehicles on future transportation. In bridging the east and west: Theories and practices of transportation in the Asia Pacific proceedings of the 11th Asia Pacific transportation development conference and the 29th ICTPA annual conference* (pp. 46–53). Reston: American Society of Civil Engineers. https://doi.org/10.1061/9780784479810.006

Liu, P., & Du, Y. (2021). Blame attribution asymmetry in human–automation cooperation. *Risk Analysis*. https://doi.org/10.1111/RISA.13674

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science, 8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Malle, B. F., Knowles, M., Scheutz, M., & Yildrim, F. (2014). Moral competence in social robots [conference presentation]. *IEEE International Symposium on Ethics in Engineering, Science, and Technology, Chicago, IL, USA.*. https://doi.org/10.1109/ETHICS.2014.6893446

Marcinkiewicz, V., Zhang, Q., & Morgan, P. L. (2023). The effects of cyber readiness and response on human trust in self-driving cars. In N. Sarkar, et al. (Eds.)*, 91. Human Factors in Cybersecurity* (pp. 50–60). AHFE Open Access. https://doi.org/10.54941/ahfe1003719.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organisational trust. *Academy of Management Review, 20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: an integrative typology. *Information Systems Research, 13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies, 27*(5–6), 527–539. https://doi.org/10.1016/S0020-7373(87)80013-5

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37*(11), 1905–1922. https://doi.org/10.1080/00140139408964957

Muzatko, S., & Bansal, G. (2018). Timing of data breach announcement and e-commerce trust. In *13th Midwest Association for Information Systems (MWAIS) Proceedings*. https://aisel.aisnet.org/mwais2018/7.

Olaverri-Monreal, C. (2020). Promoting trust in self-driving vehicles. *Nature Electronics, 3*(6), 292–294. https://doi.org/10.1038/s41928-020-0434-8

Parkers.. (2025). *Parkers trusted car reviews, valuations, and advice*. Bauer Media Group. Retrieved August 19, 2025, from https://www.parkers.co.uk/.

Payre, W., Perelló-March, J., & Birrell, S. (2023). Under pressure: effect of a ransomware and a screen failure on trust and driving performance in an automated car simulation. *Frontiers in Psychology, 14*(1), Article 1078723. https://doi.org/10.3389/fpsyg.2023.1078723

Payre, W., Perelló-March, J., Sabaliauskaite, G., Jadidbonab, H., Shaikh, S., & Birrell, S. (2022). How system failures and ransomwares affect drivers' trust and attitudes in an automated car? A simulator study. In T. Ahram, & R. Taiar (Eds.)*, 68. Human Interaction & Emerging Technologies (IHIET 2022): Artificial Intelligence & Future Applications. AHFE (2022) International Conference. AHFE Open Access*. USA: AHFE International. https://doi.org/10.54941/ahfe1002764.

Pettigrew, S., Fritschi, L., & Norman, R. (2018). The potential implications of autonomous vehicles in and around the workplace. *International Journal of Environmental Research and Public Health, 15*(9), 1916. https://doi.org/10.3390/ijerph15091876

Pham, M., & Xiong, K. (2021). A survey on security attacks and defense techniques for connected and autonomous vehicles. *Computers & Security, 109*(1), 1–29. https://doi.org/10.1016/j.cose.2021.102269

Pöllänen, E., Read, G. J. M., Lane, B. R., Thompson, J., & Salmon, P. M. (2020). Who is to blame for crashes involving autonomous vehicles? Exploring blame attribution across the road transport system. *Ergonomics, 63*(5), 525–537. https://doi.org/10.1080/00140139.2020.1744064

SAE International. (2021). Taxonomy and Definitions for Terms Related to Driving Automation Systems for on-Road Motor Vehicles (J3016_202104). https://www.sae.org/standards/content/j3016_202104/.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A Meta-analysis of factors influencing the development of Trust in Automation: Implications for understanding autonomy in future systems. *Human Factors, 58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Schoettle, B., & Sivak, M. (2014). A survey of public opinion about connected vehicles in the U.S., the U.K., and Australia. In *In 2014 International Conference on Connected Vehicles and Expo (ICCVE)* (pp. 687–692). IEEE. https://doi.org/10.1109/ICCVE.2014.7297637.

Shaver, K. G. (1985). *The Attribution of Blame: Causality, Responsibility, and Blameworthiness*. Springer. https://doi.org/10.1007/978-1-4612-5094-4

Strzelecki, A., & Rizun, M. (2022). Consumers' change in trust and security after a personal data breach in online shopping. *Sustainability, 14*(10), Article 5866. https://doi.org/10.3390/su14105866

Sun, X., Yu, F. R., & Zhang, P. (2022). A survey on cyber-security of connected and autonomous vehicles (CAVs). *IEEE Transactions on Intelligent Transportation Systems, 23*(7), 6240–6259. https://doi.org/10.1109/TITS.2021.3085297

Sung, Y.-T., & Wu, J.-S. (2018). The visual analogue scale for rating, ranking and paired-comparison (VAS-RRP): A new technique for psychological measurement. *Behavior Research Methods, 50*(4), 1694–1715. https://doi.org/10.3758/s13428-018-1041-8

Tomlinson, E. C., & Mayer, R. C. (2009). The role of causal attributions in trust repair. *Academy of Management Review, 34*(1), 85–104. https://doi.org/10.5465/AMR.2009.35713291

United Nations Economic Commission for Europe (UNECEa). (2021). *UN Regulation No. 155: Cybersecurity and cyber security management system.* Available from https://unece.org/transport/documents/2021/03/standards/un-regulation-no-155-cyber-security-and-cyber-security (Accessed 3rd May 2025).

United Nations Economic Commission for Europe (UNECEb). (2021). *UN Regulation No. 156: Software update and software update management system. UNECE.* Retrieved July 1, 2023, from https://unece.org/transport/documents/2021/03/standards/un-regulation-no-156-software-update-and-software-update (Accessed 3rd May 2025).

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology. Applied, 22*(3), 331–349. https://doi.org/10.1037/xap0000092

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From automation to autonomy: the importance of trust repair in human–machine interaction. *Ergonomics, 61*(10), 1409–1427. https://doi.org/10.1080/00140139.2018.1457725

Wallbridge, C., Zhang, Q., Marcinkiewicz, V., Bowen, L., Kozlowski, T., Jones, D., & Morgan, P. (2024). "Warning!" benefits and pitfalls of Anthropomorphising autonomous vehicle informational assistants in the case of an accident. *Multimodal Technologies Interaction, 8*, 110. https://doi.org/10.3390/mti8120110

Wang, M., Parker, J., Zhang, F., & Roberts, S. C. (2024). A simulator study assessing the effectiveness of training and warning systems on drivers' response performance to vehicle cyberattacks. *Accident Analysis & Prevention, 203*, Article 107644. https://doi.org/10.1016/j.aap.2024.107644

Wu, H., & Leung, S.-O. (2017). Can Likert scales be treated as interval scales? A simulation study. *Journal of Social Service Research, 43*(4), 527–532. https://doi.org/10.1080/01488376.2017.1329775

Yamagishi, T. (2001). Trust as a form of social intelligence. In K. S. Cook (Ed.), *Trust in society* (pp. 121–147). Russell Sage Foundation.

Yang, Y., & Kim, M.-Y. (2025). Promoting sustainable transportation: How people trust and accept autonomous vehicles - focusing on the different levels of collaboration between human drivers and artificial intelligence - an empirical study with partial least squares structural equation modeling and multi-group analysis. *Sustainability, 17*(1), 125. https://doi.org/10.3390/su17010125

Zhang, F., Wang, M., Parker, J., & Roberts, S. C. (2023). The effect of driving style on responses to unexpected vehicle cyberattacks. *Safety, 9*(1), 5. https://doi.org/10.3390/safety9010005

Zhang, Q., Wallbridge, C. D., Jones, D. M., & Morgan, P. L. (2024). Public perception of autonomous vehicle capability determines judgment of blame and trust in road traffic accidents. *Transportation Research Part A: Policy and Practice, 179*, Article 103887. https://doi.org/10.1016/j.tra.2023.103887

Zhang, Q., Wallbridge, D. C., Morgan, P., & Jones, M. D. (2022). Using simulation software-generated animations to investigate. *Procedia Computer Science, 207*, 3516–3525. https://doi.org/10.1016/j.procs.2022.09.410