Engineering 33 (2024) 121–132

Contents lists available at ScienceDirect

Engineering

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

Research
Safety for Intelligent and Connected Vehicles—Article

Not in Control, but Liable? Attributing Human Responsibility for Fully Automated Vehicle Accidents

Siming Zhai [a,#], Lin Wang [b,#], Peng Liu [a,*]

[a] Center for Psychological Sciences, Zhejiang University, Hangzhou 310063, China
[b] Department of Library and Information Science, Incheon National University, Incheon 22012, Republic of Korea

ARTICLE INFO

Article history:
Received 8 September 2022
Revised 1 May 2023
Accepted 18 October 2023
Available online 16 December 2023

Keywords:
Fully automated vehicle accidents
Responsibility attribution
Controllability
Foreseeability

ABSTRACT

Human agency has become increasingly limited in complex systems with increasingly automated decision-making capabilities. For instance, human occupants are passengers and do not have direct vehicle control in fully automated cars (i.e., driverless cars). An interesting question is whether users are responsible for the accidents of these cars. Normative ethical and legal analyses frequently argue that individuals should not bear responsibility for harm beyond their control. Here, we consider human judgment of responsibility for accidents involving fully automated cars through three studies with seven experiments ($N = 2668$). We compared the responsibility attributed to the occupants in three conditions: an owner in his private fully automated car, a passenger in a driverless robotaxi, and a passenger in a conventional taxi, where none of these three occupants have direct vehicle control over the involved vehicles that cause identical pedestrian injury. In contrast to normative analyses, we show that the occupants of driverless cars (private cars and robotaxis) are attributed more responsibility than conventional taxi passengers. This dilemma is robust across different contexts (e.g., participants from China vs the Republic of Korea, participants with first- vs third-person perspectives, and occupant presence vs absence). Furthermore, we observe that this is not due to the perception that these occupants have greater control over driving but because they are more expected to foresee the potential consequences of using driverless cars. Our findings suggest that when driverless vehicles (private cars and taxis) cause harm, their users may face more social pressure, which public discourse and legal regulations should manage appropriately.

© 2023 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Machines (e.g., automation and robots), usually powered by artificial intelligence (AI), operate alongside humans to augment human capacities or even replace them to reduce human vulnerabilities, such as fatigue and attention lapses, in various safety-critical settings (e.g., road vehicle driving, aircraft piloting, medical diagnosis, and surgery) [1–4]. As our research focus, automated vehicles (AVs) are one such important application. Their widespread adoption can facilitate greater safety, affordability, accessibility, and sustainability [5,6]. Because human errors are behind most traffic crashes [7,8], researchers claim that removing error-prone human drivers from the causal chain using automated machine drivers can significantly reduce car crashes, injuries, and deaths [9].

The roles (or levels of agency) of human drivers and machine drivers, which are essential for understanding who or what is responsible when an error occurs, are described in a six-level driving automation taxonomy [10]: no automation (level 0; hereafter, L0), driver assistance (level 1; hereafter, L1), partial automation (level 2; hereafter, L2), conditional automation (level 3; hereafter, L3; a human shares vehicle control with a machine and intervenes when required), high automation (level 4; hereafter, L4; human intervention is not required), and full automation (level 5; hereafter, L5; human intervention does not exist), as shown in Fig. 1 [10]. Similar six-level classification systems exist in other sectors, such as surgical robotics [4,11]. Unless explicitly mentioned, we refer to the driver/operator or user (occupant, passenger, or rider) as "human." In fully AVs, the human acts as a mere passenger, and the machine is the sole driver (i.e., driverless vehicles with no

[*] Corresponding author.
E-mail address: pengliu86@zju.edu.cn (P. Liu).
[#] These authors contributed equally to this work.

https://doi.org/10.1016/j.eng.2023.10.008
2095-8099/© 2023 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

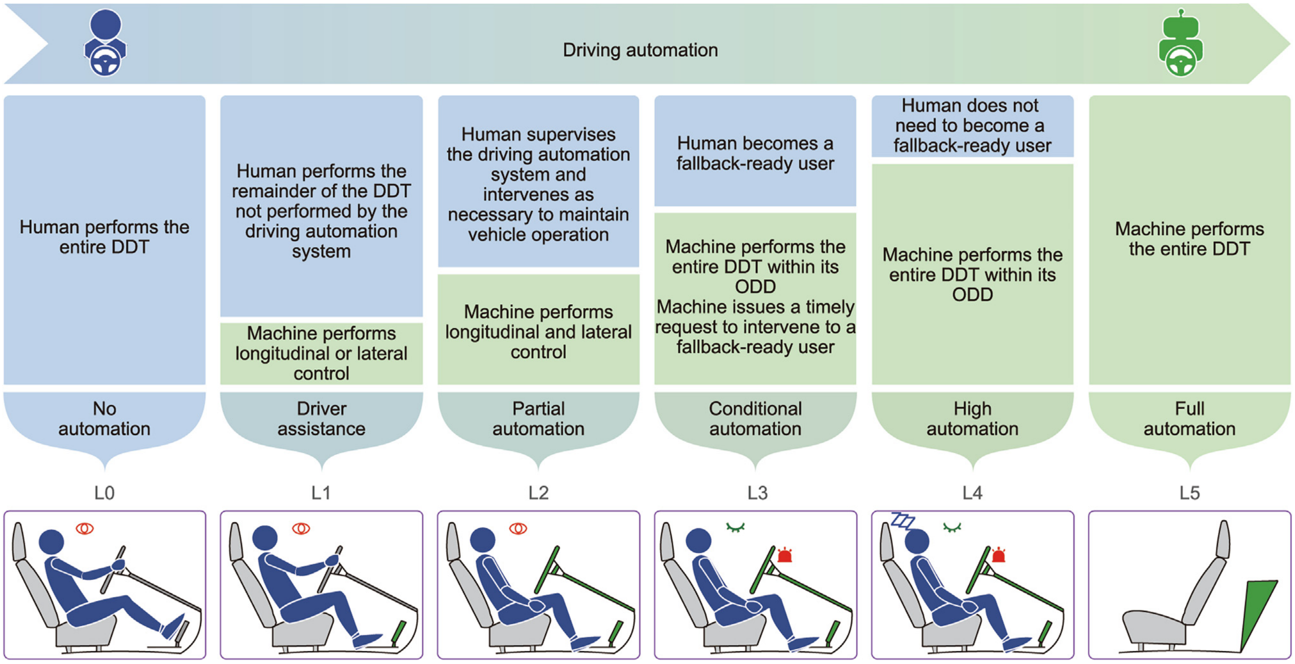
Driving automation
Human performs the entire DDT
Human performs the remainder of the DDT not performed by the driving automation system
Machine performs longitudinal or lateral control
Human supervises the driving automation system and intervenes as necessary to maintain vehicle operation
Machine performs longitudinal and lateral control
Human becomes a fallback-ready user
Machine performs the entire DDT within its ODD
Machine issues a timely request to intervene to a fallback-ready user
Human does not need to become a fallback-ready user
Machine performs the entire DDT within its ODD
Machine performs the entire DDT
No automation
Driver assistance
Partial automation
Conditional automation
High automation
Full automation
L0
L1
L2
L3
L4
L5

Fig. 1. Six levels of driving automation [10]. DDT: dynamic driving task; ODD: operational design domain.

steering wheel). Companies worldwide are developing and testing fully AVs and other AVs. For example, China, the United States, and other countries are piloting robotaxis (driverless taxis), a business model for full automation.

As with every groundbreaking technology, AV technology can introduce significant social, environmental, and economic benefits but also lead to ethical, legal, and social problems [12,13]. Responsibility attribution (particularly liability attribution) may be the most basic and important factor if an AV causes a collision. Traditionally, drivers have carried almost all moral and legal responsibilities. This driver-focused liability system would become obsolete in the AV era. Operational responsibility in AVs is transferred from the human driver to an automated driving system (ADS; i.e., a machine driver). This implies a transfer of retrospective responsibility for traffic crashes [14] and raises concerns about who should be responsible if AVs cause traffic crashes [15,16]. Vehicles are only part of a much larger sociotechnical system. For AVs to operate on public roads, the world must change [17], including the policymaking and legislation on responsibility attribution [18,19].

Traffic crashes involving different levels of automation are already occurring. In May 2016, a collision resulted in the death of a driver in his Tesla Autopilot car (L2) [20] because "neither Autopilot [an L2 system] nor the driver noticed the white side of the tractor trailer against a brightly lit sky, so the brake was not applied" [20]. In March 2018, an Uber developmental AV (L3) caused the death of a pedestrian walking across a street outside a crosswalk [21]. In this crash, the ADS inaccurately identified the pedestrian as a vehicle, unknown object, or bicyclist and thus delayed its braking. Simultaneously, the test driver did not notice the pedestrian until one second before the collision. Although the accidents were caused by the joint errors of the involved human and machine drivers, the courts cleared both companies of legal responsibility. The courts charged Uber's backup driver with negligent homicide in the first jury trial, who was the only person to face criminal charges [21]. These crashes raised concerns regarding responsibility attribution in AV accidents. A pressing concern is that humans (frequently the nearest operator) may be—accidentally or intentionally—treated as a "moral crumple zone" [22] or "legal sponge" [23] to bear the brunt of moral and legal responsibilities while having limited decision-making authority and control over faulty machines. Uncertainty regarding responsibility attribution negatively affects AV development and deployment [16,18,24] and slows the building of an insurance framework for AV crashes [14].

An interesting question is whether the users of fully AVs are subject to responsibility for their crashes. This question appears to be naïve but has non-trivial implications. L5 AV users are merely passengers [25] and perform limited strategic tasks, such as setting the destination of a trip. Current legal practice holds that a faultless passenger in a taxi, bus, or other transportation tool is not responsible for the occurrence of a crash. Following this logic, L5 AV users might expect an absolution of retrospective responsibility for any harm caused by these vehicles that drive themselves, which would be a motivation for their willingness to purchase or use L5 AVs. Normative legal and ethical analyses [26–30] argue that retrospective responsibility will likely shift from humans to manufacturers (Section 1.1) if humans do not present any illegal behaviors (e.g., hacking or other improper interventions). This shift appears to be what the industry expects. Volvo [31] and Audi [32] have publicly promised to take full responsibility for their vehicle-caused crashes when their driverless vehicles are commercially available. In contrast, society might still blame its users for currently unknown reasons (Section 1.2), which would cast a shadow over their future.

1.1. Responsibility attribution in normative analysis

Retrospective responsibility has at least three forms: causal, moral (i.e., blame), and legal (i.e., liability) [12,33,34]. They are distinct factors but have close interrelationships; for instance, causal responsibility attribution is an essential factor in blame judgment [35]. Controllability (or agency) refers to the degree to which an actor can volitionally alter a cause [36]. This lies at the center of ethical, philosophical, and legal discourses involving responsibility attribution for L5 AV accidents, which generally suggest not holding L5 AV users morally and legally responsible for the harm caused by their vehicles.

In ethics and philosophy, Aristotle [37], in *Nicomachean Ethics,* argued that exercising responsibility requires at least two conditions: the control condition (an actor must be in control of their actions) and the epistemic condition (they must know what they are doing and where). Controllability is associated with moral responsibility through the control principle: Individuals can only be morally responsible for the actions over which they have control [38–40]. This is a commonly used standard for assessing an actor's morally faulty behavior. In an ethical analysis, Hevelke and Nida-Rümelin [28] argued that blaming the rider of an L5 AV for the death of another caused by the AV when the rider did not have an actual chance to intervene is a form of defamation. They highlighted that this rider has no duty to be attentive to the road and traffic and no capacity to intervene when necessary to avoid traffic crashes. However, philosophers also note the "moral luck" phenomenon across specific cases: One can treat an actor as an object of moral judgment, even though a significant aspect of their actions depends on factors beyond their control [41,42].

The association between control and legal responsibility is also well-established [43]. The control doctrine is a legal principle in motor–vehicle accident laws. These laws are driver-centric because human drivers have complete control over their vehicles. Thus, the law considers that they bear full responsibility for crashes if there are no mitigating circumstances. Huddy [44] wrote in "The law of automobiles" that "liability for the operation of a motor vehicle is imposed on the person having 'control' of its movements." Legal scholars have expressed similar perspectives on AV accidents. As riders in L5 AVs are outside the decision-making and vehicle control loop and cannot perform any part of the dynamic driving task, certain legal scholars argue that their riders are unlikely to be liable for accidents [25,26,29]. Vladeck [27] described them as serving no different role than a "potted plant." Accordingly, Gurney [30] suggested treating manufacturers as drivers of L5 AVs; vehicle manufacturers must bear the ultimate responsibility for human drivers on current roads.

1.2. Responsibility attribution from the public opinion

As described in the "society-in-the-loop" framework for technology regulation, lawmakers should consider the values and opinions of various stakeholders affected by emerging technologies when changing regulations and legislation surrounding these technologies to make them transparent, fair, and accountable [45]. Crowdsourcing identifies and computes different stakeholders' values, choices, and opinions while solving the non-technical conundrums caused by emerging technologies. As an important crowdsourcing approach in citizen science, vignette-based experiments [18,46–53] have examined how society responds if AVs cause harm, primarily in blame attribution. Frequently, participants read about an AV crash, typically manipulated as being caused by the human driver's erroneous behaviors (e.g., being distracted), imagined themselves as an observer or victim in an AV crash involving different automation levels, and then attributed blame to multiple parties (e.g., the driver, vehicle, manufacturer, pedestrian, developer, and government).

Among these, two studies using rating measures (e.g., from no blame to a lot of blame) reported that a user [49] or passenger [18] in an L5 AV did not receive the lowest level of blame when the L5 AV caused an accident. This counter-normative finding was not deliberately discussed in these two studies but has received particular attention in other studies [47,53]. Bennett et al. [47] considered four types of vehicles (L0, L2, L3, and L4/L5). In their vignettes, the corresponding vehicle was moving on a straight street and hit a pedestrian crossing the street; the driver was on his phone. However, their vignettes did not mention whether the vehicle was faulty. From the victim's perspective, participants assigned blame to six stakeholders (driver, pedestrian, car, government, manufacturer, and programmer). Bennett et al. [47] observed that the proportion of participants who blamed the driver decreased and those who blamed the manufacturer increased as the automation level increased. However, approximately one-third of participants blamed the driver when the vehicle was fully automated. Regarding this finding, Bennett et al. [47] considered it as a form of "driver shaming." As Bennett et al. [47] still used the term "driver" in their L5 vignette, which did not match the in-vehicle human role in L5 AVs, humans in L5 AVs in Bennett et al.'s study may have still been blamed because of the "driver" role assigned to them. Aguiar et al. [53] observed that this contradicts the control doctrine (Section 1.1).

1.3. Research motivation

In contrast to normative analysis, recent public-opinion studies [18,47,49,52,53] have offered early evidence for a tendency in responsibility attribution: people still blamed the users of L5 AVs for crashes involving fully AVs when they did not practically have direct vehicle control. Previous studies may encounter a methodological critique: The obtained tendency may result from potential measurement errors in subjective measures [54]. For instance, some "inattentive" participants may have not selected the supposed "none" level in the responsibility measures in these surveys. However, if we cannot attribute this finding to the potential weaknesses of subjective measures, it reflects a deeper psychological aspect, which may indicate people's responses to fully AVs and their harm. More specifically, this may signal a social bias against future consumers of fully AVs, which has practical implications for the AV industry and legislation for AV crashes (e.g., negatively impacting consumers' incentives to use L5 AVs). It also has theoretical implications as it might indicate an asymmetry between control and responsibility, which researchers rarely examine in social psychology research related to responsibility attribution for the harm caused by fully autonomous machines. The question of responsibility for machine actions has become increasingly important. Our study also contributes to the increasing literature on the attribution of responsibility to autonomous machines [2,18,55,56].

Overall, current empirical evidence for the tendency in responsibility attribution for L5 AV crashes is tentative (owing to potential measurement errors in subjective measures). Scholars have yet to examine the boundary conditions and underlying mechanisms. In this study, we systematically investigated and determined whether—why—people require L5 AV users to be held responsible for the harm caused by their L5 AVs.

2. Theory and research questions

In three sequential studies with seven experiments (six in China and one in the Republic of Korea), we focused on two research questions. First, are L5 AV users required to take (partial) responsibility for the harm caused by these vehicles? To answer this question, we investigated its persistence and robustness under different conditions (i.e., its potential boundary conditions) in study 1. The second question was, what are the possible psychological mechanisms? We answer this question through the lens of controllability (study 2) and foreseeability (study 3) in attribution theories in social psychology. Next, we explain the theoretical foundation of these questions and the arrangement of the experiments.

2.1. Are L5 AV users required to take responsibility for the harm caused by their driverless AVs?

We avoid the methodological critique in previous public-opinion studies by considering a conventional taxi passenger—assumed to bear no responsibility for any crashes under the current traffic law—as the referent and comparing the responsibilities attributed to them and L5 AV users. We consider the responsibility attributed to an owner in a private L5 car and a passenger in an L5 driverless taxi (i.e., robotaxi, a driverless service piloted worldwide). If we found that a passenger riding in a robotaxi was also attributed more responsibility than when riding in a conventional taxi when the two different taxes led to the same crash, we offer stronger evidence for the responsibility attribution pattern related to L5 AV users.

Regarding the tendency to hold a user of an L5 AV responsible when a driverless vehicle causes harm, we manipulated different experimental conditions to guarantee its robustness. First, we specified the default conditions (as a reference). Usually, when discussing responsibility, legal scholars, ethicists, policymakers, or laypeople assess it from a third-person perspective (i.e., the observer's perspective). Regarding the default metric for responsibility attribution, we mimicked the practical method of assigning responsibility to multiple agents by a court of law or liability determination authority; thus, we requested our participants to allocate a fixed amount of responsibility to the occupant and the other responsible agent after the involved vehicle caused a crash. Under default conditions, the occupants were in the vehicle. Therefore, the default condition manipulated in study 1a was that participants from the third-person perspective allocated a fixed amount of responsibility to the in-vehicle occupant and another responsible agent (responsibility allocation × third-person perspective × occupant presence). Accordingly, we designed several experimental conditions (Fig. 2), as below.

Previous research [18,49] adopted the responsibility rating metric (i.e., participants rated their blame assigned to the responsible agents independently). This metric difference may contribute to discrepancies in responsibility attribution findings among multiple studies [46]. Thus, study 1b considered the responsibility rating metric.

Perspective-taking is crucial for responsibility attribution [57]. Previous studies have examined the second-person perspective (i.e., the victim perspective) [47] and third-person perspective [49]. According to the defensive attribution theory [58] and self-preservation tendency [59], participants from the first-person perspective (i.e., imagining themselves as L5 AV users) would think they should not bear responsibility for harm caused by vehicles beyond their control. However, if the tendency of our interest still exists from a first-person perspective, it will receive even more substantial evidence. Therefore, study 1c considered the first-person perspective.

Researchers have wondered whether this tendency is because the user is in a car (as the person nearest to the crash) [47]. We are interested in whether the presence or absence of users in driverless cars influences the attribution of responsibility. Thus, as a response, study 1d manipulated that occupants are not in the vehicle when a crash occurs.

2.2. What are potential psychological mechanisms?

If the tendency of our interest is persistent and robust, what explains it? A simple account is that the owner of an L5 car should be responsible for the adverse outcomes caused by the car (although they are not operationally responsible for the dynamic driving task), which can be supported by the association between ownership and responsibility attribution (e.g., Refs. [60,61]). If someone purchases and owns a car, society accepts their legal ownership and the law protects their right to benefit from using the vehicle. Correspondingly, society considers them responsible for the negative outcomes caused by this car if mitigating conditions do not exist (e.g., a product defect in the car). Rahwan et al. [3] expressed a similar perspective: "If a dog bites someone, the dog's owner is held responsible" while discussing attribution of responsibility after an autonomous machine causes harm. Our work compares the responsibilities attributed to passengers in a robotaxi and a conventional taxi. As taxi passengers in the two riding conditions do not own the taxis, we can show whether this tendency occurs due to property ownership or other unknown factors. As we will reveal, taxi passengers are still attributed more responsibility in the robotaxi than in the conventional taxi, indicating that this tendency is not entirely due to ownership.

Attribution theories in social psychology have been developed to understand how people identify the causes of certain events and the necessary conditions for the attribution of responsibility. We use attribution theories [36,62–65] to establish our theory to explain the responsibility attribution tendency involving L5 AV users. Specifically, we use two accounts of perceived controllability and foreseeability, both of which play crucial roles in assigning blame for negative outcomes.

Attribution theories and empirical research in social psychology support the association between control and responsibility in the control principle in ethics and philosophy (Section 1.1). Weiner's attribution theory [36] suggests that people examine three dimensions while attempting to understand the cause of an event, one of

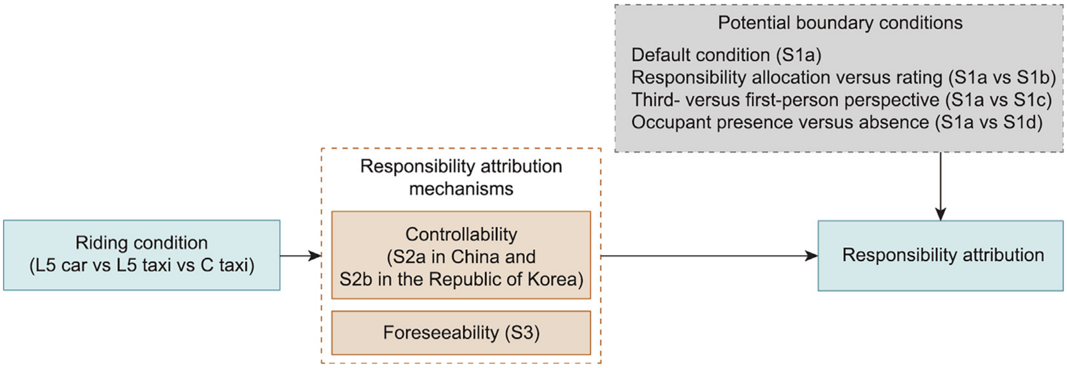
Potential boundary conditions
Default condition (S1a)
Responsibility allocation versus rating (S1a vs S1b)
Third- versus first-person perspective (S1a vs S1c)
Occupant presence versus absence (S1a vs S1d)
Responsibility attribution mechanisms
Controllability (S2a in China and S2b in the Republic of Korea)
Foreseeability (S3)
Riding condition (L5 car vs L5 taxi vs C taxi)
Responsibility attribution

Fig. 2. Theoretical framework for responsibility attribution involving L5 AV accidents. Study 1 (S1), study 2 (S2), and study 3 (S3) indicate the experimental studies testing the particular condition or mechanism. The default condition in study 1a was that participants from the third-person perspective allocated a fixed amount of responsibility to the in-vehicle occupant and the other responsible agent (responsibility allocation × third-person perspective × occupant presence). C taxi: conventional taxi.

which is controllability. According to Weiner's theory, causal control is necessary for determining responsibility. Alicke [63] proposed the culpable control model to assess the process by which blame occurs. One of its central assumptions is that ordinary people assess potentially blameworthy actions in terms of the actor's personal control over harmful events. Empirical studies support that when machine autonomy increases and human control decreases, people attribute less blame to the humans involved [47,49,52,66]. In the L5 context, people might be skeptical about whether users of L5 AVs (personally owned cars or robotaxis) cannot intervene. Thus, a potential explanation is that although the three occupants (an owner in his private L5 car, a passenger in a robotaxi, and a passenger in a conventional taxi) are common in that they literally and practically play no role in the vehicles' driving decisions and control, people perceive the two users in the two L5 conditions as having more control over driving and, thus, more responsibility for the driving outcome. Study 2 (two cross-national experiments) examined but failed to confirm this assumption, which led to follow-up study 3.

People's blame judgments are sensitive to the epistemic state of the actor [62–64]. Thus, another potential account is the occupants' foreseeability of the outcome. Lagnado and Channon [64] unpacked this construct into three varieties: subjective foreseeability (i.e., how likely an outcome is from the actor's perspective), objective foreseeability (i.e., what is in fact likely, irrespective of what the actor actually expects), and reasonable foreseeability (i.e., what is reasonable for the actor to expect, that is, what they should expect, given the information available to them). Lagnado and Channon [64] observed that subjective and objective foreseeability influence people's judgment of blame. However, few researchers have focused on empirically examining the influence of reasonable foreseeability. Shaver [62] argued that blame attribution should be affected by what the actor should have known about the outcome of their actions (reasonable foreseeability), not by what they actually knew (subjective foreseeability). Reasonable foreseeability is particularly important in legal cases, because it is a characteristic of the tort of negligence [67]. Although the lay and legal senses of reasonable foreseeability differ, we believe that the lay sense of reasonable foreseeability might affect a layperson's attribution of responsibility. Therefore, we examined reasonable foreseeability in study 3 and assume that people attribute more responsibility to L5 AV users because they believe that the users should be more aware of the consequences of using L5.

We conducted three studies sequentially (Fig. 2) to develop a deeper understanding of responsibility attribution for L5 AV accidents. The Ethics Review Board of the Center for Psychological Sciences, Zhejiang University, China, approved these studies. We ran StatCheck [68] and did not detect any inconsistencies between the different components of our statistics (e.g., t or F value, df, and p).

3. Study 1

Study 1 aimed to examine whether people attribute more responsibility to users of L5 AVs (a private L5 car and a driverless L5 taxi) than to passengers in a conventional taxi when their vehicles cause identical pedestrian injuries and whether this tendency is robust under different experimental conditions. Study 1a set the default condition: Participants from the third-person perspective allocated a fixed amount of responsibility to the in-vehicle occupants and another responsible party. Study 1b used the metric of responsibility rating (vs responsibility allocation in study 1a), study 1c considered the first-person perspective (vs the third-person perspective in study 1a), and study 1d considered the absence of occupants (vs the presence of occupants in study 1a).

3.1. Method

All experiments adopted a vignette-based design [1,18,47,49] in which participants read about a hypothetical crash and then attributed responsibility to the involved parties.

3.1.1. Participants

We adopted a between-subjects design and manipulated three riding conditions (an owner in his private L5 car, a passenger in an L5 taxi, and a passenger in a conventional taxi). We aimed to recruit at least 120 participants online for each condition in all experiments, reaching a rule of thumb for an 80% powered study (i.e., at least 100 participants per condition) [69].

We excluded data from one participant in study 1a and two in study 1c, as their ages were under 18 years (which did not meet our predetermined requirement). We recruited participants online using a sampling service provided by a platform in China (Sojump). The study's participants were as follows: study 1a: n = 393, M_{age} = 30.4 years, 45.5% women; study 1b: n = 391, M_{age} = 30.5 years, 50.4% women; study 1c: n = 389, M_{age} = 30.6 years, 48.1% women; study 1d: n = 371, M_{age} = 31.1 years, 53.4% women. A sensitivity test [70] showed that our final sample size in studies 1a—1d could provide 80% power to detect an effect of η_p^2 = 0.024–0.026 (small to medium effect size; α = 0.05). Table S1 in Appendix A provides more demographic information about our participants.

3.1.2. Procedure

We randomly allocated participants to one of the three riding conditions. Participants first read a definition of L5 under two L5 conditions (adapted from Ref. [10]): "An L5 automated vehicle refers to: A fully automated driving system can perform all aspects of the dynamic driving task, deal with all circumstances, perform fully automated driving, and free the hands and feet of the driver completely. There is no steering wheel or pedals; thus, car occupants cannot intervene in any driving task," accompanied by a graphic illustration (adapted from Diels and Bos [71]).

Subsequently, participants read about a crash scenario. Take the private L5 car, for instance; the crash in studies 1a and 1b was "On an urban road, an L5 automated driving car is carrying its owner and operating in the automated driving mode. It strikes a pedestrian suddenly crossing the road and causes injury. Before this collision, the car owner is on his phone, and the fully automated driving system does not work properly" (Refs. [47,49]). It mirrored the first pedestrian fatality caused by Uber's AV in 2018 [21]. Study 1c considered the first-person perspective. So, we asked participants to imagine themselves as the in-vehicle occupant. Thus, the crash vignette's wording had a minor change in that the in-vehicle occupant (owner in L5 car and passenger in L5/conventional taxis) was replaced by "you" in study 1c. Study 1d considered the absence of occupants: The occupant called the involved vehicle via a smartphone application and waited for its arrival, and a crash occurred during the waiting period. All the crash scenarios are listed in Table S2 in Appendix A.

After reading about the crash, participants allocated or rated three types of responsibility (causal responsibility, blameworthiness, and legal responsibility, appearing randomly) [12,34] to one of the three occupants. Study 1b measured responsibility rating, and the other three measured responsibility allocation. For instance, the three questions for responsibility rating in the private L5 car condition were (started with "You think in this crash"): "To what extent did the car owner cause the pedestrian injury?" (causal responsibility), "To what extent should the car owner be blamed for this crash?" (blameworthiness), and "To what extent should the car owner be legally responsible for this crash?" (legal responsibility) on a ten-point scale (1 = very low to 10 = very high) (adapted from Liu and Du [72]). The three questions for responsibility allocation in other studies were (started with "You think in this crash"): "To what extent was the pedestrian injury caused by the car owner and the other party (the L5 car manufacturer), respectively?" (causal responsibility), "How much blame should be allocated to the car owner and the other party (the L5 car manufacturer), respectively?" (blameworthiness), and "How much legal responsibility should be allocated to the car owner and the other party (the L5 car manufacturer), respectively?" (legal responsibility). Participants allocated a fixed amount of responsibility (fixed amount = 10) between them using a slider or direct numeric input (adapted from Kirchkamp and Strobel [73]).

In the other two riding conditions, we replaced the term "car owner" with "in-vehicle passenger." The other responsible party was the L5 taxi manufacturer and service operator in the L5 taxi condition and the taxi driver in the conventional taxi condition. We made these changes to match the roles of these responsible parties (although differences in words might influence participants' judgments). We asked participants to ignore pedestrian responsibility while allocating responsibility. We did not specify the L5 vehicles as the responsible party in the two L5 conditions, given that machine agency does not intrinsically make machines moral agents [74] and machines such as L5 vehicles cannot respond to punishment and blame [27]. In addition, before ascribing responsibility, we measured the negative affect evoked by the crash, as shown in Appendix A .

Finally, participants submitted their gender, age, and possession of a driving license and received compensation equivalent to 0.31 USD.

3.2. Results and discussion

As the judgments of the three responsibility measures were similar in magnitude and had a significant internal consistency, we averaged them to obtain a single factor of responsibility rating (Cronbach's α = 0.94 in study 1b) or allocation (Cronbach's α = 0.90–0.92 in other sub-studies), similar to previous research [18]. We conducted analysis of covariance (ANCOVA) tests with occupant responsibility as the outcome variable, riding condition (L5 car = 0, L5 taxi = 1, conventional taxi = 2) as the independent variable, and gender (male = 0, female = 1), age, and possession of a driving license (no = 0, yes = 1) as the three covariates. Fig. 3 presents the estimated marginal means (EMMs) for perceived occupant responsibility under different conditions. We selected ANCOVA rather than analysis of variance because the demographic factors of participants in study 1d were not statistically equal across the three riding conditions (which was a coincidence; Table S1). Thus, we added participants' demographic factors as covariates and controlled them in all statistical analyses (for consistency). Pairwise comparisons were performed using the least significant difference method.

The riding condition had a significant influence on occupant responsibility in all four surveys (study 1a: $F_{(2,\ 387)}$ = 42.69, p < 0.001, $\eta_p{}^2$ = 0.181; study 1b: $F_{(2,\ 385)}$ = 62.04, p < 0.001, $\eta_p{}^2$ = 0.244; study 1c: $F_{(2,\ 383)}$ = 45.70, p < 0.001, $\eta_p{}^2$ = 0.193; study 1d: $F_{(2,\ 365)}$ = 4.10, p = 0.017, $\eta_p{}^2$ = 0.022). Participants attributed more responsibility to the two L5 users than to the conventional taxi passenger (ps < 0.01; Fig. 3), regardless of the responsibility metric (responsibility rating vs allocation), participants' perspective (first-person vs third-person perspective), or their absence and presence (occupant absence vs presence). The occupant absence in study 1d was the only exception where the two taxi passengers received equal responsibility (p = 0.441); however, participants still allocated more responsibility to the owner waiting for his private L5 than the passengers waiting for the L5 taxi (ΔEMM = 0.51, t = 2.04, p = 0.042) and conventional taxi (ΔEMM = 0.70, t = 2.77, p = 0.006), as shown in Fig. 3(d).

Previous studies offered early evidence that neither users [18,49] nor drivers of L5 AVs [47] received an attribution of "none" responsibility in their measures (in terms of blame assignment) when these vehicles cause a crash. Taking a faultless passenger in a conventional taxi as the referent, we offer clear-cut and robust evidence that people generally attribute more responsibility to users in L5 AVs (private cars and robotaxis) than to conventional taxi passengers in all conditions when their vehicles cause identical crashes and damage. The obtained responsibility attribution tendency is puzzling and counter-normative because the users in L5 AVs (private cars or robotaxis) play literally and practically no role in the vehicles' driving decisions and control.

4. Study 2

To interpret the puzzling finding obtained in study 1, we extended study 1a to explore the underlying psychological mechanisms in studies 2 and 3. Study 2 assumed that observers believe that the users of driverless vehicles have more control over driving and thus more responsibility for the driving outcome. Note that such a belief is technically incorrect because in-vehicle occupants are merely passengers of driverless vehicles and do not have direct vehicle control. As study 2 considered participants from China (study 2a) and the Republic of Korea (study 2b) to examine this assumption, its cross-national design also showed whether the obtained responsibility attribution tendency exists in different countries.

4.1. Method

4.1.1. Participants

As in study 1, we adopted a between-subjects design and manipulated the three riding conditions in each survey. Participants were recruited online through Sojump in China (study 2a: n = 395, M_{age} = 30.0 years, 47.6% women) and Survey Billy in the Republic of Korea (study 2b: n = 360, M_{age} = 41.4 years, 40.3% women). In addition, we excluded one participant as he was under 18 in study 2a. The sample sizes could provide 80% power to detect an effect of $\eta_p{}^2$ = 0.024 in study 2a and 0.026 in study 2b (small to medium effect size; α = 0.05).

4.1.2. Procedure

Study 2's procedure was identical to that of study 1a, with a difference in the measures. Studies 2a and 2b requested participants to rate occupant controllability while driving before they became aware of a crash. Participants responded to three items (taking the private L5 car, for instance), "I feel that while riding in the L5 automated driving car, the whole driving process is under the control of the in-vehicle owner/the in-vehicle owner controls the whole driving process/the in-vehicle owner is in charge of the whole driving process," on a seven-point scale (1 = strongly disagree to 7 = strongly agree), for their perceived controllability of the in-vehicle humans (the owner in the L5 car or the passenger in the L5/conventional taxis). These three items for perceived controllability (Cronbach's α = 0.90 in Study 2a and 0.93 in study 2b) were adapted from Ref. [60]. As in study 1a, participants responded to the three questions for responsibility allocation (Cronbach's α = 0.94 in study 2a and 0.95 in study 2b) and one question for negative affect.

4.2. Results and discussion

Study 2 replicated the results of study 1a. An ANCOVA revealed a significant influence of the riding condition on perceived occupant responsibility (study 2a: $F_{(2,389)}$ = 41.07, p < 0.001,

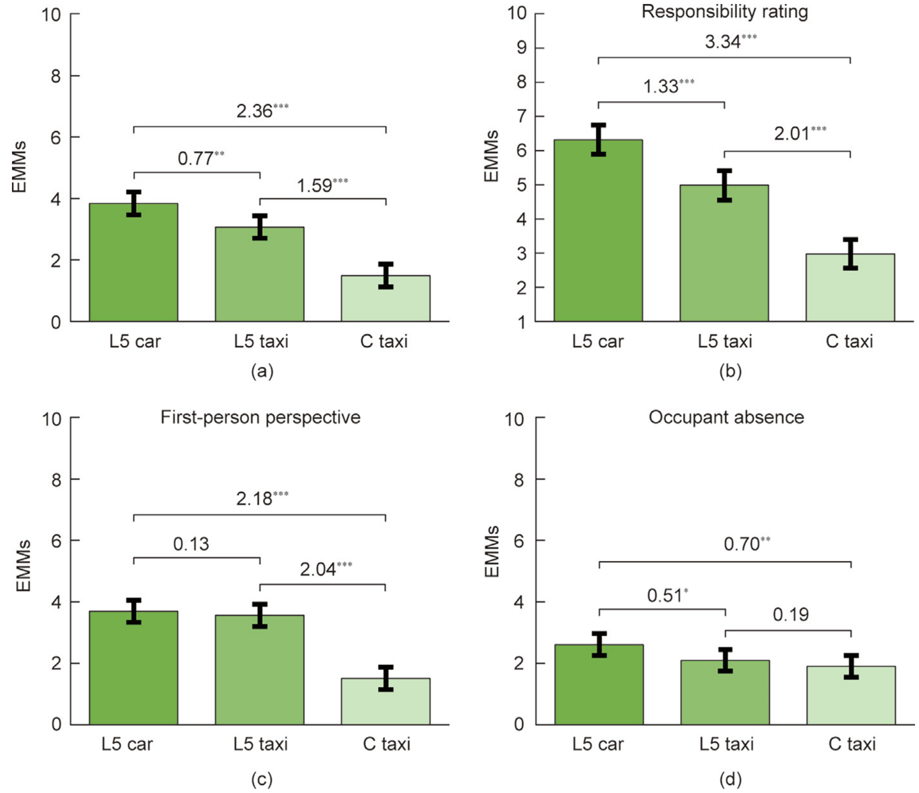

Fig. 3. Estimated marginal means (EMMs) of occupant responsibility in (a) study 1a, (b) study 1b, (c) study 1c, and (d) study 1d. Error bars = ±2 standard errors (SEs). $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. We set study 1a as the default (responsibility allocation × third-person perspective × occupant presence). The difference between the other three sub-studies and study 1a is shown in the caption.

η_p^2 = 0.174; study 2b: $F_{(2,354)}$ = 15.65, p < 0.001, η_p^2 = 0.081). Perceived occupant responsibility was different across the three riding conditions in both countries (ps < 0.01; Figs. 4(a) and (b)), which was most in the L5 car and least in the conventional taxi, with the only exception that the occupant responsibility in the L5 car condition was marginally more than that in the L5 taxi condition in the Republic of Korea (ΔEMM = 0.53, t = 1.74, p = 0.083; Fig. 4(b)).

The pooled data across the three riding conditions showed a positive association between perceived controllability and responsibility allocation (study 2a: r = 0.41, p < 0.001, 95% confidence interval (CI) [0.33, 0.49]; study 2b: r = 0.21, p < 0.001, 95% CI [0.11, 0.31]). However, study 2 rejected our assumption that observers believe that the occupants of driverless vehicles have more controllability (agency) over driving and, thus, more responsibility for the driving outcome, as they judged that the occupants had equal controllability across the three riding conditions (study 2a: p = 0.872; study 2b: p = 0.117; Fig. 4).

We also included participants' nationality (China = 0, the Republic of Korea = 1) as an independent variable based on previous ANCOVA tests. We observed non-significant main effects of nationality (p = 0.962) and riding condition (p = 0.621) on occupant controllability, with a non-significant interaction effect (p = 0.306). With respect to perceived occupant responsibility, nationality ($F_{(1,\ 746)}$ = 14.83, p < 0.001, η_p^2 = 0.019) and riding condition ($F_{(2,\ 746)}$ = 52.26, p < 0.001, η_p^2 = 0.123) had a main effect. Their interaction effect was not significant (p = 0.099). South Korean participants perceived more occupant responsibility than Chinese participants (ΔEMM = 0.79, t = 3.85, p < 0.001; see Fig. 4). Perceived occupant responsibility was different across the three riding conditions (ps < 0.01), as reported previously.

Several recent studies [46,47] adopted (but did not empirically examine) this concept of human controllability to explain why users or drivers take less responsibility when vehicle automation levels increase. Corresponding with Weiner's [36] attribution theory, when our participants perceived the occupants to have greater control over driving, they allocated more responsibility to human occupants under all three riding conditions in both countries. However, their perceived occupant controllability did not account for the differences in the responsibilities they attributed to the occupants under the three riding conditions.

5. Study 3

Study 2 failed to confirm the role of perceived controllability in explaining the observed responsibility attribution tendency involving L5 AV accidents. As explained in Section 2.2, follow-up study 3 examined the assumption that people attribute more responsibility to L5 AV users because they believe that L5 AV users should be more aware of the consequences of using driverless vehicles (reasonable foreseeability).

5.1. Method

5.1.1. Participants

As in study 1a, we recruited participants online in China and excluded two as their age was under 18, leaving a final sample of 369 participants (M_{age} = 30.6 years, 51.8% women), which could provide 80% power to detect an effect of η_p^2 = 0.026 (small to medium effect size; α = 0.05).

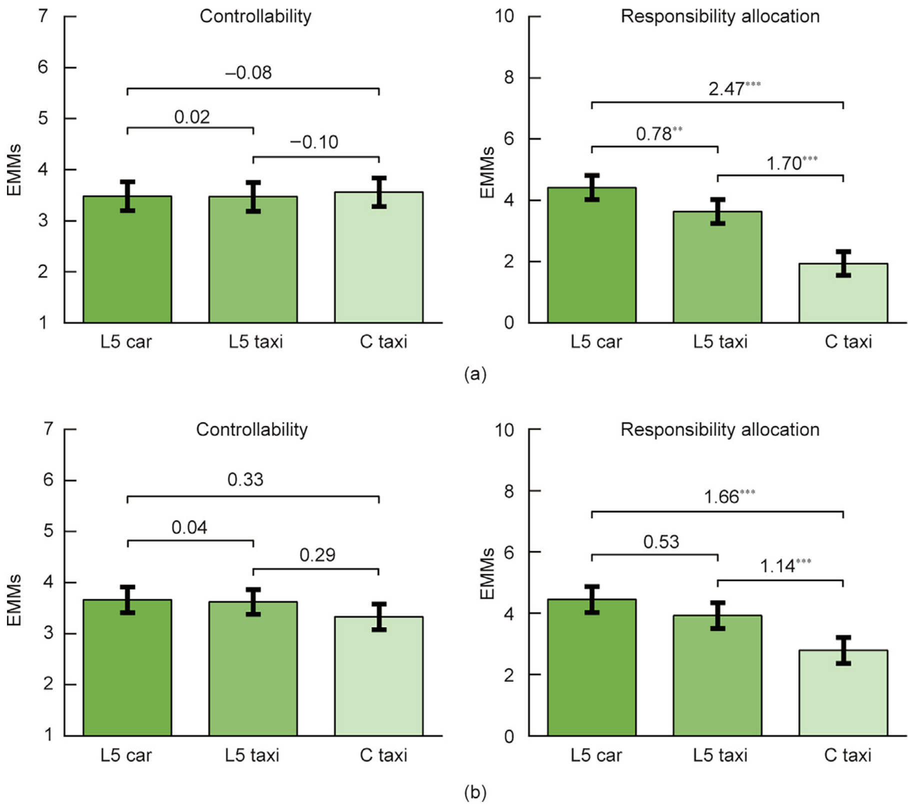

Fig. 4. EMMs of occupant responsibility and controllability in (a) study 2a in China and (b) study 2b in the Republic of Korea. Error bars = ±2 SE. $^{**}p < 0.01$; $^{***}p < 0.001$.

5.1.2. Procedure

Study 3 had the same procedure as Study 1a's procedure, with a difference in the measures. Participants responded to two items (adapted from McCaul et al. [75]) assessing the occupants' foreseeability over the crash and the consequence of their riding (Cronbach's α = 0.80) after judging responsibility allocation. They also responded to three items assessing the other responsible party's controllability (Cronbach's α = 0.83; Appendix A) before participants heard about the crash. We measured foreseeability by the following (taking the private L5 car, for instance): "I think the in-vehicle owner should be able to foresee these kinds of crashes," and "I think the in-vehicle owner should be able to foresee potential crash consequences from riding in L5 automated driving cars," on a seven-point scale (1 = strongly disagree to 7 = strongly agree). Subsequently, we measured responsibility allocation (Cronbach's α = 0.89) and negative affect as done previously.

5.2. Results and discussion

As earlier, the riding condition influenced perceived occupant responsibility ($F_{(2,363)} = 34.13$, $p < 0.001$, $\eta_p^2 = 0.158$). This differed across the three conditions ($ps < 0.001$), and was highest for the L5 car and lowest for the conventional taxi, as shown in Fig. 5(a). The riding condition also significantly influenced perceived foreseeability ($F_{(2,363)} = 37.19$, $p < 0.001$, $\eta_p^2 = 0.170$). Perceived occupant foreseeability differed across the three conditions ($ps < 0.001$) and was highest for the L5 car and lowest for the conventional taxi, as shown in Fig. 5(a). It was positively correlated with perceived occupant responsibility across the three riding conditions ($r = 0.50$, $p < 0.001$, 95% CI [0.42, 0.57]).

We conducted a mediation analysis (model 4 with 5000 resamples, following Hayes and Preacher [76]) with the riding condition as the multi-categorical independent variable (we designed two contrasts: the L5 car relative to the conventional taxi and the L5 taxi relative to the conventional taxi). Our mediation analysis showed the indirect effects of the riding condition through perceived occupant foreseeability (f_1 = 0.81, 95% CI [0.56, 1.10]; f_2 = 0.48, 95% CI [0.27, 0.73]; Fig. 5(b)). As the direct effect of the riding condition was still significant (c_1'= 1.22, $p < 0.001$; c_2' = 0.58, p = 0.015), perceived occupant foreseeability had a partial mediating effect.

Therefore, study 3 supported the idea that people attribute more responsibility to L5 AV users because they believe they should be more aware of the consequences of using L5 AVs. Thus, we identified the role of reasonable foreseeability [64] in explaining why people have different judgments of occupants' responsibility when they ride vehicles (L5 AVs vs conventional taxis) and cause identical harm.

6. Summary of the differences in responsibility attribution

Through a series of single-paper meta-analyses [77], we summarized the results of studies 1a, 1b, 2a, 2b, and 3 (all related to the third-person perspective and occupant presence). As studies 1c and 1d had unique, single-study-level moderators (first-person perspective and occupant absence), we could not analyze their results using the current single-paper meta-analyses. The point estimates of the simple effect of L5 car versus conventional taxi, L5 taxi versus conventional taxi, and L5 car versus L5 taxi were 1.07 (95% CI [0.93, 1.21]), 0.67 (95% CI [0.53, 0.82]), and 0.40 (95% CI [0.25, 0.54]) (all in standard deviation units), respectively (Fig. 6). Therefore, participants allocated the users of driverless vehicles (private cars and robotaxis) more responsibility than conventional taxi passengers when these vehicles caused a crash.

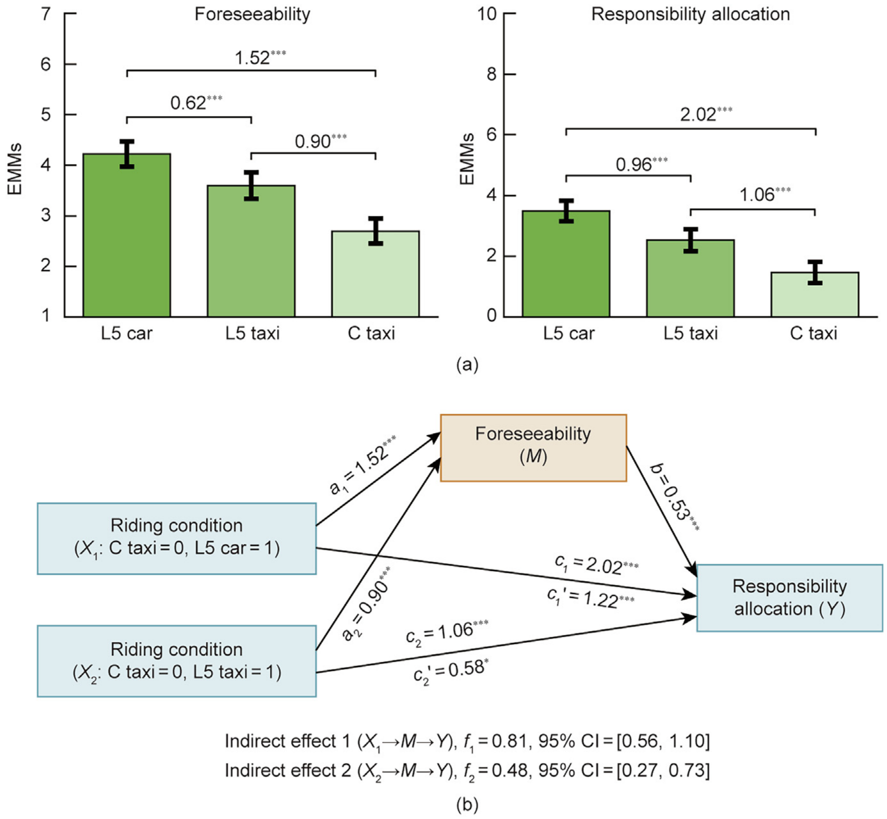

Fig. 5. (a) EMMs of occupant responsibility and foreseeability and (b) mediation analysis in study 3. Error bars = ±2 SE. $^*p$ < 0.05; $^{**}p$ < 0.01; $^{***}p$ < 0.001.

7. General discussion

Are users of fully automated and driverless vehicles subject to responsibility for their crashes? If so, then on what grounds? These questions are emerging in an era of machines and AI rising in transportation [18,19] and other safety–critical settings [1]. Our research is among the first to analyze human responsibility for the adverse outcomes of using machines with full automation. Next, we discuss the implications of our findings in both theory and practice.

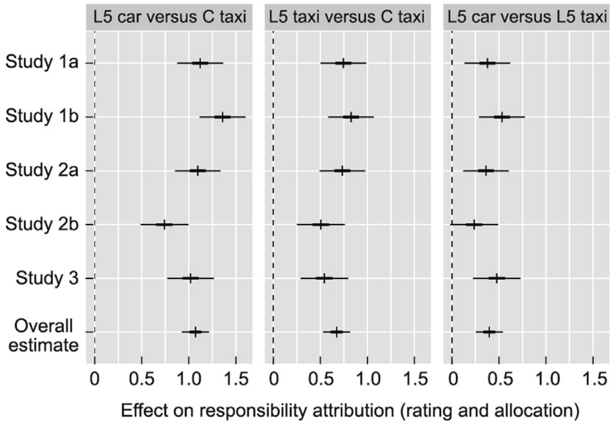

Fig. 6. Single-paper meta-analyses of the influence of riding condition on occupant responsibility attribution (rating and allocation) in studies 1–3. The thick and thin lines represent 50% and 95% confidence intervals, respectively.

Our theoretical contributions lie in the empirical investigation of the determinants and psychological processes of responsibility attribution after fully automated machines cause harm. We expect our findings to have practical implications, because responsibility attribution is essential for understanding consumer reactions [60,78] and determining AV design and pricing [18]. Our findings uncovered societal expectations of liability determinations for AV crashes, thus offering insights for building a socially acceptable regulatory scheme.

7.1. Theoretical implications

Most normative analyses from ethics, law, philosophy, and other disciplines or common sense dictate that it is unfair to hold a passenger or owner of a driverless car responsible for an act over which they have no requisite control [18,24]. In contrast, previous empirical studies [18,47,49,52,53] offered early evidence that users (or “drivers”) of L5 fully AVs did not receive an attribution of “no” responsibility when these vehicles caused a crash. Thus, a conflict might exist between objective controllability and responsibility [53]. Researchers have not yet examined its potential boundary conditions and underlying reasons.

In our three studies with seven experiments, we treated the conventional taxi passenger as the reference (who is attributed no responsibility for the involved crash according to the current tort system). We compared the responsibilities attributed to this faultless passenger and L5 AV users (L5 car owners and taxi passengers). We observed an interesting “dilemma” or tendency: Although L5 AV users have no direct vehicle control, people might require them to take partial responsibility for harm caused by these vehicles. This counterintuitive tendency was robust across the different contexts studied. In particular, participants imagining themselves as L5 AV users also assigned more responsibility to themselves than participants imagining themselves as conventional taxi passengers (Fig. 3(c)), which is surprising given the existence of the well-known self-serving tendency [59] or defensive attribution tendency [58]. It parallels the idea that society could impose strict liabilities on the owners of driverless cars [79]; however, this idea cannot explain why passengers in driverless taxis are also required to bear more responsibility than those in conventional taxis. Similarly, the concept of ownership (i.e., possession of an object) may account for more responsibility attributed to the owners of driverless cars [60,61] but cannot account for more responsibility attributed to the passengers of driverless taxis.

Therefore, our theoretical contribution is that we confirmed the persistence of this human tendency (a conflict between objective controllability and responsibility) through robust evidence and explained it through the lens of foreseeability [62,63], more specifically, reasonable foreseeability [64]. Observers are more likely to perceive that users of driverless vehicles (private cars or taxis) should be able to foresee the consequences of the trip and thus apportion more responsibility to these users for their vehicle-caused crashes. This may be counterfactual thinking or a biased intuitive reaction to these users, as they may not have more foresight in reality [80]. Underlying the role of reasonable foreseeability may be that observers think using driverless vehicles is riskier than using conventional taxis and thus blame their users more if driverless vehicles cause traffic risks and harm. Reasonable foreseeability also predicts liability in classic tort doctrines such as negligence, product liability, and strict liability [67,81,82], although its meaning in lay terms differs from that in these doctrines.

The broader literature on responsibility attribution involving autonomous machines [2,55,56] has discussed important factors influencing responsibility attribution, such as perceived intention and objective foreseeability. We contribute to the literature by examining the responsibility attributed to users when fully autonomous machines err and highlight the importance of reasonable foreseeability of machine users on the responsibility attributed to them. For instance, Hidalgo et al. [55] reported that a company hiring an AI machine (vs a human marketer) to create advertising images was attributed much more responsibility when the AI machine (vs the human marketer) created identical lewd images. According to our findings, the reasonable foreseeability of the company may partly account for Hidalgo et al.'s findings.

7.2. Practical implications

Before discussing the potential practical implications, we emphasize that we do not state that participants' judgments collected from our experiments should be directly translated into legal rules for AV accidents. However, their judgments should be appropriately anticipated and managed in public discourse and legal regulations.

Even if AVs are eventually safer than human drivers, they still cause road trauma and thus raise concerns about who or what should be responsible. Certain voices from the AV industry and legislative provisions can alleviate consumers' legal concerns resulting from the use of driverless vehicles. For instance, several automakers such as Volvo [31] and Audi [32] have publicly promised that they will take full responsibility for their vehicle-caused crashes; the UK's Automated and Electric Vehicles Act 2018 suggests that the insurer is liable for the damage caused by an insured AV when it is driving itself on the road rather than its owner if it is used appropriately [83].

However, our observed "dilemma" is that although L5 AV users (L5 car owners and taxi passengers) have no direct vehicle control, society might require them to take partial responsibility for harm caused by these vehicles—the results of our seven experiments might suggest a different possibility in the future. We do not consider it an attributional error of "naïve" participants. Instead, it might indicate a social bias against users of driverless vehicles in terms of responsibility attribution. Previous empirical studies observed that laypeople [72,84,85] and trial judges [86] assess AV-caused crashes (vs identical human-caused crashes) more severely and blame them more, thus exhibiting biases toward AVs and their usage concerning responsibility attribution. More broadly, it might speak to a negative social signal for using autonomous machines over which humans have no control [87].

In line with our discovery, specific legal regulations require owners or users of L5 AVs to bear liability for crashes caused by their vehicles, even when the accidents are beyond their control. Recently, Shenzhen City, China, issued a first-of-its-kind regulation to fill the legal gap for AVs and clarify rules for responsibility in the event of AV accidents [88]. It states that if there is a human driver in the driver's seat and the vehicle is operating autonomously, the driver will be held responsible by transportation authorities; if no driver is in the driver's seat, the owner or user of the vehicle takes the responsibility (in terms of paying compensation to the victim); if the vehicle's defects cause the accident, the owner or user of the vehicle, after paying the bill, can seek compensation from the vehicle manufacturer or seller. This regulation is friendly to the manufacturer but not to the owner or user. Under this regulation, the owners or users of L5 AVs are still the locus of liability; however, our participants from China and the Republic of Korea required them to bear partial or minor responsibility (i.e., they are not the major responsible party).

Although our central findings and the current legal regulation in Shenzhen, China [88] both support the idea that L5 AV owners or users are not clear of responsibility if these driverless cars cause harm, we believe that such a responsibility scheme might backfire. Our findings imply a potential psychological roadblock to driverless car adoption: when encountering crashes, consumers of driverless cars may face greater public pressure and moral condemnation. Judges are likely to have reactions similar to laypersons [86]. Thus, in public discourse and legal proceedings, society might regard consumers of driverless vehicles as a "moral crumple zone" [22] or "legal sponge" [23], shifting responsibility from driverless vehicle manufacturers and/or ride-sharing operators. Blaming consumers could prevent ownership or use of these fully automated and driverless vehicles, casting a shadow over their future. Specific measures, such as public discourse and insurance [18,72,80], are required to reduce this social blame against consumers of driverless vehicles after crashes and to reduce the burden of responsibility placed on them.

7.3. Research limitations

Given the theoretical and practical implications of our study, it is crucial to highlight its limitations. First, similar to other vignette-based experiments, our vignette-based experiments may lack sufficient psychological realism [18]. However, they mimic how people interact with AV accidents in the coming decades [55]: "by hearing stories about them in the news or social media." The experimental vignette methodology is widely used and well-studied in the behavioral and social sciences. Zhai et al. [51] recently analyzed public judgments of responsibility after an actual AV crash (i.e., the 2018 Uber AV crash) and then examined more variations of this crash in a vignette-based experiment. The two mixed methods produced a similar finding in that Uber's test driver was attributed more responsibility than the company (i.e., Uber). Their research provides evidence of the external validity of vignette-based studies.

Second, responsibility judgments usually occur within a richer context than the carefully controlled scenarios used in our work, influenced by more contextual factors, such as emotion-arousing details and standpoints expressed in news reports [18]. Third, our findings may have cultural limitations, as we obtained them from two Asian countries. Future studies should examine this issue in different cultural contexts. Finally, we cannot rule out that other accounts in addition to reasonable foreseeability might be effective (e.g., the occupant's "proximity" to the crash and participants' lower trust in driverless vehicles versus conventional vehicles) to explain the obtained responsibility attribution tendency involving driverless vehicles.

8. Conclusions

With a focus on fully automated driverless vehicles, our research reveals a dilemma that consumers of fully AVs (owners of fully automated cars and passengers of driverless taxis) versus passengers in conventional taxis are attributed more responsibility when these vehicles cause identical harm (although none of these riders have direct control over driving). This responsibility difference is partly because they are expected to have more foreseeability over the riding consequences (reasonable foreseeability). The observed dilemma indicates the potential social blame against consumers of driverless vehicles after crashes. Public discourse, insurance, and other societal measures should focus on reducing the social blame and burden of responsibility placed on driverless-vehicle consumers.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (72071143).

Compliance with ethics guidelines

Siming Zhai, Lin Wang, and Peng Liu declare that they have no conflicts of interest or financial conflicts to disclose.

Appendix A. Supplementary data

All data, codes, and results are publicly available on the *Open Science Framework* at https://osf.io/58s42/?view_only=6b5b8d4bade449a8a0f8d0cc8836ac57.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2023.10.008.

References

[1] Jamjoom AAB, Jamjoom AMA, Marcus HJ. Exploring public opinion about liability and responsibility in surgical robotics. Nat Mach Intell 2020;2(4):194–6.

[2] Bigman YE, Waytz A, Alterovitz R, Gray K. Holding robots responsible: the elements of machine morality. Trends Cogn Sci 2019;23(5):365–8.

[3] Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon JF, Breazeal C, et al. Machine behaviour. Nature 2019;568(7753):477–86.

[4] Yang GZ, Cambias J, Cleary K, Daimler E, Drake J, Dupont PE, et al. Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy. Sci Robot 2017;2(4):eaam8638.

[5] Fagnant DJ, Kockelman K. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transp Res Policy Pract 2015;77:167–81.

[6] Wang J, Huang H, Li K, Li J. Towards the unified principles for level 5 autonomous vehicles. Engineering 2021;7(9):1313–25.

[7] Dingus TA, Guo F, Lee S, Antin JF, Perez M, Buchanan-King M, et al. Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proc Natl Acad Sci USA 2016;113(10):2636–41.

[8] Wang X, Liu Q, Guo F, Fang S, Xu X, Chen X. Causation analysis of crashes and near crashes using naturalistic driving data. Accid Anal Prev 2022;177:106821.

[9] National Highway Traffic Safety Administration (NHTSA). Automated vehicles for safety [Internet]. Washington, DC: National Highway Traffic Safety Administration; 2020 [cited 2020 Oct 18]. Available from: https://shorturl.at/mtyI0.

[10] Society of Automotive Engineers (SAE). J3016. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Washington, DC: SAE International/ISO; 2021.

[11] Fosch-Villaronga E, Khanna P, Drukarch H, Custers BHM. A human in the loop in surgery automation. Nat Mach Intell 2021;3(5):368–9.

[12] Bonnefon JF, Černy D, Danaher J, Devillier N, Johansson V, Kovacikova T, et al. Ethics of connected and automated vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility. Brussels: EU Commission; 2020.

[13] Bonnefon JF, Shariff A, Rahwan I. The social dilemma of autonomous vehicles. Science 2016;352(6293):1573–6.

[14] Pattinson JA, Chen H, Basu S. Legal issues in automated vehicles: critically considering the potential role of consent and interactive digital interfaces. Humanit Soc Sci Commun 2020;7(1):153.

[15] Marchant G, Lindor R. The coming collision between autonomous vehicles and the liability system. Santa Clara Law Rev 2012;52(4):1321–40.

[16] Liu P, Du M, Li T. Psychological consequences of legal responsibility misattribution associated with automated vehicles. Ethics Inf Technol 2021;23(4):763–76.

[17] Stilgoe J. Self-driving cars will take a while to get right. Nat Mach Intell 2019;1(5):202–3.

[18] Awad E, Levine S, Kleiman-Weiner M, Dsouza S, Tenenbaum JB, Shariff A, et al. Drivers are blamed more than their automated cars when both make mistakes. Nat Hum Behav 2020;4(2):134–43.

[19] Hancock PA, Nourbakhsh I, Stewart J. On the future of transportation in an era of automated and autonomous vehicles. Proc Natl Acad Sci USA 2019;116(16):7684–91.

[20] A tragic loss [Internet]. Austin: Tesla; 2016 Jun 30 [cited 2019 May 1]. Available from: https://www.tesla.com/blog/tragic-loss.

[21] McFarland M. Uber self-driving car operator charged in pedestrian death [Internet]. Atlanta: CNN; 2020 [cited 2021 Jan 23]. Available from: https://rb.gy/hkskb.

[22] Elish MC. Moral crumple zones: cautionary tales in human–robot interaction. Engag Sci Technol Soc 2019;5:40–60.

[23] Holford WD. An ethical inquiry of the effect of cockpit automation on the responsibilities of airline pilots: dissonance or meaningful control? J Bus Ethics 2022;176(1):141–57.

[24] Geistfeld MA. A roadmap for autonomous vehicles: state tort liability, automobile insurance, and federal safety regulation. Calif LRev 2017;105(6):1611.

[25] Grieman K. Hard drive crash: an examination of liability for self-driving vehicles. J Intell Prop Info Tech Elec Com L 2018;9(3):294–309.

[26] Mackie T. Proving liability for highly and fully automated vehicle accidents in Australia. Comput Law Secur Rev 2018;34(6):1314–32.

[27] Vladeck DC. Machines without principals: liability rules and artificial intelligence. Wash Law Rev 2014;89(1):117–50.

[28] Hevelke A, Nida-Rümelin J. Responsibility for crashes of autonomous vehicles: an ethical analysis. Sci Eng Ethics 2015;21(3):619–30.

[29] Marchant G, Bazzi R. Autonomous vehicles and liability: what will juries do? B U J Sci Tech L 2020;26(1):67–119.

[30] Gurney JK. Imputing driverhood: applying a reasonable driver standard to accidents caused by autonomous vehicles. In: Lin P, Abney K, Jenkins R, editors. Robot ethics 20: from autonomous cars to artificial intelligence. Oxford: Oxford University Press; 2017.

[31] Atiyeh C. Volvo will take responsibility if its self-driving cars crash [Internet]. Harlan: Car and Driver; 2015 [cited 2019 May 1]. Available from: https://shorturl.at/cIRW9.

[31] Atiyeh C. Volvo will take responsibility if its self-driving cars crash [Internet]. Harlan: Car and Driver; 2015 [cited 2019 May 1]. Available from: https://shorturl.at/cIRW9.

[32] Maric P. Audi to take full responsibility in event of autonomous vehicle crash [Internet]. Drive; 2017 Sep 11 [cited 2019 May 1]. Available from: https://shorturl.at/pEPTX.

[33] Lima G, Grgić-Hlača N, Cha M. Human perceptions on moral responsibility of AI: a case study in AI-assisted bail decision-making. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; 2021 May 8–13; Yokohama, Japan; 2021.

[34] van de Poel I. Moral responsibility. In: van de Poel I, Royakkers L, Zwart SD, editors. Moral responsibility and the problem of many hands. New York City: Routledge; 2015.

[35] Cushman F. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. Cognition 2008;108(2):353–80.

[36] Weiner B. Social motivation, justice, and the moral emotions: an attributional approach. Mahwah: Lawrence Erlbaum Associates; 2006.

[37] Aristotle. Nicomachean ethics. In: Barnes J, editor. The complete works of Aristotle. Princeton: Princeton University Press; 1984.

[38] Johnson DG. Technology with no human responsibility? J Bus Ethics 2015;127(4):707–15.

[39] Fischer JM, Ravizza M. Responsibility and control: a theory of moral responsibility. Cambridge: Cambridge University Press; 1998.

[40] Nelkin DK. Moral luck. In: Zalta EN, editor. The Stanford encyclopedia of philosophy. Stanford: Stanford University; 2004.

[41] Nagel T. Mortal questions. Cambridge: Cambridge University Press; 1979.
[42] Williams B. Moral luck. Cambridge: Cambridge University Press; 1981.
[43] Howe JST. Towards a control-centric account of tort liability for automated vehicles. Torts Law J 2021;26(3):221–43.
[44] Huddy XP. The law of automobiles. 6th ed. Albany: Matthew Bender; 1922.
[45] Rahwan I. Society-in-the-loop: programming the algorithmic social contract. Ethics Inf Technol 2018;20(1):5–14.
[46] Copp CJ, Cabell JJ, Kemmelmeier M. Plenty of blame to go around: attributions of responsibility in a fatal autonomous vehicle accident. Curr Psychol 2023;42(8):6752–67.
[47] Bennett JM, Challinor KL, Modesto O, Prabhakharan P. Attribution of blame of crash causation across varying levels of vehicle automation. Saf Sci 2020;132:104968.
[48] McManus RM, Rutchick AM. Autonomous vehicles and the attribution of moral responsibility. Soc Psychol Personal Sci 2019;10(3):345–52.
[49] Pöllänen E, Read GJM, Lane BR, Thompson J, Salmon PM. Who is to blame for crashes involving autonomous vehicles? Exploring blame attribution across the road transport system. Ergonomics 2020;63(5):525–37.
[50] Li J, Cho MJ, Zhao X, Ju W, Malle BF. From trolley to autonomous vehicle: perceptions of responsibility and moral norms in traffic accidents with self-driving cars. In: SAE 2016 World Congress and Exhibition; 2016 Apr 12–14; Detroit, MI, USA; 2016.
[51] Zhai S, Gao S, Wang L, Liu P. When both human and machine drivers make mistakes: whom to blame? Transp Res Policy Pract 2023;170:103637.
[52] Zhai S, Wang L, Liu P. Human and machine drivers: sharing control, sharing responsibility. Accid Anal Prev 2023;188:107096.
[53] Aguiar F, Hannikainen IR, Aguilar P. Guilt without fault: accidental agency in the era of autonomous vehicles. Sci Eng Ethics 2022;28(2):11.
[54] Jahedi S, Méndez F. On the advantages and disadvantages of subjective measures. J Econ Behav Organ 2014;98:97–114.
[55] Hidalgo CA, Orghiain D, Canals JA, de Almeida F, Martin N. How humans judge machines. Cambridge: MIT Press; 2021.
[56] Franklin M, Ashton H, Awad E, Lagnado D. Causal framework of artificial autonomous agent responsibility. In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society; 2022 Aug 1–3; Oxford, UK; 2022. p. 276–84.
[57] Malter MS, Kim SS, Metcalfe J. Feelings of culpability: just following orders versus making the decision oneself. Psychol Sci 2021;32(5):635–45.
[58] Shaver KG. Defensive attribution: effects of severity and relevance on the responsibility assigned for an accident. J Pers Soc Psychol 1970;14(2):101–13.
[59] Bradley GW. Self-serving biases in the attribution process: a reexamination of the fact or fiction question. J Pers Soc Psychol 1978;36(1):56–71.
[60] Jörling M, Böhm R, Paluch S. Service robots: drivers of perceived responsibility for service outcomes. J Serv Res 2019;22(4):404–20.
[61] Palamar M, Le DT, Friedman O. Acquiring ownership and the attribution of responsibility. Cognition 2012;124(2):201–8.
[62] Shaver KG. The attribution of blame: causality, responsibility, and blameworthiness. New York City: Springer; 1985.
[63] Alicke MD. Culpable control and the psychology of blame. Psychol Bull 2000;126(4):556–74.
[64] Lagnado DA, Channon S. Judgments of cause and blame: the effects of intentionality and foreseeability. Cognition 2008;108(3):754–70.
[65] Heider F. The psychology of interpersonal relations. New York City: John Wiley & Sons; 1958.
[66] Furlough C, Stokes T, Gillan DJ. Attributing blame to robots: I. the influence of robot autonomy. Hum Factors 2021;63(4):592–602.
[67] Khoury L, Smyth S. Reasonable foreseeability and liability in relation to genetically modified organisms. Bull Sci Technol Soc 2007;27(3):215–32.
[68] Nuijten MB, Hartgerink CHJ, van Assen MALM, Epskamp S, Wicherts JM. The prevalence of statistical reporting errors in psychology (1985–2013). Behav Res Methods 2016;48(4):1205–26.
[69] Brysbaert M. How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. J Cogn 2019;2(1):16.
[70] Faul F, Erdfelder E, Lang AG, Buchner AG. *Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175–91.
[71] Diels C, Bos JE. Self-driving carsickness. Appl Ergon 2016;53:374–82.
[72] Liu P, Du Y. Blame attribution asymmetry in human–automation cooperation. Risk Anal 2022;42(8):1769–83.
[73] Kirchkamp O, Strobel C. Sharing responsibility with a machine. J Behav Exp Econ 2019;80:25–33.
[74] Banks J. A perceived moral agency scale: development and validation of a metric for humans and social machines. Comput Human Behav 2019;90:363–71.
[75] McCaul KD, Veltum LG, Boyechko V, Crawford JJ. Understanding attributions of victim blame for rape: sex, violence, and foreseeability. J Appl Soc Psychol 1990;20(1):1–26.
[76] Hayes AF, Preacher KJ. Statistical mediation analysis with a multicategorical independent variable. Br J Math Stat Psychol 2014;67(3):451–70.
[77] McShane BB, Böckenholt U. Meta-analysis of studies with multiple contrasts and differences in measurement scales. J Consum Psychol 2022;32(1):23–40.
[78] Weiner B. Attributional thoughts about consumer behavior. J Consum Res 2000;27(3):382–7.
[79] Shavell S. On the redesign of accident liability for the world of autonomous vehicles. J Legal Stud 2020;49(2):243–85.
[80] Shariff A, Bonnefon JF, Rahwan I. Psychological roadblocks to the adoption of self-driving vehicles. Nat Hum Behav 2017;1(10):694–6.
[81] Karnow CEA. The application of traditional tort theory to embodied machine intelligence. In: Calo R, Froomkin AM, Kerr I, editors. Robot law. Northampton: Edward Elgar; 2016.
[82] Van Uytsel S. Different liability regimes for autonomous vehicles: one preferable above the other? In: Van Uytsel S, Vasconcellos Vargas D, editors. Autonomous vehicles: business, technology and law. Singapore: Springer; 2021.
[83] Automated vehicles: joint report. Law Commission of England and Wales and Scottish Law Commission; 2022.
[84] Liu P, Du Y, Xu Z. Machines versus humans: people's biased responses to traffic accidents involving self-driving vehicles. Accid Anal Prev 2019;125:232–40.
[85] Franklin M, Awad E, Lagnado D. Blaming automated vehicles in difficult situations. iScience 2021;24(4):102252.
[86] Rachlinski JJ, Wistrich AJ. Judging autonomous vehicles. Yale J Law Technol 2022;24:706–66.
[87] Bigman YE, Gray K. People are averse to machines making moral decisions. Cognition 2018;181:21–34.
[88] Ma S. Shenzhen gives green light to fully autonomous vehicles [Internet]. Beijing: China Daily; [updated 2022 Jul 11; cited 2022 Aug 1]. Available from: https://shorturl.at/mvO18.