Self-Driving Vehicles against Human Drivers: Equal Safety is Far from Enough

Peng Liu[1,2], Lin Wang[3], & Charles Vincent[2]

[1]Tianjin University, China

[2]University of Oxford, UK

[3]Incheon National University, South Korea

Author Note

Peng Liu, College of Management and Economics, Tianjin University and Department of Experimental Psychology, University of Oxford; Lin Wang, Department of Library and Information Science, Incheon National University; Charles Vincent, Department of Experimental Psychology, University of Oxford. Peng Liu and Lin Wang contributed equally to this work.

Correspondence concerning this article should be addressed to Peng Liu, College of Management and Economics, Tianjin University, Tianjin 300072, China. E-mail: pengliu@tju.edu.cn.

Abstract

We examined the acceptable risk of self-driving vehicles (SDVs) compared with that of human-driven vehicles (HDVs) and the psychological mechanisms influencing the decision-making regarding acceptable risk through four studies conducted in China and South Korea. Participants from both countries required SDVs to be 4–5 times as safe as HDVs (Studies 1 and 4). When an SDV and an HDV were manipulated to exhibit equivalent safety performance, participants' lower trust in the SDV, rather than the higher negative affect evoked by the SDV, accounted for their lower risk acceptance of the SDV (Studies 2 and 3). Both lower trust and higher negative affect accounted for why participants were less willing to ride in the SDV (Study 3). These reproducible findings improve the understanding of public assessment of acceptable risk of SDVs and offer insights for regulating SDVs.

Public Significance Statement

This study suggests that people require self-driving cars to be safer than conventional cars. Further, it explains that people trust less in self-driving cars than conventional cars with equivalent safety performance, which in turn leads them to be less willing to accept the risk of self-driving cars.

Keywords*:* Acceptable risk, trust, affect heuristic, self-driving vehicles

Policymakers, scientists, and road safety organizations are enthusiastic about the potential for widespread adoption of self-driving vehicles (SDVs) to reduce traffic accidents, traffic congestion, and air pollution and to increase fuel efficiency, space utilization, and human mobility (Anderson et al., 2016; NHTSA, 2016; Waldrop, 2015). However, SDVs also pose risks and challenges related to safety, security, liability, and regulation (Anderson et al., 2016; Bonnefon, Shariff, & Rahwan, 2016; Fagnant & Kockelman, 2015; Liu, Yang, & Xu, 2019b; Nunes, Reimer, & Coughlin, 2018; Xu et al., 2018). Among dozens of studies on public perceptions, attitude, and acceptance, some found that participants held positive attitudes toward SDVs (e.g., Penmetsa, Adanu, Wood, Wang, & Jones, 2019; Schoettle & Sivak, 2014), whereas others reported participants' resistance and negative attitude to SDVs (e.g., Nielsen & Haustein, 2018; Smith & Anderson, 2017). In particular, participants were concerned about potential risks of SDVs (e.g., hacking) (Liu, Yang, et al., 2019b). The present series of studies builds on these earlier findings to consider what people would regard as an acceptable level of risk for SDVs, an issue which has so far not been well addressed.

Removing control from human drivers is assumed to make SDVs much safer than conventional human-driven vehicles (HDVs) (Mervis, 2017), but this has not yet been confirmed (Banerjee, Jha, Cyriac, Kalbarczyk, & Iyer, 2018). Then, *how safe is safe enough for SDVs*? This popularized question has been widely debated (Halsey, 2017; Hook, 2017; Mervis, 2017). Some lawmakers and regulators have been reported to consider allowing SDVs to be deployed on roads provided they are deemed either as safe as human drivers (Mervis, 2017) or twice as safe as human drivers (Demers, 2018). Others claim that SDVs need to be multiple times (between 2 to 100 times) safer than

HDVs, but, so far, this claim lacks a scientific foundation (Shladover & Nowakowski, 2019). Policy researchers (Kalra & Groves, 2017) argued that a less stringent policy (e.g., allowing SDVs to be slightly safer than the average human driver) should be considered to save more human lives. From a utilitarian standpoint, this policy seems sensible as it will increase the vehicle miles traveled by SDVs and consequently improve the safety performance of SDVs and save more lives in the long-run. However, such a policy may backfire as SDVs allowed on roads under this policy would still cause many traffic accidents, which will lead to people's over-reaction and deter them from adopting SDVs (Liu, Du, & Xu, 2019). The public must be informed about the safety requirements of SDVs for them to make rational choices regarding SDVs (Hancock, Nourbakhsh, & Stewart, 2019). As of now, no law or regulation policy in transportation has specified the safety requirements of SDVs, an obstacle to their proliferation.

The posed question is essentially a problem of determining the acceptable risk of a technology (Fischhoff, Lichtenstein, Slovic, Derby, & Keeney, 1981). A number of approaches have been suggested, including the revealed-preference approach (Starr, 1969), expressed-preference approach (Fischhoff, Slovic, Lichtenstein, Read, & Combs, 1978), cost–benefit analysis, and natural standards. The revealed-preference method (Starr, 1969) assumes that society has already reached an essentially optimal balance between the risks and benefits of any existing technology and that the risks of a new technology are deemed acceptable if they do not exceed the risks of existing technologies providing similar benefits to society. According to this assumption, as SDVs promise to yield more benefits (e.g., environmental benefits) than HDVs, they would be acceptable if they achieve the same safety level achieved by human drivers in current traffic

conditions. However, SDVs probably need to be demonstrably safer than HDVs (Coelingh, Nilsson, & Buffum, 2018; Shladover & Nowakowski, 2019; Waycaster, Matsumura, Bilotkach, Haftka, & Kim, 2018); otherwise, they will not be accepted by the public. Furthermore, in the absence of information on the costs and benefits of SDVs, a formal cost–benefit analysis is impossible. In contrast, the expressed-preference approach, which considers public attitudes, concerns, interests, and values in determining acceptable risk (Fischhoff et al., 1978), is a more appropriate method of assessing the acceptable risk of SDVs before they have been widely deployed. An expressed-preference approach (Liu, Yang, & Xu, 2019a) has been developed to assess the acceptable risk of SDVs (which will be introduced later). Using this approach, a survey in China (Liu, Yang, et al., 2019a) found that participants implicitly wanted SDVs to be 4–5 times as safe as HDVs in terms of fatality risk. This finding on its own has direct policy implications. However, several issues remain unknown, including whether this finding can be replicated in other countries and what factors account for the difference between SDVs and HDVs in terms of acceptable risk.

Affective responses associated with a technology may determine its acceptable risk. The affect heuristic (Finucane, Alhakami, Slovic, & Johnson, 2000; Slovic, Finucane, Peters, & MacGregor, 2004) suggests that the affect evoked by a technology, either consciously or unconsciously, will strongly influence judgments and decision-making related to the technology (e.g., risk and benefit judgments), especially when resources for judgments or decisions are limited and when judgments or decisions are complex (Finucane et al., 2000). If people have higher negative affect triggered by a technology, their willingness to accept the risks of that technology will be lower (Siegrist & Sütterlin,

2014). Trust, a core concept in human-machine interaction and human-automation interaction, has been acknowledged also to be a key determinant of reliance on and acceptance of automation technology (Lee & See, 2004). Considering that accepting an automation technology implies the acceptance of its risks, we assume trust as a determinant of acceptability of risks associated with automation usage. Lower trust in a technology may render the associated risks of the technology to be less acceptable (Siegrist & Sütterlin, 2017). Although these above theoretical explanations are intuitively appealing, they have not been well-tested empirically in research on acceptable risk and risk acceptance of technologies.

We have two research objectives. The major one is to determine the acceptable risk of SDVs to help policymakers and regulators form appropriate policies for managing and regulating SDVs. Considering its policy significance, current related efforts (e.g., Liu, Yang, et al., 2019a) should be considered in multiple countries and confirmed with empirical evidence. The second one is to extend knowledge about the unknown mechanisms underlying the difference between SDVs and HDVs in terms of risk acceptance. This knowledge could offer insights into increasing risk acceptance and potentially support the wider adoption of SDVs. To serve these purposes, four studies were conducted in two Asian countries— China (a middle-income country) and South Korea (a high-income country)—to investigate the socially acceptable risk of self-driving (Studies 1 and 4) and the psychological factors that explain the difference between self-driving and human driving in terms of acceptable risk (Studies 2 and 3). Their data are available at the Open Science Framework: https://osf.io/b9zu4/.

Study 1

Study 1 assesses whether participants from two countries (South Korea and China) have similar acceptable risk of SDVs against HDVs.

Method

The method for measuring the acceptable risk of SDVs was introduced previously (Liu, Yang, et al., 2019a).

Expressed-preference approach to measure acceptable risk.

This approach (Liu, Yang, et al., 2019a) assumes that the risk acceptable rate, that is the proportion of people accepting a risk (such as injury or fatality), is a function of the frequency of this risk. A risk with a higher frequency will be accepted by fewer people. A close relationship between risk acceptance rate and risk frequency (Huang et al., 2013) will enable us to inversely predict acceptable risk frequencies given specific risk acceptance rates. As shown in Figure 1a, given the same risk acceptance rate, the predicted acceptable risk frequency in *A* is lower than that in *B*, indicating a lower level of acceptable risk in *A*. By calculating the ratio of the predicted acceptable risk frequencies between *A* and *B*, we can determine their quantitative difference in acceptable risk. The major steps for assessing the acceptable risk of SDVs versus HDVs are shown in Figure 1b.

Participants.

In both South Korea and China, participants were randomly assigned to one of two vehicle scenarios (HDV or SDV). They were invited through direct contact by interviewers at recreational areas and other areas (e.g., communities, parks, bus stations, and shopping malls). In Study 1 and other three studies, the research purposes were blind

to the interviewers responsible for data collection. All participants were compensated with a monetary reward. In South Korea, 300 participants from each of the HDV and SDV groups submitted responses. Responses of 28 participants in the HDV group and 23 in the SDV group were excluded because they submitted abnormal responses (e.g., identical responses to most questions). In China, 263 participants in the HDV group and 271 participants in the SDV group submitted their responses (Liu, Yang, et al., 2019a). Responses of 24 participants in the HDV group and 11 in the SDV group were excluded due to the same reasons as for those in South Korea. Appendix Table 1 summarizes the demographic information of participants (South Korea, $n_{HDV} = 272$, $n_{SDV} = 277$; China, $n_{HDV} = 239$, $n_{SDV} = 260$).

Material and procedure.

The questionnaire consisted of three parts. In Part I, participants read a textual description about the statistics for road traffic injuries and fatalities worldwide, in their own country, and in the United States (Liu, Yang, et al., 2019a). Participants in the SDV group were also given a textual description and a graphic scenario for SDVs and were told that future daily transportation would be done by SDVs (Liu, Yang, et al., 2019a). In Part II, participants were presented with a series of risk scenarios with various severity levels and frequencies. Three severity levels were designated: *property damage only* (Level 1), *injury* (Level 2), and *fatality* (Level 3) (Shankar, Mannering, & Barfield, 1996). Risk frequency was expressed as one victim (that is, with property damaged, injured, or dead) per social population. An example was "*for every 10,000 persons per year, 1 person died in road traffic crashes*" and its risk frequency was 1 in 10,000. For Level 1 and Level 2, the eight risk frequencies ranged between 1 in 100 to 1 in 1000,000;

for Level 3, they ranged between 1 in 1000 to 1 in 10,000,000 (see https://osf.io/b9zu4/). Participants were asked to rate their acceptance according to one of four levels: *never accept*, *hard to accept*, *easy to accept*, and *fully accept*. In Part III, participants provided demographic information.

The only difference between the surveys conducted in China (Liu, Yang, et al., 2019a) and South Korea was that in the former study risk frequency was also expressed as one victim per vehicle-kilometers traveled. Similar results regarding the acceptable risk of SDVs versus HDVs were reported when risk frequency was expressed in the two different formats (that is, per social population and per vehicle-kilometers traveled) in China. The results for fatality risk in China were reported previously (Liu, Yang, et al., 2019a). In Study 1, the results for all three risk severities were analyzed.

Statistical analysis.

The risk acceptance rate of a risk scenario was defined as the percentage of participants choosing "*easy to accept*" or "*fully accept*." Logarithmic relationships between the risk acceptance rate and risk frequency (Huang et al., 2013) were examined. A regression analysis was conducted to examine the effects of risk frequency, vehicle type (HDV = 0, SDV = 1), and country (Korea = 0, China = 1) on the risk acceptance rate. The interaction effect between risk frequency and country was also examined. Risk frequency was log-transformed and then standardized to analyze the interaction effect. We predicted the acceptable risk frequency for specified risk acceptance rates using the R package *investr* (Greenwell, 2016) and then compared the acceptable risk frequencies for SDVs with those for HDVs and the real traffic fatality risk (see Figure 1b).

Results

All logarithmic regression models were significant (all p values < 0.001 and all R^2 values > .900; see Figure 2). Risk acceptance rate decreased with increasing risk frequency at all risk severity levels. Given that SDVs and HDVs had the same risk frequency, fewer participants accepted the risks associated with the former vehicles (see Table 1). The interaction effect between risk frequency and country was significant at Level 3 and marginally significant at the other two levels (Level 1: $\beta = 0.050$, $p = .056$; Level 2: $\beta = 0.056$, $p = .068$).

Table 2 shows the predicted acceptable risk frequencies for risk acceptance rates ranging from 30% to 80% (with an interval of 10%), and the corresponding ratio for HDVs versus SDVs in each case. This ratio represents the acceptable safety level of SDVs versus HDVs. For Level 1 (property damage only), the mean ratio is 3.5 for both Korea and China; for Level 2 (injury), it is 4.0 for Korea and 2.4 for China; for Level 3 (fatality), it is 5.5 for Korea and 4.7 for China. Remarkably, these ratio values are convergent for all three severity levels and in both countries. Concerning fatality risk, participants from both countries implicitly required SDVs to be 4–5 times as safe as HDVs. This finding was further supported in Study 4.

We also compared the predicted acceptable risk frequencies with the current traffic mortality rate. The acceptable risk frequency for half of Korean participants was 1.0E–5 per person for HDVs and 2.2E–6 for SDVs. The actual fatality risk for South Korea was 8.5E–5 (Korean National Police Agency, 2016), which is eight times (8.5E–5/1.0E–5) as high as that for HDVs and 38 times (8.5E–5/2.2E–6) as high as that for SDVs that was accepted by half of Korean participants. The acceptable risk frequency for half of

Chinese participants was 2.1E–6 for HDVs and 5.0E–7 for SDVs. The World Health Organization estimated that there are 18.8 deaths per 100,000 population in China (18.8E–5) (WHO, 2015), which is 89.5 times (18.8E–5/2.1E–6) as high as the risk frequency for HDVs and 376 times (18.8E–5/5.0E–7) as high as that for SDVs that was accepted by half of Chinese participants.

Summary

Human-machine interaction research suggests that people have a natural propensity to mindlessly apply social rules and expectations to machines (Nass & Moon, 2000), probably implying that we can use similar safety requirements for the regulation of both SDVs and HDVs. However, participants from South Korea and Chine wanted SDVs to be safer than HDVs, which is more in line with research on acceptable risk. Acceptable risk of technologies and behaviors is determined by their risk characteristics (e.g., uncontrollable and newness) (Fischhoff et al., 1978; Slovic, 1987; Slovic, Fischhoff, & Lichtenstein, 1979). A technology which is seen as new, uncontrollable or potentially catastrophic will be perceived as more risky and less acceptable (Otway & von Winterfeldt, 1982). From this perspective, considering self-driving as a new, less-controllable activity than human driving, people would require higher safety standards for SDVs. Like Chinese participants (Liu, Yang, et al., 2019a), South Korean participants also wanted self-driving to be 4–5 times as safe as human driving. This central finding from the cross-national survey is corroborated in Study 4.

The acceptable fatal risk for HDVs was lower than the actual mortality risk in both countries. The result for South Korea is similar to a finding that the risk for motor vehicles would need to be 5–8 times lower in order to be acceptable from the perspective

of lay people (Fischhoff et al., 1978; Fox-Glassman & Weber, 2016). To secure the benefits and utilities of vehicles, people have to tolerate traffic risks that are higher than their stated acceptable risks (Liu, Yang, et al., 2019a).

Study 2

Study 1 indicated that if SDVs have the same risk level (e.g., the same safety performance) as HDVs, they will be less likely to be accepted. Studies 2 and 3 were conducted to identify the factors that could determine the acceptable risk of SDVs and improve our understanding of the decision-making regarding acceptable risk. More specifically, Study 2 aimed to understand whether trust and negative affect mediate the relationship between vehicle type (HDV vs. SDV) and risk acceptance.

Method

Participants.

A total of 292 students (104 females) in a Chinese university were recruited through social media tools to participate in the vignette-based, online survey (https://www.sojump.com/) and randomly assigned to the HDV or SDV scenario. They were compensated with a small monetary reward ($0.3).

Material and procedure.

Participants were asked to read the following text, in which the phrases in brackets indicated the text on the SDV questionnaire (translated from Chinese): “Assume you are a taxi passenger and planning to take a taxi to your destination. A taxi driver [a self-driving car] has the average safety performance of all human drivers.” They then answered four questions, for which the scales were labeled “*very low*” on the left (= 1) and “*very high*” on the right (= 10). The two questions of fear and dread were: “If you are

riding in the taxi driver's car [the self-driving car], how much fear/dread do you feel?" They were averaged to measure negative affect (Cronbach's $\alpha = 0.94$) (Midden & Huijts, 2009).Trust was assessed by asking: "If you are riding in the taxi driver's car [the self-driving car], how much do you trust the driving performance of the taxi driver [the self-driving car]?" The question of risk acceptance (Siegrist & Sütterlin, 2014) was: "A trip by car could be risky. If you are riding in the driver's car [the self-driving car], how acceptable do you think the traffic risks are for this trip?" In the SDV questionnaire, before responded to the questions, participants were given a textual description of SDVs to demonstrate what SDVs are (Kyriakidis, Happee, & de Winter, 2015) and what non-driving activities can be performed in an SDV (see Appendix 2). Participants finally submitted demographic information (sex, age, and driving experience). Age was evaluated on four levels: "< 20" ($n = 66$), "20–25" ($n = 210$), "26–30" ($n = 11$), "> 30" ($n = 5$). Driving experience was evaluated on four levels: "0 year" ($n = 207$), "1–3 years" ($n = 77$), "4–7 years" ($n = 6$), "> 7 years" ($n = 2$). These levels describing age and driving experience were assigned values from 1 to 4 in statistical analysis.

Statistical analysis.

An analysis of covariance (ANCOVA), with negative affect, trust, and risk acceptance as dependent variables, vehicle type (HDV = 0, SDV = 1) as the independent variable, and gender, age, and driving experience as the covariates, examined whether participants reported different responses for the SDV and HDV with the same driving safety. We tested for mediation using bootstrapping procedures (Hayes, 2013) with 5,000 bootstrapped samples and Hayes's PROCESS Macro (model 4), with vehicle type as the

independent variable, trust and negative affect as the mediators, risk acceptance as the dependent variable, and gender, age, and driving experience as the covariates.

Results

Negative affect was correlated with trust ($r = –.36$, $p < .001$) but not with risk acceptance ($r = –.10$, $p = .083$). Trust was correlated with risk acceptance ($r = .43$, $p < .001$). According to a rule of thumb (Evans, 1996) for determining the correlation strength (< .20, very weak; .20–.39, weak; .40–.59, moderate; .60–.79, strong), their correlations were moderate or lower in magnitude. Figure 3a shows the estimated marginal means (EMM) of these responses. For the SDV, participants reported higher negative affect ($EMM_{HDV} = 3.96$, $EMM_{SDV} = 6.31$, $F(1,287) = 75.92$, $p < .001$, $\eta_p^2 = .21$), lower trust ($EMM_{HDV} = 6.38$, $EMM_{SDV} = 5.37$, $F(1,287) = 15.65$, $p < .001$, $\eta_p^2 = .05$), and lower risk acceptance ($EMM_{HDV} = 5.44$, $EMM_{SDV} = 4.71$, $F(1,287) = 6.74$, $p = .010$, $\eta_p^2 = .02$).

After entering the mediators, the path of vehicle type on risk acceptance became non-significant, $b = –0.49$, standard error (SE) = 0.29, $p = .094$, 95% confidence interval (CI) [–1.07, 0.08]. The indirect effect of vehicle type through negative affect had a point estimate of 0.23 and a bias-corrected 95% CI between –0.11 and 0.56 that included zero, while its indirect effect through trust had a point estimate of –0.48 and a bias-corrected 95% CI between –0.79 and –0.22 that did not include zero (see Figure 3b). Thus, instead of the higher negative affect evoked by SDVs, lower trust was the dominant factor accounting for lower risk acceptance for SDVs versus HDVs.

Summary

When an HDV and an SDV exhibited equivalent safety performance, participants expressed lower risk acceptance of the SDV, consistent with the finding of Study 1 that the acceptable risk frequency of SDVs was lower than that of HDVs. Thus, although previous studies (Awad et al., in press; Kohn, Quinn, Pak, de Visser, & Shaw, 2018) claimed that people would treat the risks of human drivers and machine drivers equally, our participants were less willing to accept the risks of machine drivers.

Study 3

Study 3 aimed to check the reproducibility of the results in Study 2 using a South Korean sample and to examine how vehicle type influences participants' willingness to ride (WTR) in the vehicle.

Method

Participants.

A total of 273 Korean students (103 females; M_{age} = 23.7, SD_{age} = 5.1; 142 with a driving license) participated in the offline survey (between-subjects design). They were recruited from a university through direct contact by interviewers and compensated with a small monetary reward ($1.5).

Material and procedure.

We measured negative affect ($\alpha = 0.95$), trust, and risk acceptance using the same items in Study 2. The term "self-driving car" in the questions in Study 2 was replaced by the term "self-driving taxi" in Study 3. We believed the latter term could make the used vignette more specific. Study 3 also measured participants' WTR through the question "Are you willing to ride in the driver's taxi [the self-driving taxi]?" on a scale of 1 to 10

(1 = *very low*; 10 = *very high*). Participants reported their gender, exact age, and whether they held a driving license.

Statistical analysis.

As for Study 2.

Results

Negative affect was negatively correlated with trust ($r = –.42$, $p < .001$), risk acceptance ($r = –.18$, $p = .003$), and WTR ($r = –.54$, $p < .001$). Trust was positively associated with risk acceptance ($r = .28$, $p < .001$) and WTR ($r = .52$, $p < .001$). Risk acceptance was positively correlated with WTR ($r = .41$, $p < .001$). These correlations were moderate or low in magnitude (Evans, 1996).

While responding to questions concerning riding in a self-driving taxi (vs. a human-driven taxi), participants reported higher negative affect ($EMM_{HDV} = 3.45$, $EMM_{SDV} = 6.44$, $F(1,268) = 136.87$, $p < .001$, $\eta_p^2 = .34$), lower trust ($EMM_{HDV} = 6.19$, $EMM_{SDV} = 5.14$, $F(1,268) = 15.57$, $p < .001$, $\eta_p^2 = .05$), lower risk acceptance ($EMM_{HDV} = 4.57$, $EMM_{SDV} = 4.01$, $F(1,268) = 4.96$, $p = .027$, $\eta_p^2 = .02$), and lower WTR ($EMM_{HDV} = 6.12$, $EMM_{SDV} = 4.64$, $F(1,268) = 27.02$, $p < .001$, $\eta_p^2 = . 09$) (see Figure 4a); these results replicated those in Study 2.

After entering the two mediators, the path of vehicle type on risk acceptance became non-significant, $b = –0.25$, $SE = 0.30$, $p = .405$, 95% CI [–0.85, 0.34]. The indirect effect of vehicle type through negative affect had a point estimate of –0.09 and a bias-corrected 95% CI between –0.52 and 0.33 that included zero, while its indirect effect through trust had a point estimate of –0.22 and a bias-corrected 95% CI between –0.44 and –0.06 that

did not include zero (see Figure 4b); these results replicated the findings of Study 2 (see Figure 3b).

After entering the two mediators, the path of vehicle type on WTR became non-significant, $b = 0.10$, $SE = 0.29$, $p = .726$, 95% CI [–0.47, 0.67]. The indirect effect of vehicle type through negative affect had a point estimate of –1.19 and a bias-corrected 95% CI between –1.66 and –0.79, while its indirect effect through trust had a point estimate of –0.39 and a bias-corrected 95% CI between –0.68 and –0.17 (see Figure 4c). Therefore, both negative affect and trust mediated the association between vehicle type and WTR. The non-overlapping 95% CIs of these two indirect effects indicated a greater indirect effect through negative affect.

Summary

The offline survey in South Korea in Study 3 replicated the online survey in China in Study 2 and produced similar findings indicating participants' lower risk acceptance of an SDV compared with an HDV with equivalent safety performance, partly because participants placed lower trust in the SDV (vs. HDV). Participants were less willing to use this SDV compared with its human counterpart, probably due to their lower trust in the SDV and higher negative affect evoked by it.

Study 4

The cross-national survey in Study 1 found that SDVs need to be 4–5 times as safe as HDVs. If this finding is robust, then people should have similar risk acceptance for an HDV as that of an SDV with 4–5 times the safety of the HDV. To check this, we conducted an additional survey in South Korea.

Method

The data of HDV in Study 3 was re-used. Thus, we only introduced the methodological issues in the survey involving a safer SDV.

Participants.

As in Study 3, a total of 120 Koreans (64 females; M_{age} = 20.5, SD_{age} = 1.7; 37 with a driving license) participated in the vignette-based, offline survey.

Material and procedure.

These are the same as those used in Study 3, with the only change being that the self-driving taxi was manipulated to exhibit 5 times the average safety performance of all human drivers (i.e., 5 times as safe as the human-driven taxi in Study 3).

Statistical analysis.

An ANCOVA was conducted similar to Study 3.

Results

Comparison between the human-driven taxi in Study 3 and this safer self-driving taxi in Study 4 showed that participants reported non-significant differences in terms of risk acceptance (EMM_{HDV} = 4.62, EMM_{SDV5} = 4.61, $F(1,252) = 0.002$, $p = .961$, $\eta_p^2 < .01$), WTR (EMM_{HDV} = 6.04, EMM_{SDV5} = 5.58, $F(1,252) = 1.29$, $p = .257$, $\eta_p^2 = .01$), and trust (EMM_{HDV} = 6.12, EMM_{SDV5} = 6.02, $F(1,252) = 0.12$, $p = .729$, $\eta_p^2 < .01$), but still reported higher negative affect towards the safer self-driving taxi (EMM_{HDV} = 3.51, EMM_{SDV5} = 5.36, $F(1,252) = 39.80$, $p < .001$, $\eta_p^2 = .14$) (see Figure 5).

Summary

When an SDV was manipulated to be superior to an HDV, i.e., 5 times as safe as the HDV, participants equally rated their risk acceptance of these two vehicles and

willingness to ride in these vehicles, supporting the central finding of Study 1 that SDVs need to be 4–5 times as safe as their human counterparts.

General Discussion

Tolerability of Risk for Self-Driving Vehicles

The public are skeptical about SDVs (Nielsen & Haustein, 2018; Stepp, 2017). No law or regulation policy has yet specified the required safety of SDVs. The United States Congress was reportedly considering legislation that would allow SDVs to be deployed on roads so long as they were deemed as safe as HDVs (Mervis, 2017). Our results suggest that such legislation would erode support for the technology from an already skeptical public.

Based on our findings, we propose a conceptual model of tolerability of risk for regulating and legislating SDVs. We distinguish three levels of risk: "*unacceptable*," "*tolerable*," and "*broadly acceptable*" (see Figure 6). We define the *unacceptable risk* criterion as SDVs not being safer than human drivers (Coelingh et al., 2018; Shladover & Nowakowski, 2019; Waycaster et al., 2018). Comparison between the acceptable risk frequencies of SDVs and HDVs in Study 1 indicated that both Korean and Chinese participants wanted SDVs to be 4–5 times as safe as HDVs (which was further confirmed in Study 4), which might be an indication of the boundary of the *tolerable* region. Furthermore, comparison between the risk frequency of SDVs acceptable to half of participants and the real traffic risk frequency indicated that SDVs are ideally needed to be roughly forty times and hundreds of times safer than the current road traffic for Korean and Chinese participants, respectively. Thus, we would like to refer the *broadly*

acceptable region to that SDVs are ideally dozens to hundreds of times safer than current road traffic.

Psychological Underpinning of Acceptable Risk

When the risk of SDVs and HDVs is specified as the same, the risk of SDVs is less acceptable than that of HDVs. This difference was assumed to be caused by participants' negative affect associated with these vehicles and their trust in them. According to the affect heuristic (Finucane et al., 2000; Slovic et al., 2004), people would consult their affective feelings associated with a technology to make risk and benefit judgments of the technology (Hadjichristidis, Geipel, & Savadori, 2015; Pachur, Hertwig, & Steinmann, 2012) and determine their willingness to accept the risks of the technology (Siegrist & Sütterlin, 2014). In Studies 2 and 3, a higher negative affect was indeed evoked when participants were asked to imagine themselves riding in an SDV (vs. HDV); however, higher negative affect did not account for participants' lower risk acceptance of SDVs, inconsistent with prior research on negative affect's role in reducing risk acceptance (Siegrist & Sütterlin, 2014). Instead, higher negative affect in part accounted for participants' lower willingness to ride in a self-driving car, in line with the affect heuristic.

Our results confirm the importance of public trust for facilitating the widespread adoption of SDVs (Hengstler, Enkel, & Duelli, 2016; Shariff, Bonnefon, & Rahwan, 2017). Lower trust in SDVs (vs. HDVs)—with the SDVs and HDVs manipulated to have the same safety performance—was found to motivate participants to not only be less willing to ride in an SDV but also be less willing to accept the risks associated with a trip in the SDV. In essence, trust in a technology represents a willingness to accept some risk

(Mayer, Davis, & Schoorman, 1995; Rousseau, Sitkin, Burt, & Camerer, 1998). Public distrust in the technology (Hutson, 2017; Stepp, 2017) is one of the biggest psychological obstacles to the widespread adoption of SDVs (Hutson, 2017; Shariff et al., 2017). It will also undermine people's willingness to accept their inevitable risks.

Several directions for mitigating the negative emotions toward SDVs (Hohenberger, Spörrle, & Welpe, 2016) and fostering public trust in SDVs (Hengstler et al., 2016; Shariff et al., 2017) have been suggested. Increasing the technical safety and reliability of SDVs, of course, can promote the public's trust and confidence in SDVs and diminish the public's negative impressions of SDVs, but it is not the only way that these goals can be accomplished. For example, improving the transparency in the decision-making processes associated with SDVs and communicating information about the benefits of SDVs will play important roles (Shariff et al., 2017). These non-technical strategies will be critical when the technical safety of SDVs cannot meet the public's acceptable level in their initial stages.

Policy Considerations

The proposed model of tolerability of risk (see Figure 6) describes the macroscopic safety requirement from a perspective of lay people, which implies that existing suggestions on legislation, e.g., allowing SDVs to be deployed on roads so long as they are deemed as safe as average human drivers (Mervis, 2017) or twice as safe as human drivers (Demers, 2018), will not gain sufficient public support. We furthermore discuss the introduction policy for SDVs. Policy researchers (Kalra & Groves, 2017) compared several introduction policies, including a less stringent policy (i.e., SDVs are required to be just 10% safer than average human drivers) and a more stringent policy (i.e., SDVs are

required to be 75% or 90% safer than average human drivers). Kalra and Groves predicted that more lives cumulatively saved under the less stringent policy could be as large as hundreds of thousands in the short term (over 15 years) to more than half a million in some cases in the long term (over 30 years). Thus, Kalra and Groves (2017) concluded that policymakers and regulators should allow SDVs on roads once their safety performance is better than that of average human drivers in a utilitarian society.

However, as Kalra and Groves (2017) also admitted, we do not live in a utilitarian society. For instance, Bonnefon et al. (2016) found a social dilemma that people hope others to buy utilitarian SDVs (i.e., SDVs that sacrifice their passengers for the greater good), but they themselves prefer to ride in selfish SDVs that protect passengers at all costs. This dilemma might also arise while discussing the acceptable safety of SDVs. It will be difficult to persuade people to entrust their safety to an SDV that only slightly safer than the average human driver or themselves. In addition, we should concern potential unexpected consequences of a less stringent policy. People's over-reactions to and lower tolerance for traffic crashes involving SDVs (e.g., they judge these crashes more severe and are less willing to accept them than equivalent crashes involving conventional cars) (Liu, Du, et al., 2019) suggests that a less stringent policy may backfire, because SDVs under this policy will still cause many crashes and fatalities, which may deter more people from adopting SDVs. Therefore, it seems to be unwise to hold a pure utilitarian standpoint in making acceptable risk decisions on SDVs.

According to our results, most people may not approve of a less stringent policy as proposed by Kalra and Groves (2017). The conceptual model of the tolerability of risk for SDVs in Figure 6 can offer insights for designing more appreciate introduction policies.

However, a more stringent policy—for instance, SDVs are required to be 4–5 times as safe as HDVs—could deprive SDVs of the opportunity to save many lives (Kalra & Groves, 2017). Such a policy would deny SDVs the real-world driving experience necessary to reach the safety level required by the public. A middle ground of the acceptable safety of SDVs might be needed for various stakeholders in the era of safer transportation. Thus, the conceptual model of the tolerability of risk for SDVs is not intended for immediate regulatory application.

Limitations

The four studies are not free from limitations. First, our samples in Study 1 are not representative; Studies 2–4 involved young participants, who might be more trusting of new technologies and more willing to accept risks than older people. We only considered participants from two Asian countries (i.e., South Korea and China), and thus we are uncertain whether our results from Study 1 could be applicable in other countries for the purpose of policy making. However, we are confident about the generalizability of the results because (1) the two central findings that SDVs are required to be 4–5 times as safe as HDVs and that lower trust in SDVs accounted for the lower acceptable risk of SDVs were reproduced and (2) the adoption of controlled between-subjects designs in these studies mitigated the potential negative impact of the unrepresentative samples. We look forward to the replication of these studies in other countries to provide specific empirical knowledge for policy making and regulation of SDVs in these countries. Second, people often struggle to grasp numerical concepts and lack sufficient numeracy (Garcia-Retamero & Cokely, 2013; Peters, 2012), which may make them difficult to make good judgments and decisions on the basis of very small numbers. Thus, their limitations in

numeracy (e.g., being biased in the interpretation of very small numbers) (Cohen, Ferrell, & Johnson, 2002) could impact the quantitative relationship between risk frequency and risk acceptance rate in Study 1 (Liu, Yang, et al., 2019a). Third, that people cannot yet experience a real SDV or other types of automated vehicles could influence their acceptable risk of SDVs. Further studies can provide participants with experience of self-driving and then survey their acceptable risk of SDVs.

Conclusions

Our participants implied that they wanted SDVs to be 4–5 times as safe as HDVs (see Studies 1 and 4). Their lower trust in SDVs, rather than the higher negative affect evoked by SDVs, explained their lower risk acceptance of SDVs (see Studies 2 and 3). The acceptable risk of SDVs will not only be determined by public attitude but also by the government and the transportation industry. Determining the level of acceptable risk is a political matter. We hope our observations can spur an effective public discourse among the many stakeholders to arrive at a consensus about the acceptable risk of SDVs.

References

Anderson, J. M., Kalra, N., Stanley, K. D., Sorensen, P., Samaras, C., & Oluwatola, O. A. (2016). *Autonomous Vehicle Technology: A Guide for Policymakers*. Santa Monica, CA: RAND Corporation.

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

Banerjee, S. S., Jha, S., Cyriac, J., Kalbarczyk, Z. T., & Iyer, R. K. (2018). Hands off the wheel in autonomous vehicles? A systems perspective on over a million miles of field data. In: 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Luxembourg City, Luxembourg.

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

Coelingh, E., Nilsson, J., & Buffum, J. (2018). Driving tests for self-driving cars. *IEEE Spectrum, 55*(3), 40–45. doi: 10.1109/MSPEC.2018.8302386

Cohen, D. J., Ferrell, J. M., & Johnson, N. (2002). What very small numbers mean. *Journal of Experimental Psychology: General, 131*(3), 424–442. doi: 10.1037//0096-3445.131.3.424

Demers, J. (2018). Self-driving cars will kill people and we need to accept that. Retrieved September 26, 2018, from https://thenextweb.com/contributors/2018/06/02/self-driving-cars-will-kill-people-heres-why-you-need-to-get-over-it/

Evans, J. D. (1996). *Straightforward Statistics for the Behavioral Sciences*. Pacific Grove, CA: Brooks/Cole Publishing.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice, 77*, 167–181. doi: 10.1016/j.tra.2015.04.003

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*(1), 1–17. doi: 10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L., & Keeney, R. (1981). *Acceptable Risk*. Cambridge: Cambridge University Press.

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences, 9*(2), 127–152. doi: 10.1007/BF00143739

Fox-Glassman, K. T., & Weber, E. U. (2016). What makes risk acceptable? Revisiting the 1978 psychological dimensions of perceptions of technological risks. *Journal of Mathematical Psychology, 75*, 157–169. doi: 10.1016/j.jmp.2016.05.003

Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science, 22*(5), 392–399. doi: 10.1177/0963721413491570

Greenwell, B. M. (2016). Package 'investr'. Retrieved December 28, 2016, from https://github.com/bgreenwell/investr

Hadjichristidis, C., Geipel, J., & Savadori, L. (2015). The effect of foreign language in judgments of risk and benefit: The role of affect. *Journal of Experimental Psychology: Applied, 21*(2), 117–129. doi: 10.1037/xap0000044

Halsey, A. (2017). How safe is 'safe enough' to put driverless cars on the nation's roadways? *The Washington Post*. Retrieved Auguest 1, 2018, from https://www.washingtonpost.com/local/trafficandcommuting/how-safe-is-safe-enough-to-put-driverless-cars-on-the-nations-roadways/2017/12/10/9a1aa348-d519-11e7-b62d-d9345ced896d_story.html?utm_term=.35c58fc84482

Hancock, P. A., Nourbakhsh, I., & Stewart, J. (2019). On the future of transportation in an era of automated and autonomous vehicles. *Proceedings of the National Academy of Sciences, 116*(16), 7684–7691. doi: 10.1073/pnas.1805770115

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. London: The Guilford Press.

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change, 105*, 105–120. doi: 10.1016/j.techfore.2015.12.014

Hohenberger, C., Spörrle, M., & Welpe, I. M. (2016). How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups. *Transportation Research Part A: Policy and Practice, 94*, 374–385. doi: 10.1016/j.tra.2016.09.022

Hook, L. (2017). For driverless cars, how safe is safe enough? *Financial Times*. Retrieved Auguest 1, 2018, from https://www.ft.com/content/70924ace-cf0d-11e7-b781-794ce08b24dc

Huang, L., Zhou, Y., Han, Y., Hammitt, J. K., Bi, J., & Liu, Y. (2013). Effect of the Fukushima nuclear accident on the risk perception of residents near a nuclear power plant in China.

Proceedings of the National Academy of Sciences, 110(49), 19742–19747. doi: 10.1073/pnas.1313825110

Hutson, M. (2017). People don't trust driverless cars. Researchers are trying to change that. *Science*. Retrieved January 16, 2018, from http://www.sciencemag.org/news/2017/12/people-don-t-trust-driverless-cars-researchers-are-trying-change

Kalra, N., & Groves, D. G. (2017). *The Enemy of Good: Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles*. Santa Monica, CA: RAND Corporation.

Kohn, S. C., Quinn, D., Pak, R., de Visser, E. J., & Shaw, T. H. (2018). Trust repair strategies with self-driving vehicles: An exploratory study. In: Proceedings of the Human Factors and Ergonomics Society 2018 Annual Meeting, Philadelphia, PA.

Korean National Police Agency. (2016). Traffic Accident Statistics. from http://www.police.go.kr/portal/main/contents.do?menuNo=200814#

Kyriakidis, M., Happee, R., & de Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour, 32*, 127–140. doi: 10.1016/j.trf.2015.04.014

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

Liu, P., Du, Y., & Xu, Z. (2019). Machines versus humans: People's biased responses to traffic accidents involving self-driving vehicles. *Accident Analysis & Prevention, 125*, 232–240. doi: 10.1016/j.aap.2019.02.012

Liu, P., Yang, R., & Xu, Z. (2019a). How safe is safe enough for self-driving vehicles? *Risk Analysis, 39*(2), 315–325. doi: 10.1111/risa.13116

Liu, P., Yang, R., & Xu, Z. (2019b). Public acceptance of fully automated driving: Effects of social trust and risk/benefit perceptions. *Risk Analysis, 39*(2), 326–341. doi: 10.1111/risa.13143

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

Mervis, J. (2017). Not so fast. *Science, 6369*, 1370–1374. doi: 10.1126/science.358.6369.1370

Midden, C. J. H., & Huijts, N. M. A. (2009). The role of trust in the affective evaluation of novel risks: The case of CO2 storage. *Risk Analysis, 29*(5), 743–751. doi: 10.1111/j.1539-6924.2009.01201.x

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. doi: 10.1111/0022-4537.00153

NHTSA. (2016). *Federal Automated Vehicles Policy: Accelerating the Next Revolution in Roadway Safety*. Washington, D.C.: National Highway Traffic Safety Administration (NHTSA), U.S. Department of Transportation.

Nielsen, T. A. S., & Haustein, S. (2018). On sceptics and enthusiasts: What are the expectations towards self-driving cars? *Transport Policy, 66*, 49–55. doi: 10.1016/j.tranpol.2018.03.004

Nunes, A., Reimer, B., & Coughlin, J. F. (2018). People must retain control of autonomous vehicles. *Nature, 556*, 169–171. doi: 10.1038/d41586-018-04158-5

Otway, H. J., & von Winterfeldt, D. (1982). Beyond acceptable risk: On the social acceptability of technologies. *Policy Sciences, 14*(3), 247–256. doi: 10.1007/BF00136399

Pachur, T., Hertwig, R., & Steinmann, F. (2012). How do people judge risks: Availability heuristic, affect heuristic, or both? *Journal of Experimental Psychology: Applied, 18*(3), 314–330. doi: 10.1037/a0028279

Penmetsa, P., Adanu, E. K., Wood, D., Wang, T., & Jones, S. L. (2019). Perceptions and expectations of autonomous vehicles – A snapshot of vulnerable road user opinion. *Technological Forecasting and Social Change, 143*, 9–13. doi: 10.1016/j.techfore.2019.02.010

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science, 21*(1), 31–35. doi: 10.1177/0963721411429960

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393–404. doi: 10.5465/amr.1998.926617

Schoettle, B., & Sivak, M. (2014). *Public Opinion about Self-Driving Vehicles in China, India, Japan, the U.S., the U.K., and Australia*. Ann Arbor, MI: Transportation Research Institute, University of Michigan.

Shankar, V., Mannering, F., & Barfield, W. (1996). Statistical analysis of accident severity on rural freeways. *Accident Analysis & Prevention, 28*(3), 391–401. doi: 10.1016/0001-4575(96)00009-7

Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2017). Psychological roadblocks to the adoption of self-driving vehicles. *Nature Human Behaviour, 1*(10), 694–696. doi: 10.1038/s41562-017-0202-6

Shladover, S. E., & Nowakowski, C. (2019). Regulatory challenges for road vehicle automation: Lessons from the California experience. *Transportation Research Part A: Policy and Practice, 122*, 125–133. doi: 10.1016/j.tra.2017.10.006

Siegrist, M., & Sütterlin, B. (2014). Human and nature-caused hazards: The affect heuristic causes biased decisions. *Risk Analysis, 34*(8), 1482–1494. doi: 10.1111/risa.12179

Siegrist, M., & Sütterlin, B. (2017). Importance of perceived naturalness for acceptance of food additives and cultured meat. *Appetite, 113*, 320–326. doi: 10.1016/j.appet.2017.03.019

Slovic, P. (1987). Perception of risk. *Science, 236*(4799), 280–285. doi: 10.1126/science.3563507

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis, 24*(2), 311–322. doi: 10.1111/j.0272-4332.2004.00433.x

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1979). Rating the risks. *Environment, 21*(3), 14–20, 36–39. doi: 10.1007/978-1-4899-2168-0_17

Smith, A., & Anderson, M. (2017). *Automation in Everyday Life*. Washington, DC: Pew Research Center.

Starr, C. (1969). Social benefit versus technological risk. *Science, 165*(3899), 1232–1238. doi: 10.1126/science.165.3899.1232

Stepp, E. (2017). Americans feel unsafe sharing the road with fully self-driving cars. *American Automobile Association*. Retrieved Auguest 1, 2018, from https://newsroom.aaa.com/2017/03/americans-feel-unsafe-sharing-road-fully-self-driving-cars/

Waldrop, M. (2015). Autonomous vehicles: No drivers required. *Nature, 518*, 20–23. doi: 10.1038/518020a

Waycaster, G. C., Matsumura, T., Bilotkach, V., Haftka, R. T., & Kim, N. H. (2018). Review of regulatory emphasis on transportation safety in the United States, 2002–2009: Public versus private modes. *Risk Analysis, 38*(5), 1085–1101. doi: 10.1111/risa.12693

WHO. (2015). *Global Status Report on Road Safety 2015*. Geneva, Switzerland: World Health Organization.

Xu, Z., Zhang, K., Min, H., Wang, Z., Zhao, X., & Liu, P. (2018). What drives people to accept automated vehicles? Findings from a field experiment. *Transportation Research Part C: Emerging Technologies, 95*, 320–334. doi: 10.1016/j.trc.2018.07.024

Appendix 1

Appendix Table 1

Demographic Information (Percentage) for Participants from South Korea (n_{HDV} = 272, n_{SDV} = 277) and China (n_{HDV} = 239, n_{SDV} = 260) in Study 1

Variables	South Korea		China		Variables	South Korea		China	
	HDV	SDV	HDV	SDV		HDV	SDV	HDV	SDV
Have heard of SDVs					*Education*				
Yes	77.9	72.2	90.0	91.5	Middle school and below	2.2	1.8	1.7	3.8
No	22.1	27.8	10.0	8.5	High school	25.8	22.1	14.6	10.0
Gender					Junior college	17.7	19.6	22.6	18.1
Female	50.2	53.1	43.5	46.5	Undergraduate	43.9	46.4	50.6	58.5
Male	49.8	46.9	56.5	53.5	Graduate	10.3	10.1	10.5	9.6
Age					*Occupation*				
≤ 20	2.6	4.0	8.4	6.9	Company employee	39.7	41.9	35.1	39.6

21−29	31.6	36.5	35.6	42.3	Civil servant	9.9	11.2	8.8	7.3
30−39	27.9	26.7	38.1	31.2	Public-sector employee	2.9	5.8	18.4	17.3
40−49	17.6	16.2	13.8	13.1	Self-employed	12.9	7.6	11.7	9.6
50−59	14.3	13.0	3.3	4.6	Retired	2.2	4.0	1.7	2.3
≥ 60	5.9	3.6	0.8	1.9	Student	11.0	11.9	15.5	16.5
Driver license holder					Other	21.3	17.7	8.8	7.3
Yes	71.3	80.1	75.3	68.8					
No	28.7	19.9	24.7	31.2					
Monthly income (KRW)					*Monthly income* (CNY)				
< 1,000,000	26.2	25.5			1,000−3,000			20.5	23.8
1,000,000~3,000,000	42.7	38.7			3,000−5,000			32.2	31.5
> 3,000,000	31.1	35.8			5,000−7,000			19.2	21.9
					7,000−10,000			18.4	16.5
					10,000−20,000			8.4	3.1
					> 20,000			1.3	3.1

HDV, human-driven vehicle; SDV, self-driving vehicle. KRW, Korea won, 1,000 KRW = $0.89. 1 CNY = $0.145. Demographic Information of Chinese participants has been shown previously (Liu, Yang, et al., 2019a).

Appendix 2

The textual description of SDVs in Studies 2-4:

The automated driving system takes over speed and steering control completely and permanently, on all roads and in all situations. The driver or passenger sets a destination via a touchscreen. The driver or passenger cannot drive manually or perform interventions because the vehicle has no steering wheel nor pedals. Self-driving vehicles allow drivers (passengers) to perform non-driving activities, such as reading a book, watching a film, surfing the Internet, playing with phones, dealing with work affairs, sleeping, and so on.

In Studies 3 & 4, Korean participants were also given the graphic picture of SDVs used in Study 1.

Table 1

A summary of the Regression Results in Study 1

Predictors	Risk severity					
	Level 1: Property damage only		Level 2: Injury		Level 3: Fatality	
	β	*SE*	*β*	*SE*	*β*	*SE*
Constant	0.670***	0.022	0.538***	0.027	0.458***	0.025
In(Risk frequency)[a]	–0.304***	0.018	–0.324**	0.022	–0.313***	0.021
Vehicle type (SDV = 1)	–0.116***	0.025	–0.109**	0.031	–0.116***	0.029
Country (China = 1)	–0.065*	0.025	–0.036	0.031	–0.089**	0.029
In(Risk frequency) × Country	0.050	0.025	0.059	0.031	0.096**	0.029
F(4,27)	130.3***		93.17***		91.37***	
R^2	.951		.932		.931	

[a]Risk frequency was log-transformed and then standardized to analyze the interaction effect. *β*: standardized coefficients; *SE*: standard error. HDV: human-driven vehicle; SDV: self-driving vehicle. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 2

Predicted Acceptable Risk Frequencies Given Risk Acceptance Rates

Severity	Risk acceptance rate	Predicted acceptable risk frequencies					
		Korea			China		
		HDVs	SDVs	Ratio[a]	HDVs	SDVs	Ratio[a]
Level 1: Property damage only	30%	2.3E–3	6.0E–4	3.8	2.1E–3	4.8E–4	4.4
	40%	8.4E–4	2.3E–4	3.7	6.3E–4	1.6E–4	4.0
	50%	3.1E–4	8.7E–5	3.6	1.9E–4	5.1E–5	3.6
	60%	1.2E–4	3.3E–5	3.5	5.5E–5	1.7E–5	3.3
	70%	4.3E–5	1.3E–5	3.4	1.6E–5	5.4E–6	3.0
	80%	1.6E–5	4.8E–6	3.3	4.7E–6	1.8E–6	2.7
	Mean			3.5			3.5
Level 2: Injury	30%	4.8E–4	1.7E–4	2.8	4.5E–4	1.8E–4	2.5
	40%	2.0E–4	6.3E–5	3.2	1.5E–4	5.9E–5	2.5
	50%	8.6E–5	2.4E–5	3.6	4.7E–5	1.9E–5	2.5
	60%	3.6E–5	8.8E–6	4.1	1.5E–5	6.3E–6	2.4
	70%	1.5E–5	3.3E–6	4.7	5.0E–6	2.1E–6	2.4
	80%	6.5E–6	1.2E–6	5.3	1.6E–6	6.8E–7	2.4
	Mean			4.0			2.4
Level 3: Fatality	30%	5.3E–5	1.8E–5	3.0	2.7E–5	8.1E–6	3.3
	40%	2.3E–5	6.3E–6	3.7	7.5E–6	2.0E–6	3.7
	50%	1.0E–5	2.2E–6	4.6	2.1E–6	5.0E–7	4.3
	60%	4.5E–6	7.9E–7	5.7	6.0E–7	1.2E–7	4.9
	70%	2.0E–6	2.8E–7	7.0	1.7E–7	3.0E–8	5.6
	80%	8.6E–7	9.8E–8	8.7	4.8E–8	7.5E–9	6.4
	Mean			5.5			4.7

[a]Ratio = Acceptable risk frequency for HDVs / Acceptable risk frequency for SDVs.

HDV: human-driven vehicle; SDV: self-driving vehicle.

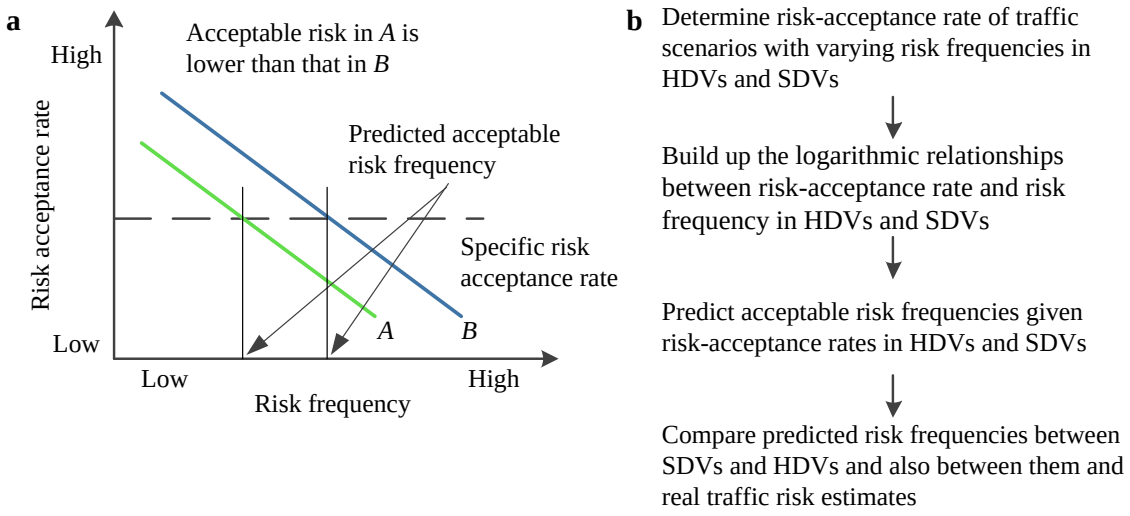

Figure 1. Expressed-preference approach for measuring the acceptable risks of SDVs

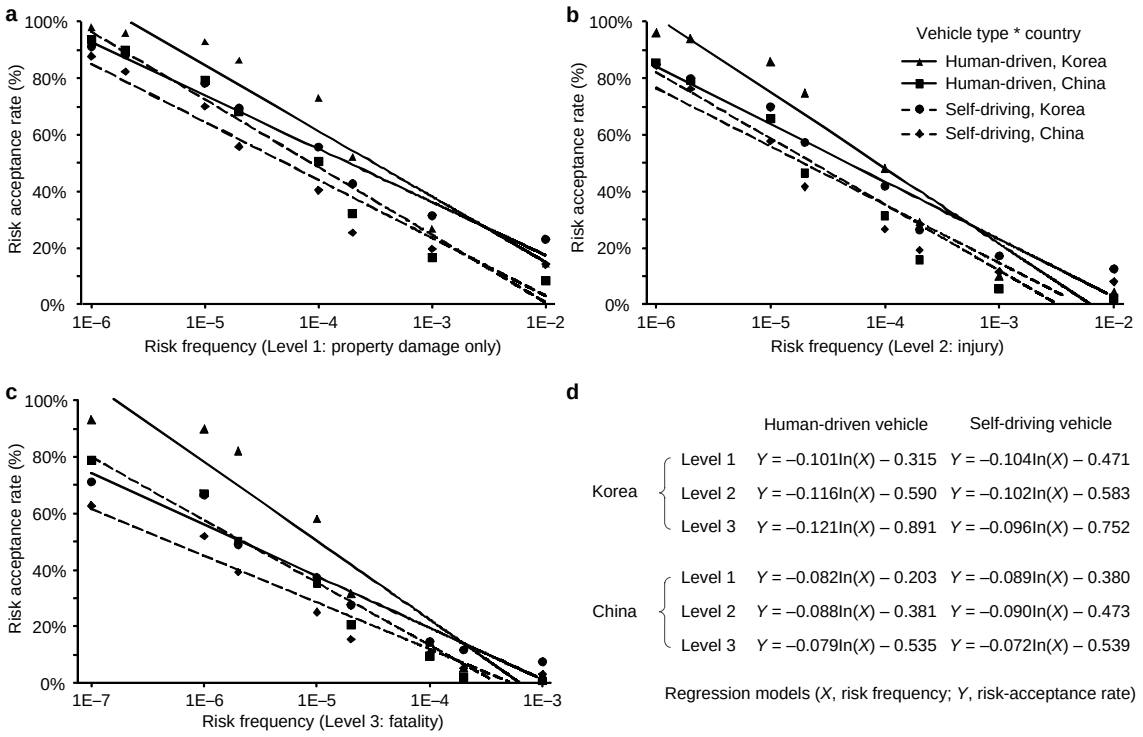

Figure 2. Relationships between risk frequency and risk acceptance rate in Korea (n_{HDV} = 272, n_{SDV} = 277) and China (n_{HDV} = 239, n_{SDV} = 260). (a) relationships for Level 1 (property damage only); (b) relationships for Level 2 (injury); (c) relationships for Level 3 (fatality) (data of Level 3 in China from Liu et al., 2019a); (d) summary of the regression models (all p values < 0.001, R^2 values > 0.900; X, risk frequency, and Y, risk acceptance rate). HDV: human-driven vehicle; SDV: self-driving vehicle.

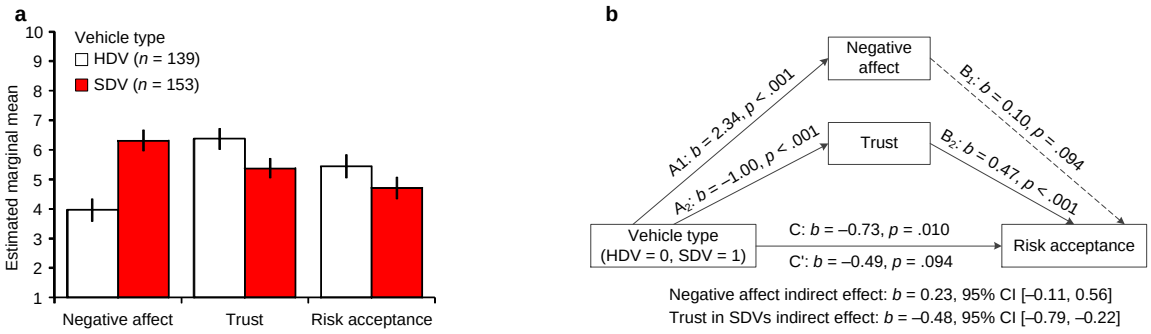

Figure 3. Results of estimated marginal means for the three responses (a) and the bootstrapped mediation analysis (b) in Study 2. Error bars indicate 95% CIs in (a). Non-standardized coefficients are shown in (b) and non-significant paths are shown as dotted lines. HDV: human-driven vehicle; SDV: self-driving vehicle.

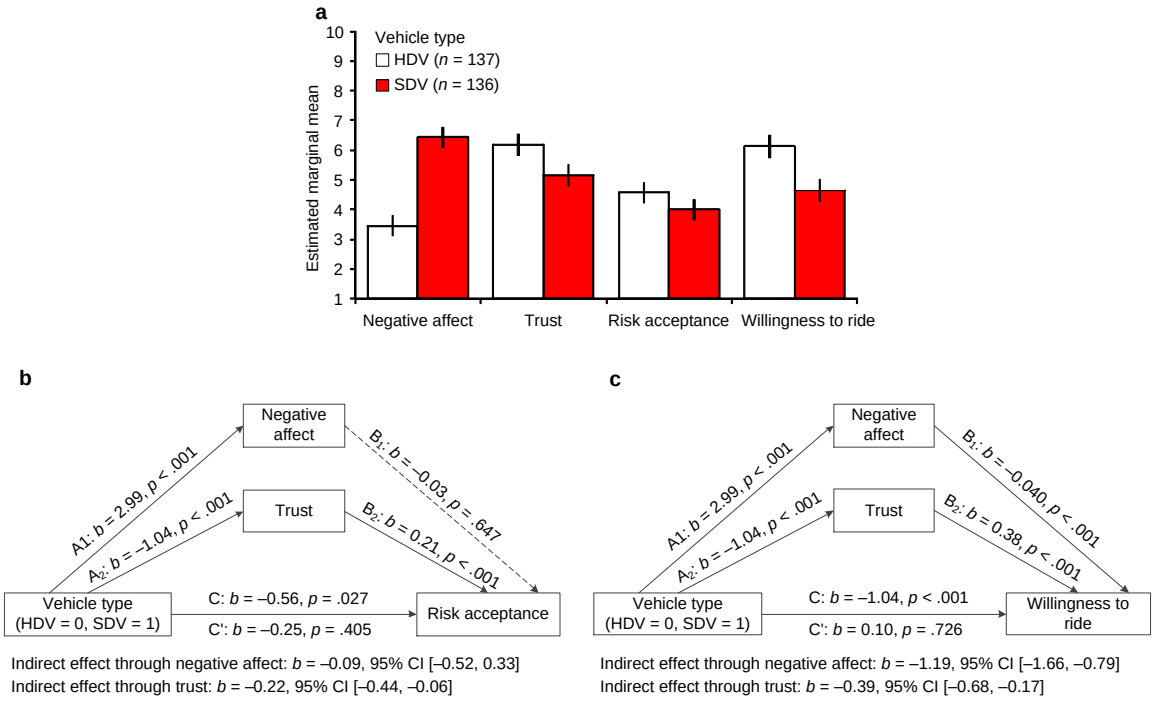

Figure 4. Results of estimated marginal means for the four responses (a) and the bootstrapped mediation analysis with risk acceptance (b) and willingness to ride (c) as the dependent variable in Study 3. Error bars indicate 95% CIs in (a). Non-standardized coefficients are shown in (b) and (c) and non-significant paths are shown as dotted lines. HDV: human-driven vehicle; SDV: self-driving vehicle.

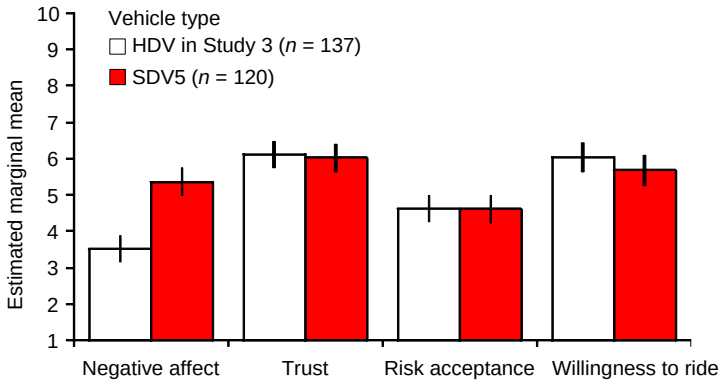

Figure 5. Results of estimated marginal means for the four responses in Study 4. Error bars indicate 95% CIs. Data of HDV is from Study 3. HDV: human-driven vehicle; SDV5: self-driving vehicle with 5 times the safety of average human drivers.

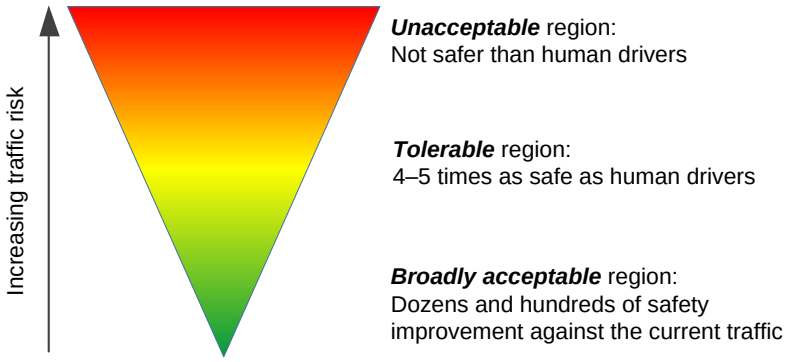

Figure 6. Conceptual model of tolerability of risk for self-driving vehicles