
International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

# Selfish or Utilitarian Automated Vehicles? Deontological Evaluation and Public Acceptance

Peng Liu & Jinting Liu

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
2021, VOL. 37, NO. 13, 1231–1242
https://doi.org/10.1080/10447318.2021.1876357



# Selfish or Utilitarian Automated Vehicles? Deontological Evaluation and Public Acceptance

Peng Liu and Jinting Liu

College of Management and Economics, Tianjin University, Tianjin, P.R. China

**ABSTRACT**

This research involves a controversial topic in the public sphere: should automated vehicles (AVs) be programmed with selfish algorithms to protect their passengers at all costs or utilitarian algorithms to minimize social loss in crashes involving moral dilemmas? Among a growing number of studies on what AVs should do in sacrificial dilemmas from the perspective of laypeople, few have considered how laypeople respond to AVs programmed with these crash algorithms. Our survey collected participants' deontological evaluation (i.e., evaluations of the moral righteousness of the decisions made by these AVs and of adopting these AVs), their perceived benefit and risk of these AVs, and their behavioral intention to use and willingness to pay (WTP) a premium for these AVs. The participants ($N$ = 580) perceived greater benefits from selfish AVs and reported a greater intention to use and higher WTP a premium for selfish AVs than utilitarian AVs. Deontological evaluation and perceived risk were non-significantly different between these AVs. Overall, selfish AVs were more acceptable to our participants. Deontological evaluation, perceived benefit, and perceived risk were predictive of behavioral intention. Additionally, after controlling for them, vehicle type still exerted a direct influence on behavioral intention. Perceived benefit was the dominant predictor of WTP a premium. Remarkably, participants expressed an insufficient intention to adopt both AVs, probably indicating that in regards to AV deployment, non-positive public attitudes toward AVs are more pressing than the challenge of deciding upon their ethical behaviors in rare moral dilemmas.

## 1. Introduction

Spatial mobility is a basic need for humans. However, the negative consequences of automobility are increasingly a concern in public debates. Our society struggles with traffic safety, traffic congestion, air and noise pollution, degradation of urban landscapes, and global warming due to greenhouse gas emissions (Banister, 2005). A potential solution to these issues is the mass adoption of autonomous vehicles (AVs) (Anderson et al., 2016; Bansal et al., 2016; Fagnant & Kockelman, 2015). As AVs will perform some or all of the tasks of fallible human drivers, who are complicit in more than 70% of traffic crashes (Dhillon, 2007; National Highway Traffic Safety Administration, 2013), AVs are expected to considerably reduce traffic crashes (National Highway Traffic Safety Administration, 2016). Vehicle automation is classified into six levels (Society of Automotive Engineers, 2018): no automation (Level 0), driver assistance (Level 1), partial automation (Level 2), conditional automation (Level 3), high automation (Level 4), and full automation (Level 5). Level 5 AVs are also called self-driving vehicles, fully automated vehicles, fully autonomous vehicles, or driverless vehicles; in them, all dynamic driving tasks are automated under all roadway and environmental conditions, without any human inputs.

Because of unforeseen outside circumstances (e.g., careless pedestrians, bad weather, wild animals) or unforeseen software and hardware faults, AVs cannot eliminate all traffic crashes and fatalities (Kalra & Paddock, 2016; Nyholm & Smids, 2016). Unlike their human counterparts, who base their traffic decisions in unwanted situations mostly on reflexes and instincts, AVs are expected to have the ability to form deliberate decisions. They can potentially estimate the outcomes of various options within milliseconds, taking many factors into account. Being considered implicit moral agents (Bonnefon et al., 2019), they are required to distribute harms in a socially acceptable manner. Therefore, how AVs are programmed to make moral decisions and how they behave in different scenarios where harms are highly likely or unavoidable, as well as what kinds of ethical and legal principles should be referred to when building the crash algorithms for guiding the moral decisions of AVs needs to be considered (Lin, 2016). These considerations raise important ethical and legal questions (Gless et al., 2016); for instance, should AVs be programmed to minimize the number of casualties, or to protect their passengers at all costs? Who is legally responsible for unavoidable harms caused by an AV (e.g., the manufacturer, the programmer, the passenger)?

Two kinds of crash algorithms guiding the ethical behaviors of AVs in moral dilemmas attract particular attention: *selfish algorithms*, which guide AVs to protect their passengers at all costs (even to the extent of having to sacrifice others),

**CONTACT** Peng Liu pengliu@tju.edu.cn College of Management and Economics, Tianjin University, Tianjin 300072, P.R. China

and *utilitarian algorithms*, which guide AVs to minimize all social losses (even to the extent of sacrificing their passengers to spare more road users). Coca-Vila (2018) called AVs with selfish algorithms "selfish AVs" (or "passenger-protective AVs") and those with utilitarian algorithms "utilitarian AVs." There is no consensus yet on what type of crash algorithms should be preferred, a dispute that might inhibit this technology's proliferation (Mordue et al., 2020). The question has become the elephant in the room, argued over incessantly by social scientists, philosophers, and ethicists (see Section 2).

As machine ethics should be aligned with human values, a growing number of studies has tried to understand laypeople's moral preferences regarding the decision-making of AVs in dilemmas (see Section 2). However, other types of responses to selfish and utilitarian AVs have been under investigated. For instance, how do people evaluate the respective benefits and risks of these AVs in society? Are they more or less willing to ride in one or the other? Are they willing to pay more for one or the other? Although the number of studies investigating public attitudes and responses to AVs is growing (see Becker & Axhausen, 2017; Choi & Ji, 2015; Elvik, 2020; Gkartzonikas & Gkritza, 2019; Hegner et al., 2019; Jing et al., 2020; Kim et al., 2020; Moody et al., 2020; Motamedi et al., 2020; Nordhoff et al., 2019; Panagiotopoulos & Dimitrakopoulos, 2018; Zhang et al., 2020), few of them have examined public responses to AVs programmed with different crash algorithms.

To address this gap, we surveyed public responses to selfish and utilitarian AVs, aiming to offer insights into choosing socially acceptable crash algorithms for AVs (not to recommend specific programming for AVs). As such, this paper is organized as follows. Section 2 reviews the moral preferences of laypeople regarding these AVs and addresses our research motivation. Section 3 explains methodological issues. Section 4 presents the results. Section 5 discusses the theoretical and practical implications of the results and the research weaknesses. The study concludes in Section 6.

## 2. Literature review and current research

### 2.1. Utilitarian AVs

The traffic scenarios that AVs might encounter have been likened to the trolley problem (Foot, 1967; Thomson, 1985) by some social scientists. In this thought experiment, the driver of a runaway trolley can only steer from one track to another. On the straight track ahead, five persons are working. Simultaneously, one person is working on another track. Anyone on the track the trolley driver enters is bound to be killed. Should the driver steer for the less occupied track? Most people who are asked this question answer, "Yes" (Thomson, 1985), indicating (from the layperson's perspective) that it is morally permissible to sacrifice one life to save more lives in impersonal situations.

The public's preference regarding the moral decisions AVs make in dilemmas has been discussed through different versions of the trolley problem. For instance, consider the case where an AV is driving at the correct speed along an inner-city road when five persons in the lane ahead start to cross the road rapidly, and the AV cannot stop in time to save them. The only way to save the group of five is to swerve into another lane where one person is walking and will be killed by the crash. In an alternative situation, the only way to save the five persons is to swerve and crash into an obstacle, though the AV's passenger will die. What should the AV do? Is it morally permissible to spare the five pedestrians over the single pedestrian or the passenger?

Most participants thought that AVs should make utilitarian decisions (Awad et al., 2018; Faulhaber et al., 2019; Bonnefon et al., 2016; Kallioinen et al., 2019; Li et al., 2019; Wintersberger et al., 2017) or that human drivers should make utilitarian decisions in similar situations (Bergmann et al., 2018; Kallioinen et al., 2019). Bonnefon et al. (2016) found that American participants overwhelmingly approved of utilitarian AVs that sacrifice their passengers or other road users for the greater good. Participants did not think that AVs should sacrifice one passenger when only one pedestrian could be saved; however, they did think that it was moral for AVs to sacrifice one passenger to save more pedestrians. Participants' approval of passenger sacrifice was even robust in treatments where the participant or another close person (a coworker or family member) was the passenger in the AV; the sacrifice's morality was still rated above the scale's midpoint. Furthermore, in their Moral Machine experiment, Awad et al. (2018) gathered 40 million decisions on AV choices in moral dilemmas in ten languages from 233 countries and territories. Their research noted several global moral preferences, including very strong preferences for sparing more lives, sparing humans over animals, and sparing young lives. Participants preferred a decision with the lesser evil.

Classic utilitarianism mandates sacrificing oneself. Bergmann et al. (2018) examined one-versus-more dilemmas in which participants imagining themselves as a human driver in a conventional car had to choose between driving into a group of people that suddenly appeared or driving toward a large chasm and risking their own life. Participants chose to save two persons in 52% of cases, three persons in 57% of cases, four persons in 63%, and five to seven persons in approximately 70% of cases. That those participants chose to save other people in more than 50% of cases seems to violate the self-preservation tendency. As Bergmann et al. did not examine the one-versus-one dilemma, whether participants were willing to sacrifice themselves to save a single person remained unknown.

### 2.2. Selfish AVs

Current studies have reported mixed results regarding moral decisions in dilemmas involving the sacrifice of passengers and pedestrians. The dominant outcome was that participants preferred to sacrifice passengers to preserve more road users (Awad et al., 2018; Kallioinen et al., 2019), even when they were the passenger (Bergmann et al., 2018; Frank et al., 2019). Awad et al. (2018) reported a preference for sparing pedestrians over passengers, although this preference was weak. As noted above, Bergmann et al. (2018) found that more than 50% of participants agreed to sacrifice themselves to save two or more other people in the human-driving context; as

Bergmann et al. (2018) did not examine the one-versus-one dilemma, this "self-sacrifice tendency" might be a result of a utilitarian attitude.

Frank et al. (2019) found a strong tendency for self-sacrifice. In their Study 1 and Study 2, they found that regardless of personal perspectives (that is, of whether participants acted as a passenger, pedestrian, or observer) and decision-making modes (deliberate, i.e., decision time is sufficient, vs. intuitive, i.e., decision time is limited), more than 50% of participants chose to sacrifice the passenger rather than an innocent pedestrian crossing the street. Frank et al. (2019) also tested participants' preferences in a dilemma with two passengers and a single innocent pedestrian crossing the street. They assumed that participants would sacrifice the single pedestrian in the crosswalk if they employed a utilitarian moral doctrine. However, their Study 3 showed that the prevalent choice remained in favor of saving the single pedestrian. For instance, from the passenger perspective, more than 70% of participants chose to sacrifice the two passengers when decision time was limited, and 50% of participants chose to sacrifice the two passengers when the decision time was sufficient. Thus, Frank et al. (2019) concluded that participants preferred a deontological moral doctrine "motivated by the reasoning that it is simply not right to sacrifice an innocent person in the street."

In contrast, in a virtual reality experiment, Kallioinen et al. (2019) found a self-preservation tendency when they manipulated two-versus-two dilemmas (2 adult occupants vs. 2 adult pedestrians) from three perspectives (passenger, observer, and pedestrian) involving two types of motorists (AVs and human drivers). The participants chose to stay on course (endangering pedestrians), or to veer toward a parked van (in the parked van trial) or a cliff edge (in the cliff trial), endangering the car's occupants in the latter two cases. In the parked van trial, most participants, regardless of their perspective, preferred to endanger the occupants (probably because veering toward the parked van is less dangerous). However, in the cliff trial, the occupants preferred the car to remain on course and endanger the pedestrians rather than veering toward a cliff edge, whereas, pedestrians preferred the opposite. It corresponds to a self-preservation effect for both occupants and pedestrians. This strong self-preservation tendency is not in line with the strong self-sacrifice tendency found by Frank et al. (2019). The observers also agreed with the pedestrians in this case.

### 2.3. Selfish or utilitarian AVs?

Whether to make AVs act in a utilitarian or a selfish manner is a challenging question. There seems to be strong agreement among laypeople that AVs should make moral decisions in a utilitarian manner (Awad et al., 2018; Faulhaber et al., 2019; Bonnefon et al., 2016; Kallioinen et al., 2019; Li et al., 2019; Wintersberger et al., 2017). Thus, Faulhaber et al. (2019) suggested implementing utilitarian ethics as a mandatory setting in future AVs. However, programming AVs with utilitarian ethics is unacceptable from the consumer and user perspectives (Bonnefon et al., 2016; Kallioinen et al., 2019), as well as in the context of certain ethical guidelines (Luetge, 2017) and legal requirements (Belay, 2015; Coca-Vila, 2018), as discussed below.

As utilitarian AVs will sacrifice the interests of AV consumers and users, they will not be accepted by those who pay for them. As Kallioinen et al. (2019) and Bonnefon et al. (2019) indicated, participants were unwilling to sacrifice or endanger themselves. Thus, another social dilemma emerges: while people would prefer that others buy utilitarian AVs, they are more willing to ride in selfish AVs (Bonnefon et al., 2016).

Ethical Guideline Number 9 by the German Ethics Commission on Automated and Connected Driving (Luetge, 2017) states that offsetting victims against one another is prohibited. In other words, the sacrifice of noninvolved parties to save more involved parties is not allowed. Accordingly, under this guideline, AVs should not be programmed to sacrifice the few to spare the many. The German Ethics Commission on Automated and Connected Driving also states that "it would be incompatible with the human dignity to expect people to sacrifice themselves in unavoidable dilemma situations" (see Bergmann et al., 2018, p. 7).

Similarly, from the perspective of current legal systems, which recognize humans as free agents with rights and responsibilities, maximizing social utility does not justify harmful interference into a person's legal sphere (Coca-Vila, 2018). As Belay (2015) claimed, "the legislature should determine that all self-driving cars must act in the interest of their passengers over anything else" (p. 129).

In addition, a few studies have begun to model the risks of AVs under different crash algorithms, offering insights for the selection of these algorithms. For instance, Mordue et al. (2020) modeled the risks related to AVs programmed on the basis of different ethical frameworks, where selfish AVs (i.e., AVs programmed based on ethical egoism) and utilitarian AVs were placed in varying traffic situations. In these models, the AV confronted three decisions (steering right and going into a ditch, resulting in the AV rolling over; proceeding straight, resulting in a rear-end collision with another vehicle; and steering left, resulting in a head-on collision with an oncoming vehicle). The situational variants were manipulated in terms of the number of occupants, the age of all occupants, the fatality rate by age for all occupants, and whether the occupants in any of the vehicles were parents. In Mordue et al.'s simulation, the utilitarian AV occupant dies at a frequency of at least 37% higher than that of occupants in a selfish AV, although utilitarian AVs lead to fewer deaths overall. Their results further manifest the extent of the challenge related to crash algorithms for policymakers, regulators, and automotive companies.

### 2.4. Current research

The large discrepancies among public preferences, ethical guidelines, and legal requirements multiply and complicate dilemmas regarding which kinds of crash algorithms should be programmed into AVs. We thus look at this question from another view. These crash algorithms, in principle, will ultimately be installed into AVs; therefore, to decide between them, we should know how the public evaluates and responds to the AVs installed with different crash algorithms specifically.

Our research focuses on the public's deontological and teleological evaluations regarding the acceptance of AVs equipped with utilitarian or selfish algorithms. Based on the General Theory of Marketing Ethics (Hunt & Vitell, 1986), deontological and teleological evaluations associated with an ethical object are assumed to predict public acceptance of this object. Deontological evaluations assess the inherent righteousness of this object (Chan et al., 2008). In the AV context, we evaluated the perceived righteousness of the AV's moral decision and the consumer's use of the AV, which is largely ignored in the AV literature. In contrast, teleological evaluations refer to the estimated goodness or badness of the consequences because of the ethical object. Perceived benefits and risks/costs of AVs are indicative of teleological evaluations. Finally, public acceptance can be measured by behavioral intention to use AVs or the willingness to pay (WTP) a premium for them (Liu et al., 2019a).

In summary, we surveyed five responses to selfish and utilitarian AVs: deontological evaluation, perceived benefit and risk (i.e., two teleological evaluations), behavioral intention, and WTP a premium. We examined whether participants would have different responses to these two AV types in a between-subjects survey and what kind of evaluations determined behavioral intention and WTP a premium. We expected that deontological evaluation and perceived benefit would positively predict behavioral intention and WTP a premium, while perceived risk would negatively predict them. We did not intend to formulate implementable ethical rules but instead suggest that these public responses deserve consideration when stakeholders of AVs decide on ethical AV behaviors. We adopted a between-subject design to collect these responses and expect it to overcome one limitation of current experimental ethics studies: participants facing a forced-choice with two alternatives were asked to indicate which choice is moral and their preference. This limitation might lead to biased results (Bigman & Gray, 2020; De Freitas et al., 2020a; De Freitas & Cikara, 2020).

We expect our work to make two contributions to the AV literature on AV ethics and public acceptance. First, the experimental ethics studies reviewed in Sections 2.1 and 2.2 mainly focused on moral preferences regarding AVs, and a very few of them (e.g., Bonnefon et al., 2016) have considered other limited public responses (e.g., behavioral intention and moral judgment) to AVs programmed with different ethical principles. To anticipate how the public will think about and treat these different AVs when they hit the market, we need to consider more public reactions (e.g., benefit and risk perceptions of these AVs and the willingness to pay a price premium). Second, few of the growing number of AV acceptance studies have investigated the public acceptance and attitude toward AVs programmed with different crash algorithms. Thus, our study fills this gap in AV acceptance literature. Notably, to the best of our knowledge, the influence of deontological evaluation on public acceptance of AVs is examined for the first time.

## 3. Methodology

### 3.1. Participants

Our between-subjects survey was conducted in Tianjin (a large metropolitan city in China) between August and October 2019. Participants were recruited through direct intercept at recreational areas and randomly assigned to one of the two AV groups. Of the sample of 605 Tianjin residents, 13 (2.1%) were eliminated because they answered almost all questions with the same answer in Part II of the questionnaire (see Section 3.3), while 12 (2.0%) were eliminated because they were observed not to read the description of AVs before they responded to Part II. The valid respondents' ($N$ = 580) demographic information is presented in Table 1. The expected numbers for each of the six age groups defined in Table 1 are 21, 163, 107, 106, 98, and 85, respectively, according to the 2010 Tianjin census data (Tianjin Municipal Bureau of Statistics, 2012). Compared between the actual and expected age distributions indicated that the sample was representative in terms of age ($\chi^2$ = 2.05, $p$ = .842). No significant differences were detected in terms of demographic information between the two groups in chi–squared tests ($ps$ > .05).

### 3.2. Procedure and measure design

This questionnaire consists of three parts. Part III required participants to submit their demographic information (see

**Table 1.** Demographic information.

| | Selfish AVs ($n$ = 295) | | Utilitarian AVs ($n$ = 285) | | Total ($N$ = 580) | |
|---|---|---|---|---|---|---|
| Variables | $n$ | % | $n$ | % | $n$ | % |
| *Gender* | | | | | | |
| Female | 130 | 44.1 | 129 | 45.3 | 259 | 44.7 |
| Male | 165 | 55.9 | 156 | 54.7 | 321 | 55.3 |
| *Age* | | | | | | |
| < 20 | 15 | 5.1 | 16 | 5.6 | 31 | 5.4 |
| 20 − 29 | 84 | 28.5 | 78 | 27.4 | 162 | 27.9 |
| 30 − 39 | 52 | 17.6 | 54 | 18.9 | 106 | 18.3 |
| 40 − 49 | 53 | 18.0 | 51 | 17.9 | 104 | 17.9 |
| 50 − 59 | 51 | 17.3 | 44 | 15.4 | 95 | 16.4 |
| ≥ 60 | 40 | 13.6 | 42 | 14.7 | 82 | 14.1 |
| *Education* | | | | | | |
| Middle school and below | 25 | 8.5 | 29 | 10.2 | 54 | 9.3 |
| High school | 50 | 16.9 | 51 | 17.9 | 101 | 17.4 |
| Junior college | 50 | 16.9 | 32 | 11.2 | 82 | 14.1 |
| Undergraduate | 87 | 29.5 | 98 | 34.4 | 185 | 31.9 |
| Graduate | 83 | 28.1 | 75 | 26.3 | 158 | 27.3 |
| *Monthly income (Chinese Yuan)* | | | | | | |
| < 3,000 | 75 | 25.4 | 86 | 30.2 | 161 | 27.8 |
| 3,000 − 5,000 | 69 | 23.4 | 58 | 20.4 | 127 | 21.9 |
| 5,000 − 7,000 | 58 | 19.7 | 52 | 18.2 | 110 | 19.0 |
| 7,000 − 10,000 | 47 | 15.9 | 46 | 16.1 | 93 | 16.0 |
| 10,000 − 20,000 | 35 | 11.9 | 34 | 11.9 | 69 | 11.9 |
| > 20,0000 | 11 | 3.7 | 9 | 3.2 | 20 | 3.4 |
| *Occupation* | | | | | | |
| Company employee | 64 | 21.7 | 76 | 26.7 | 140 | 24.1 |
| Civil servant | 15 | 5.1 | 12 | 4.2 | 27 | 4.7 |
| Public-sector employee | 65 | 22.0 | 50 | 17.5 | 115 | 19.8 |
| Self-employed | 18 | 6.1 | 20 | 7.0 | 38 | 6.6 |
| Retired | 52 | 17.6 | 45 | 15.8 | 97 | 16.7 |
| Student | 51 | 17.3 | 62 | 21.8 | 113 | 19.5 |
| Other | 30 | 10.2 | 20 | 7.0 | 50 | 8.6 |
| *Driving license holder* | | | | | | |
| Yes | 188 | 63.7 | 183 | 64.2 | 371 | 64.0 |
| No | 107 | 36.3 | 102 | 35.8 | 209 | 36.0 |

Percentages may not total 100% due to rounding. 1 Chinese Yuan ≈ 0.145 USD.

Table 1). Next we introduce the other two parts. In Part I, participants were instructed to read a short description of fully automated vehicles and the ethical issues and moral dilemmas relevant to AVs. Its description was adapted from Kyriakidis et al. (2015): "*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 by a touchscreen. The driver or passenger cannot drive manually and perform interventions because the vehicle does not have a steering wheel.*"

Considering that the ethical issues and moral dilemmas relevant to AVs are usually discussed in terms of life-or-death scenarios in public discourses and academic publications (e.g., see Section 2), we asked participants to read a narrative involving a life-or-death scenario to understand the moral dilemmas fully automated vehicles will face in the future and the potential ethical decisions involved:

Self-driving vehicles cannot guarantee the absolute safety of *passengers and other road users. Sometimes they may encounter moral dilemmas. Consider the following unexpected situation: a high-speed, self-driving vehicle is driving a passenger to the destination. Suddenly, three pedestrians are crossing the road in front of the self-driving vehicle. It is too late to brake. If the self-driving vehicle stays in the current lane, it will result in the death of the three pedestrians, but the passenger inside will not be harmed. If the self-driving vehicle swerves suddenly, it will hit the guardrail and cause the death of the passenger inside, but the three pedestrians will not be harmed.*

Then, participants read a text about the choice of fully automated vehicles. In the selfish AV scenario, participants read the following text: "*Assume the core of the algorithms for self-driving vehicles is to protect the benefit of their passengers. Therefore, in the abovementioned unexpected situation, the self-driving vehicle does not swerve to guarantee passenger safety.*" In the utilitarian AV scenario, the text was, "*Assume the core of the algorithms for self-driving vehicles is to achieve minimum social losses or maximum social benefits. Therefore, in the abovementioned unexpected situation, the self-driving vehicle swerves to minimize the number of deaths in this crash.*" The texts for the two AVs were adapted from Coca-Vila (2018).

In Part II, participants were asked to express their agreement or disagreement with items measuring behavioral intention, deontological evaluation, perceived benefit, perceived risk (see Table 2), and WTP a premium for their AV type. The items were assessed on a five-point Likert-type scale (1 = *totally disagree*, 5 = *totally agree*). Four behavioral intention items were adapted from previous works (Choi & Ji, 2015; Liu et al., 2019a). Six deontological evaluation items, adapted from Chan et al. (2008), measured the moral righteousness of the decisions made by these AVs and of adopting these AVs. Three perceived benefit items and three perceived risk items were modified from Liu and Xu (2020). To measure participants' WTP a premium for fully automated vehicles, participants were asked the question, "*Assume you will buy a new vehicle; how much extra money would you be willing to pay to purchase fully self-driving technology installed in your next vehicle?*" This question was adopted from previous studies (Kyriakidis et al., 2015; Schoettle & Sivak, 2014). A total of 12 alternatives were provided, from "Chinese Yuan 0" to "> Chinese Yuan 200,000" (Liu et al., 2019a).

**Table 2.** Psychological factors and their items (translated from Chinese).

| Factors | Item descriptions |
|---|---|
| Behavioral intention (BI) | BI1: I intend to use these self-driving vehicles in the future.<br>BI2: I intend to buy these self-driving vehicles in the future.<br>BI3: I will recommend that my family members and friends ride in these self-driving vehicles.<br>BI4: I hope that the government will promote these self-driving vehicles. |
| Deontological evaluation (DE) | DE1: This choice of self-driving vehicles is fair.<br>DE2: This choice of self-driving vehicles is morally right.<br>DE3: This choice of self-driving vehicles is just.<br>DE4: It is fair to use these self-driving vehicles.<br>DE5: It is morally right to use these self-driving vehicles.<br>DE6: It is just to use these self-driving vehicles. |
| Perceived benefit (PB) | PB1: Mass adoption of these self-driving vehicles will benefit our society.<br>PB2: Mass adoption of these self-driving vehicles will benefit my family and me.<br>PB3: Mass adoption of these self-driving vehicles will benefit road traffic safety. |
| Perceived risk (PR) | PR1: Mass adoption of these self-driving vehicles will threaten our society.<br>PR2: Mass adoption of these self-driving vehicles will threaten my family and me.<br>PR3: Mass adoption of these self-driving vehicles will threaten road traffic safety. |

## 4. Results

The raw materials, data, code, and results are available at the Open Science Framework (https://osf.io/yr78k/?view_only=5e05a98639a44c9aaeee6f3f0512e819).

### 4.1. Exploratory factor analysis

We examined the dimensional structure of the rating items for the four psychological factors in the exploratory factor analysis. We used parallel analysis (Horn, 1965) to determine the number of factors to extract. The "fa.parallel" function of the R package "psych" was used, which found four latent factors; thus, we used this number in the exploratory factor analysis. The Kaiser–Meyer–Olkin (KMO) factor adequacy index (KMO = .92) and Bartlett's sphericity tests ($\chi^2$ = 7671.94, $p$ < .001) supported the suitability of the data for exploratory factor analysis. Second, to determine the correct percentage of explained common variance in the item scores, we used a minimum rank exploratory factor analysis (Shapiro & Ten Berge, 2002). Except for the three perceived benefit items, the other items had low cross-loadings on other factors (< .40). Considering that the cross-loadings of the three perceived benefit items on another factor (i.e., behavioral intention) were equal or very close to the cutoff value of .40, we decided to keep these three items. As shown in Table 3, Factor I was interpreted as behavioral intention (Cronbach's α = .77); Factor II as deontological evaluation (α = .83); Factor III as perceived benefit (α = .74); and Factor IV as perceived risk (α = .84). All α values were

Table 3. Rotated factor patterns.

| Item | Factor patterns I | II | III | IV | Communalities ($h^2$) |
|---|---|---|---|---|---|
| BI1 | .82 | .26 | .24 | .21 | .84 |
| BI2 | .89 | .24 | .17 | .21 | .92 |
| BI3 | .77 | .26 | .25 | .25 | .79 |
| BI4 | .67 | .32 | .37 | .24 | .74 |
| DE1 | .18 | .73 | .13 | .12 | .60 |
| DE2 | .15 | .78 | .09 | .11 | .64 |
| DE3 | .17 | .79 | .07 | .16 | .68 |
| DE4 | .22 | .72 | .22 | .17 | .65 |
| DE5 | .21 | .74 | .25 | .10 | .67 |
| DE6 | .23 | .78 | .22 | .12 | .72 |
| PB1 | .40 | .31 | .59 | .30 | .69 |
| PB2 | .42 | .32 | .72 | .23 | .84 |
| PB3 | .40 | .30 | .61 | .31 | .71 |
| PR1 | −.20 | −.15 | −.14 | −.83 | .77 |
| PR2 | −.19 | −.17 | −.13 | −.88 | .85 |
| PR3 | −.25 | −.14 | −.22 | −.79 | .75 |

greater than .70, supporting internal consistency reliability (Kankanhalli et al., 2005). The four factors cumulatively explained 74% of the variance and were significantly correlated with each other (see Table 4).

## 4.2. Differences in psychological factors

The item scores were averaged to obtain the intended psychological factors. The mean values of the four factors associated with selfish and utilitarian AVs are presented in Figure 1. For selfish AVs, deontological evaluation (mean = 3.26, $t$ = 5.39, $p$ < .001) and perceived benefit (mean = 3.24, $t$ = 4.15, $p$ < .001) were greater than the scale middle of 3, while perceived risk was lower than 3 (mean = 2.75, $t$ = −4.46, $p$ < .001) and behavioral intention was not significantly different from 3 (mean = 3.11, $t$ = 1.97, $p$ = .0502). For utilitarian AVs, deontological evaluation was greater than 3 (mean = 3.16, $t$ = 3.35, $p$ = .001), while perceived risk (mean = 2.79, $t$ = −3.93, $p$ < .001) and behavioral intention (mean = 2.87, $t$ = −2.13, $p$ = .034) were lower than 3, and perceived benefit was not different from 3 (mean = 3.06, $t$ = 1.00, $p$ = .319).

Table 4. Mean and standard deviation (SD) of four psychological factors and their correlation matrix.

| Psychological factor | Mean | SD | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| 1. Behavioral intention | 2.99 | 1.02 | (.77) | | | |
| 2. Deontological evaluation | 3.21 | 0.82 | .56*** | (.83) | | |
| 3. Perceived benefit | 3.15 | 1.00 | .76*** | .61*** | (.74) | |
| 4. Perceived risk | 2.77 | 0.94 | −.53*** | −.39*** | −.57*** | (.84) |

Numbers in parentheses are Cronbach's α value. ***$p$ < .001 (two-sided).

Several ordinary least squares regression tests were conducted with the four psychological factors as dependent variables and with the vehicle type (selfish AV = 0, utilitarian AV =1) and six demographic factors as the predictors. In these tests and following logistic regression tests, the demographic factors played a role as covariates; thus, adding them will examine the robustness of the estimated effects of vehicle type (Lu & White, 2014). As the effects of these demographic factors are not our focus, we will not discuss them. Two regression models were built for behavioral intention: Model 1 (vehicle type and demographic factors as potential predictors) and Model 2 (vehicle type, demographic factors, and three psychological factors as potential predictors).

The regression tests revealed the differences in terms of perceived benefit and behavioral intention between selfish and utilitarian AVs (see Table 5): participants perceived less benefit from utilitarian AVs ($B$ = −0.17, $t$ = −2.12, $p$ = .035) and reported less intention to use utilitarian AVs (Model 1: $B$ = −0.24, $t$ = −2.89, $p$ = .004). Non-significant differences between these two AVs were detected in terms of deontological evaluation ($B$ = −0.10, $t$ = −1.50, $p$ = .133) or perceived risk ($B$ = 0.04, $t$ = 0.54, $p$ = .588).

In Model 2 for behavioral intention, vehicle type was still predictive of behavioral intention ($B$ = −0.11, $t$ = −2.10, $p$ = .036) after controlling for demographic and psychological predictors. Perceived benefit ($B$ = 0.61, $t$ = 15.75, $p$ < .001) and deontological evaluation ($B$ = 0.19, $t$ = 4.56, $p$ < .001) were positive predictors of behavioral intention, and perceived risk ($B$ = −0.14, $t$ = −3.91, $p$ < .001) was a negative predictor (see Model 2 for behavioral intention).

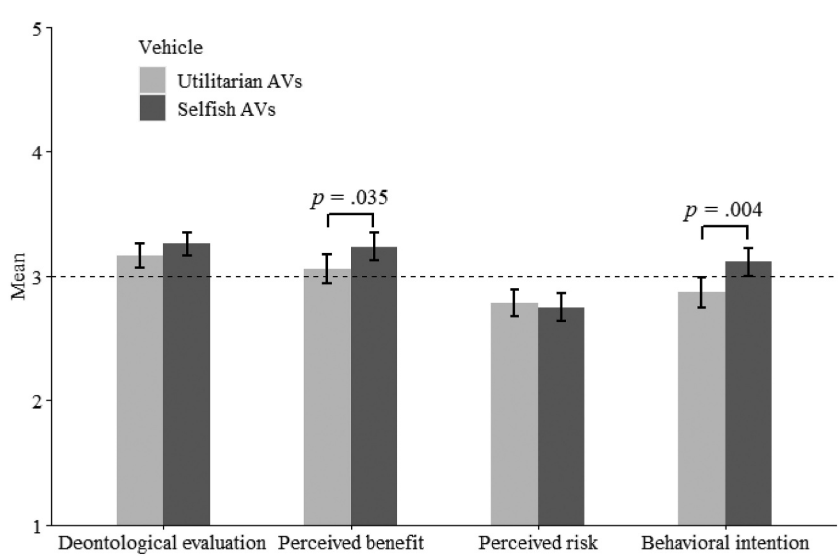


Figure 1. Psychological factors related to selfish and utilitarian AVs. Error bars = Mean ± 2 standard error.

**Table 5.** Results of ordinary least squares regression tests.

| | Deontological evaluation | | Perceived benefit | | Perceived risk | | Behavioral intention | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Model 1 | | Model 2 | |
| Predictor | *B* | *t* | *B* | *t* | *B* | *t* | *B* | *t* | *B* | *t* |
| Gender[a] | −0.12[†] | −1.68 | −0.31*** | −3.77 | 0.33*** | 4.12 | −0.29*** | −3.40 | −0.03 | −0.61 |
| Age[b] | 0.17* | 2.16 | 0.12 | 1.26 | −0.03 | −0.40 | 0.15 | 1.57 | 0.04 | 0.70 |
| Education[c] | 0.13 | 1.45 | 0.04 | 0.41 | −0.21* | −2.03 | 0.04 | 0.32 | −0.04 | −0.62 |
| Occupation[d] | −0.10 | −1.14 | −0.01 | −0.11 | 0.06 | 0.60 | −0.04 | −0.35 | 0.00 | −0.04 |
| Income[e] | −0.07 | −0.83 | −0.08 | −0.86 | 0.03 | 0.34 | −0.04 | −0.38 | 0.03 | 0.46 |
| Driver[f] | −0.17* | −2.04 | −0.21* | −2.13 | 0.09 | 0.91 | −0.10 | −0.98 | 0.07 | 1.10 |
| Vehicle[g] | −0.10 | −1.50 | −0.17* | −2.12 | 0.04 | 0.54 | −0.24** | −2.89 | −0.11* | −2.10 |
| Deontological evaluation | | | | | | | | | 0.19*** | 4.56 |
| Perceived benefit | | | | | | | | | 0.61*** | 15.75 |
| Perceived risk | | | | | | | | | −0.14*** | −3.91 |
| *F* | 2.32 | | 3.73 | | 3.21 | | 3.41 | | 89.31 | |
| *p* | .024 | | .001 | | .002 | | .001 | | < .001 | |
| $R^2$ | .028 | | .044 | | .038 | | .040 | | .611 | |

*B* = unstandardized coefficients.
[a]male = 0, female = 1.
[b] < 40 years old = 0, ≥ 40 years old = 1.
[c]undergraduate or graduate = 1, others = 0.
[d] < 7,000 Chinese Yuan per month = 0, others = 1.
[e]civil servant or public-sector employee = 1, others = 0.
[f]not driver = 0, driver = 1.
[g]selfish AV = 0, utilitarian AV =1.
[†]*p* < .10, **p* < .05, ***p* < .01, ****p* < .001.

### 4.3. Differences in WTP a premium

Figure 2 shows the distribution of participants' WTP a premium for the two fully automated vehicles. In line with previous work (Liu et al., 2019b), we roughly classified participants into three groups: participants unwilling to pay extra money for fully automated vehicles (33.2% for a selfish AV and 43.9% for a utilitarian AV), participants willing to pay (the equivalent of) a premium less than 2,900 USD (36.6% for a selfish AV and 30.9% for a utilitarian AV), and participants willing to pay a premium more than 2,900 USD (30.2% for a selfish AV and 25.3% for a utilitarian AV). The distribution of participants among the three groups was different between selfish and utilitarian AVs ($\chi^2_{(2,\ N\ =\ 580)} = 6.94$, $p = .031$).

As participants' WTP a premium was coded with three categories, we first conducted an ordinal logistic regression analysis to examine the influences of vehicle type, demographic factors, and psychological factors (deontological evaluation, perceived benefit, and perceived risk) on WTP a premium. One of its assumptions (i.e., parallel regression assumption or proportional odds assumption) is that the relationship between each pair of outcome groups is the same. For instance, the coefficients describing the relationship between the lowest versus all higher categories of the outcome variable are the same as those describing the relationship between the second-lowest versus all higher categories. If this assumption is met, only one set of coefficients (only one model) is needed. The Brant test (Brant, 1990) through the R package "brant" (Schlegel & Steenbergen, 2018) indicated the violation of the parallel regression assumption in the ordinal logistic model. Thus, we needed different models to describe the relationship between each pair of outcome groups. We ran two tests of binary logistic regression analysis with WTP a premium set at 0 as the reference group through the "glm" function in the R package "MASS" (Ripley et al., 2019). Two comparisons of WTP a premium were involved: Comparison 1 (WTP a premium of 0 vs. WTP a premium ≤ 2,900 USD) and Comparison 2 (WTP a premium of 0 vs. WTP a premium > 2,900 USD). For each comparison, two models were considered to demonstrate the influence of vehicle type and three psychological factors. As shown in Table 6, six involved demographic factors and vehicle type acted as predictors of WTP a premium in Model 1; three psychological factors were then added in Model 2.

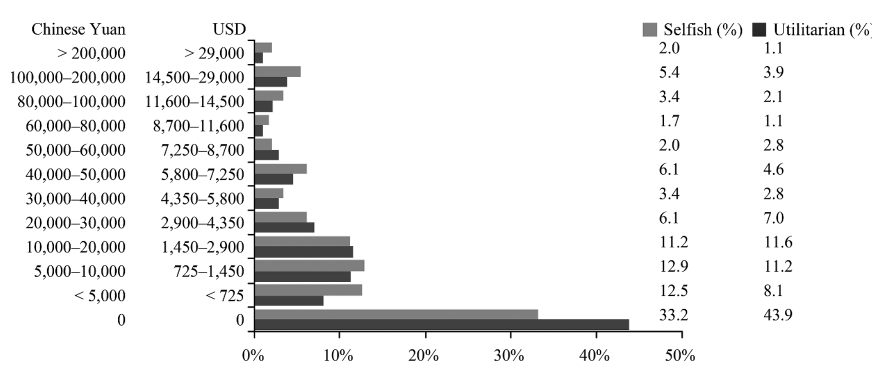


**Figure 2.** Participants' willingness to pay extra money for a selfish and for a utilitarian AV (1 Chinese Yuan ≈ 0.145 USD).

**Table 6.** Results of logistic regression analysis.

| | Comparison 1: WTP a premium of 0 (reference group) vs. WTP a premium ≤ 2,900 USD | | | | Comparison 2: WTP a premium of 0 (reference group) vs. WTP a premium > 2,900 USD | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 | | Model 2 | | Model 1 | | Model 2 | |
| Predictor | *B* | *z* | *B* | *z* | *B* | *z* | *B* | *z* |
| Gender[a] | −0.13 | −0.65 | 0.19 | 0.84 | −0.74*** | −3.37 | −0.54* | −2.13 |
| Age[b] | −0.23 | −0.99 | −0.43† | −1.70 | −0.18 | −0.72 | −0.32 | −1.12 |
| Education[c] | 0.02 | 0.06 | −0.08 | −0.29 | −0.26 | −0.92 | −0.27 | −0.84 |
| Occupation[d] | −0.05 | −0.19 | 0.10 | 0.36 | −0.06 | −0.22 | −0.30 | −0.95 |
| Income[e] | −0.21 | −0.84 | 0.01 | 0.03 | 0.46† | 1.82 | 0.69 | 2.36 |
| Driver[f] | 0.16 | 0.65 | 0.46† | 1.69 | 0.19 | 0.76 | 0.52† | 1.77 |
| Vehicle[g] | −0.46* | −2.32 | −0.37† | −1.71 | −0.43* | −2.00 | −0.33 | −1.34 |
| Deontological evaluation | | | −0.04 | −0.26 | | | −0.01 | −0.06 |
| Perceived benefit | | | 1.07*** | 6.08 | | | 1.26*** | 6.39 |
| Perceived risk | | | 0.04 | 0.28 | | | 0.00 | 0.01 |
| AIC[h] | 586.4 | | 515.3 | | 515.2 | | 428.3 | |

*B* = unstandardized coefficients.
[a]male = 0, female = 1.
[b] < 40 years old = 0, ≥ 40 years old = 1.
[c]undergraduate or graduate = 1, others = 0.
[d] < 7,000 Chinese Yuan per month = 0, others = 1.
[e]civil servant or public-sector employee = 1, others = 0.
[f]not driver = 0, driver = 1.
[g]selfish AV = 0, utilitarian AV =1.
[h]AIC, Akaike's information criterion (Akaike, 1973); lower AIC means greater goodness-of-fit of the model.
†$p$ < .10, *$p$ < .05, **$p$ < .01, ***$p$ < .001.

Several major findings emerged from the logistic regression analysis (see Table 6). First, vehicle type (selfish AV = 0, utilitarian AV = 1) was a negative predictor in Model 1 in both comparisons, even after controlling for the influences of demographic factors. This indicates that participants self-reported a lower WTP a premium for the utilitarian AV. After the inclusion of three psychological factors, vehicle type lost its influence in Model 2. Second, gender (male = 0, female = 1) was a negative predictor in Comparison 2, indicating that female participants were less willing to pay extra money for AVs. Third, perceived benefit was the only significant psychological predictor in both comparisons.

## 5. Discussion

### *5.1. Theoretical implications*

Deployment of AVs will encounter social and juristic challenges for which we are poorly prepared. One of them is the moral dilemmas encountered by AVs: in extreme situations, the machine driver has to determine whether to act in favor of its occupants or (a larger number of) others (Bonnefon et al., 2016). Sacrificing the few (even if they are the occupants of the AV) to save the many has been more often considered morally right in most ethics experiments (Awad et al., 2018; Faulhaber et al., 2019; Bonnefon et al., 2016; Kallioinen et al., 2019; Li et al., 2019; Wintersberger et al., 2017). Even when participants have been asked to answer from the passenger's perspective, they demonstrated an evident tendency for self-sacrifice in these moral dilemmas (Frank et al., 2019), which seems to run against a person's temptation to act in their own best self-interest. Therefore, certain researchers (Faulhaber et al., 2019) strongly support programming AVs in a utilitarian manner.

Contrary to the previous results, however, in our participants' evaluation, selfish and utilitarian AVs did not differ in moral righteousness (i.e., deontological evaluation). This difference between our results and the existing literature may be due to several factors. The dominant one, we believe, is related to the different levels of pressure that these participants faced during their judgments. In previous studies (e.g., Awad et al., 2018; Frank et al., 2019; Bonnefon et al., 2016), participants were forced to directly compare the two options (protecting the passenger at all costs vs. minimizing the number of casualties) and supposedly faced pressure to give the "socially desirable" response; thus, they had to show they more approved of the choice of minimizing the number of causalities. The direct comparison between the two options cannot reveal how participants evaluated each option's moral righteousness. Several researchers (Bigman & Gray, 2020; De Freitas et al., 2020a; De Freitas & Cikara, 2020) have made criticisms that participants would make biased responses when forced to choose one of the two options. In contrast, in our between-subjects design, participants only responded regarding one type of fully automated vehicles and thus, in principle, did not have this decision pressure. Another factor may be that our participants took the AV user's perspective and thus gave similar moral righteousness judgments regarding both AV types.

To the best of our knowledge, no previous studies have performed teleological evaluations on these AV types, that is, assigning goodness or badness to them. We performed teleological evaluations by collecting participants' benefit and risk perceptions for both AV types. Participants believed that selfish AVs create more benefits than utilitarian AVs and that they produce equal risks. Thus, participants had more positive teleological evaluations regarding selfish AVs.

As mentioned above, previous ethical studies (e.g., Awad et al., 2018; Bonnefon et al., 2016) indicated that participants' moral preferences could largely be described in a utilitarian manner. Accordingly, some of those studies (e.g., Faulhaber et al., 2019) have suggested implementing a utilitarian ethic setting for AVs. However, this might create a disincentive for individuals to adopt this technology. As Bonnefon et al. (2016) found, although participants thought selecting a utilitarian AV was more morally right, they would not ride in an AV if the car prioritized pedestrian safety over their own. There is a clear distinction between people's own self-interested choices as consumers and their ethical views as citizens in society. A similar finding was obtained in our study: participants expressed a greater intention to use selfish AVs than utilitarian AVs and were more willing to pay for the former.

Benefit perception (but not risk perception) was a steady predictor of behavioral intention and WTP a premium, which corroborates the importance of benefit perception for public acceptance of AVs (Liu et al., 2019a). Deontological evaluation was a steady predictor of behavioral intention but lost its predictive ability for WTP a premium in the presence of benefit perception. Previous ethical studies (e.g., Chan et al., 2008; Mayo & Marks, 1990) also indicated that teleological evaluation (as opposed to deontological evaluation) of a technology, behavior, or practice is more influential for the formation of acceptance of that technology, behavior, or practice.

### 5.2. Practical implications

Automakers who elected to prioritize occupant safety have been censured in the public sphere. A manager of active safety at Mercedes-Benz stated that the world's oldest automaker intends to program its AVs to save occupants inside the car preferentially and explained that, as the likelihood of saving occupants is higher compared to people outside the vehicle, priority should be given to occupants (Taylor, 2016). According to our results, this decision to weight occupant lives more heavily than pedestrian lives will make AVs more attractive to consumers. However, this statement evoked a public outcry, and the automaker ultimately purported that such statements were misquoted: "There is no instance in which we've made a decision in favor of vehicle occupants" (Orlove, 2016).

Public sentiments regarding crash algorithms or an AV's choice in the trolley problem would hinder AVs' technological advancement. Using the trolley problem in discussions of AV ethics has recently been criticized as useless. Himmelreich (2018) argued that the two assumptions of unavoidability and control advanced in the trolley problem (i.e., that harm is unavoidable and that the AV has a choice regarding the distribution of this unavoidable harm) are in tension with each other. For instance, if a crash is unavoidable, then there is no choice about distributing harms because the AV is already out of control, given AVs' engineering design and their failure modes. Thus, Himmelreich (2018) argued that trolley problems hardly represent situations that might occur. De Freitas et al. (2020a) similarly argued that trolley problems assume a fundamental paradox (e.g., an AV has the freedom to make a considered decision about whom to harm yet does not have enough control to take action to avoid harming anyone) and concluded that "trolley dilemmas are incredibly unlikely to occur on real roads." Nyholm (2018) also pointed out that people's intuitive reactions to an imagined moral dilemma offer limited value for programming AVs. Discussions on selfish and utilitarian algorithms might not offer insights for programming AVs in engineering design or related policymaking (Dewitt et al., 2019; De Freitas et al., 2020a), although their usefulness for AV design has been defended elsewhere (Keeling, 2020). Others (e.g., Coca-Vila, 2018) have suggested developing emergency algorithms for AVs from a legal and legislative perspective, which is out of our research scope.

However, our survey indicated a more pressing challenge. Participants' behavioral intention to use selfish AVs is very mild (mean = 3.11), and their intention to use utilitarian AVs is even lower (mean = 2.87); thus, participants showed insufficient acceptance of fully automated vehicles regardless of their type. Participants' deontological evaluations (slightly greater than 3) also indicate that the moral righteousness of both AV types was moderately acceptable. Public opinion surveys have previously indicated participants' ambivalent or negative attitudes or resistance toward AVs (e.g., Haboucha et al., 2017; Liu, 2020; Liu & Xu, 2020; Nielsen & Haustein, 2018). Notably, a growing number of longitudinal surveys show that public attitudes toward AVs are becoming more negative (AAA, 2018; Abraham et al., 2018; Hobbs, 2018; J.D. Power, 2017; Wang & Akar, 2019). We should, therefore, prevent AVs from becoming another controversial technology in society.

It may take a very long time (or be impossible) to instill in AVs the ability to detect and solve moral dilemmas (De Freitas et al., 2020a). Until that day comes, we should focus on whether people will accept AVs running on public roads or entrust their lives to AVs. People hold AVs to a higher safety standard than conventional human-driven cars (Liu et al., 2020, 2019c). AVs' safety benefits are only achieved after enough consumers purchase and adopt them. People will not buy AVs that do not protect them from any eventuality (Coca-Vila, 2018). For AVs to gain enough consumer support in the early stages, policymakers and manufacturers might need to prioritize occupant safety.

### 5.3. Research limitations and remarks

Several limitations and remarks are noted. First, the five types of responses to AVs identified might be formed from the users' self-interest. Our participants might implicitly act as future AVs users and thus give more favorable evaluations to selfish AVs that protect car occupants and express greater acceptance of these AVs in terms of behavioral intention and WTP a premium. Responses and perspectives from other road users (e.g., pedestrians and cyclists) should be considered. Second, due to difficulties in conducting a face-to-face survey and the great resource requirements it would entail, we surveyed participants from a single city in China, which could limit the generalizability of our findings. National differences in public attitudes toward AVs have been reported previously (Kyriakidis et al., 2015; Lang et al., 2016; Nordhoff et al., 2018; Schoettle & Sivak, 2014); in general, people from developed countries seem to be less positive toward this technology than those from developing countries (Liu et al., 2019b; Moody et al., 2020; Schoettle & Sivak, 2014). Awad et al. (2018) particularly reported cross-cultural variations in moral preference. Participants from individualistic cultures (e.g., North America) seemed to have a stronger preference for sparing the greater number of lives than those from collectivistic cultures (e.g., China), implying that culture might influence a person's attitudes toward AVs programmed as utilitarian or selfish. Thus, replications in other countries will be necessary and helpful. Third, considering that the public is educated on the ethical issues relevant to AVs usually through life-or-death situations similar to trolley cases (e.g., killing one to save five), we used a life-or-death scenario to describe ethical challenges that AVs could face. However, such life-or-death scenarios would be very rare. The ethics of AVs in mundane situations (e.g., pedestrian crosswalks or left-turns at intersections) also deserve attention (De Freitas et al., 2020b; Himmelreich, 2018) and have been largely ignored. Future studies may investigate how life-or-death scenarios influence participants' responses compared to mundane situations and whether AVs' different behaviors could influence participants' responses.

## 6. Conclusions

AV stakeholders struggle with a moral dilemma: should AVs be configured according to the principle of always prioritizing their occupants or minimizing social damage in traffic situations involving conflicts of interest? Regarding this dilemma, major discrepancies in public opinion, ethical guidelines, and legal requirements exist. We do not intend to recommend any specific programming for AVs but instead suggest that public responses to AVs programmed with different crash algorithms deserve consideration before programming any AVs. Our participants perceived more benefits from selfish AVs, showing a higher intention to use and greater willingness to pay extra money for these AVs, probably because they took passengers' perspective. They gave similar moral judgments of selfish and utilitarian AVs in terms of deontological evaluation. This finding is not in line with most previous studies, which have shown that laypeople view utilitarian AVs as more acceptable morally. For both types of AVs, their moral righteousness was moderately acceptable in our study. We also noted that even for selfish AVs, the participants only expressed a very limited intention to use them. This probably implies that building public acceptance is a more pressing issue than programming AVs with the ability to make ethical decisions in rare moral dilemmas, provided that AVs can achieve the purported benefits. Thus, prioritizing passenger safety might be needed for the diffusion of this technology in its early stages.

## Funding

This study was supported by the National Natural Science Foundation of China [Project no. 72071143].

## ORCID

Peng Liu http://orcid.org/0000-0003-4929-0531
Jinting Liu http://orcid.org/0000-0002-1910-2811

## References

AAA. (2018). *American trust in autonomous vehicles slips. American Automobile Association (AAA)*. Retrieved December 12 2018, from https://newsroom.aaa.com/2018/05/aaa-american-trust-autonomous-vehicles-slips/

Abraham, H., Reimer, B., Seppelt, B., Fitzgerald, C., Mehler, B., & Coughlin, J. F. (2018). Consumer interest in automation: Change over one year. In *Transportation research board (TRB) 97th annual meeting*, January 7–11, 2018. Washington D.C.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & B. F. Csaki (eds.), *Second international symposium on information theory*. Budapest: Academiai Kiado.

Anderson, J. M., Kalra, N., Stanley, K. D., Sorensen, P., Samaras, C., & Oluwatola, O. A. (2016). *Autonomous vehicle technology: A guide for policymakers*. RAND Corporation.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine Experiment. *Nature*, *563*, 59–64. https://doi.org/10.1038/s41586-018-0637-6

Banister, D. (2005). *Unsustainable transport: City transport in the new century*. Routledge.

Bansal, P., Kockelman, K. M., & Singh, A. (2016). Assessing public opinions of and interest in new vehicle technologies: An Austin perspective. *Transportation Research Part C: Emerging Technologies*, *67*, 1–14. https://doi.org/10.1016/j.trc.2016.01.019

Becker, F., & Axhausen, K. W. (2017). Literature review on surveys investigating the acceptance of automated vehicles. *Transportation*, *44*, 1293–1306. https://doi.org/10.1007/s11116-017-9808-9

Belay, N. (2015). Robot ethics and self-driving cars: How ethical determinations in software will require a new legal framework. *Journal of the Legal Profession*, *40*(1), 119–130. https://heinonline.org/HOL/LandingPage?handle=hein.journals/jlegpro40&div=9&id=&page=

Bergmann, L. T., Schlicht, L., Meixner, C., König, P., Pipa, G., Boshammer, S., & Stephan, A. (2018). Autonomous vehicles require socio-political acceptance—An empirical and philosophical perspective on the problem of moral decision making. *Frontiers in Behavioral Neuroscience*, *12*, Article 31. https://doi.org/10.3389/fnbeh.2018.00031

Bigman, Y. E., & Gray, K. (2020). Life and death decisions of autonomous vehicles. *Nature*, *579*, E1–E2. https://doi.org/10.1038/s41586-020-1987-4

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

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2019). The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars. *Proceedings of the IEEE*, *107*(3), 502–504. https://doi.org/10.1109/JPROC.2019.2897447

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, *46*(4), 1171–1178. https://doi.org/10.2307/2532457

Chan, R. Y. K., Wong, Y. H., & Leung, T. K. P. (2008). Applying ethical concepts to the study of "green" consumer behavior: An analysis of Chinese consumers' intentions to bring their own shopping bags. *Journal of Business Ethics*, *79*, 469–481. https://doi.org/10.1007/s10551-007-9410-8

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

Coca-Vila, I. (2018). Self-driving cars in dilemmatic situations: An approach based on the theory of justification in criminal law. *Criminal Law Philosophy*, *12*(1), 59–82. https://doi.org/10.1007/s11572-017-9411-3

De Freitas, J., Anthony, S. E., Censi, A., & Alvarez, G. A. (2020a). Doubting driverless dilemmas. *Perspectives on Psychological Science*, *15*(5), 1284–1288. https://doi.org/10.1177/17456916209222

De Freitas, J., Censi, A., Anthony, S. E., Di Lillo, L., & Frazzoli, E. (2020b). *From driverless dilemmas to more practical ethics tests for autonomous vehicles*. Retrieved September 19 2020, from https://psyarxiv.com/ypbve

De Freitas, J., & Cikara, M. (2020). *Deliberately prejudiced self-driving cars elicit the most outrage*. Retrieved October 4, 2020, from https://psyarxiv.com/2bxju

Dewitt, B., Fischhoff, B., & Sahlin, N.-E. (2019). Policy flaw in moral machine experiment. *Nature*, *567*, 31. https://doi.org/10.1038/d41586-019-00766-x

Dhillon, B. S. (2007). *Human reliability and error in transportation systems*. Springer.

Elvik, R. (2020). The demand for automated vehicles: A synthesis of willingness-to-pay surveys. *Economics of Transportation*, *23*, 100179. https://doi.org/10.1016/j.ecotra.2020.100179

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. https://doi.org/10.1016/j.tra.2015.04.003

Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., Stephan, A., Pipa, G., & König, P. (2019). Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles. *Science and Engineering Ethics*, *25*(2), 399–418. https://doi.org/10.1007/s11948-018-0020-x

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, *5*, 5–15.

Frank, D.-A., Chrysochou, P., Mitkidis, P., & Ariely, D. (2019). Human decision-making biases in the moral dilemmas of autonomous vehicles. *Scientific Reports*, *9*(1), Article 13080. https://doi.org/10.1038/s41598-019-49411-7

Gkartzonikas, C., & Gkritza, K. (2019). What have we learned? A review of stated preference and choice studies on autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *98*, 323–337. https://doi.org/10.1016/j.trc.2018.12.003

Gless, S., Silverman, E., & Weigend, T. (2016). If Robots cause harm, who is to blame? Self-driving cars and criminal liability. *New Criminal Law Review*, *19*(3), 412–436. https://doi.org/10.1525/nclr.2016.19.3.412

Haboucha, C. J., Ishaq, R., & Shiftan, Y. (2017). User preferences regarding autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *78*, 37–49. https://doi.org/10.1016/j.trc.2017.01.010

Hegner, S. M., Beldad, A. D., & Brunswick, G. J. (2019). In automatic we trust: Investigating the impact of trust, control, personality characteristics, and extrinsic and intrinsic motivations on the acceptance of autonomous vehicles. *International Journal of Human–Computer Interaction*, *35*(19), 1769–1780. https://doi.org/10.1080/10447318.2019.1572353

Himmelreich, J. (2018). Never mind the trolley: The ethics of autonomous vehicles in mundane situations. *Ethical Theory and Moral Practice*, *21*(3), 669–684. https://doi.org/10.1007/s10677-018-9896-4

Hobbs, A. (2018). *Survey reveals damaged public perception of autonomous cars*. Internet of Business. Retrieved April 18 2019, from https://internetofbusiness.com/survey-reveals-damaged-public-perception-of-autonomous-cars/

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/bf02289447

Hunt, S. D., & Vitell, S. (1986). A general theory of marketing ethics. *Journal of Macromarketing*, *6*(1), 5–16. https://doi.org/10.1177/027614678600600103

J. D. Power (2017). *Hands off? Not quite*. Consumers fear technology failures with autonomous vehicles. Retrieved October 29 2019, from https://www.jdpower.com/business/press-releases/jd-power-2017-us-tech-choice-study

Jing, P., Xu, G., Chen, Y., Shi, Y., & Zhan, F. (2020). The determinants behind the acceptance of autonomous vehicles: A systematic review. *Sustainability*, *12*(5), Article 1719. https://doi.org/10.3390/su12051719

Kallioinen, N., Pershina, M., Zeiser, J., Nezami, F. N., Pipa, G., Stephan, A., & König, P. (2019). Moral judgements on the actions of self-driving cars and human drivers in dilemma situations from different perspectives. *Frontiers in Psychology*, *10*, Article 2415. https://doi.org/10.3389/fpsyg.2019.02415

Kalra, N., & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice*, *94*, 182–193. https://doi.org/10.1016/j.tra.2016.09.010

Kankanhalli, A., Tan, B. C. Y., & Wei, -K.-K. (2005). Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly*, *29*(1), 113–143. https://doi.org/10.2307/25148670

Keeling, G. (2020). Why trolley problems matter for the ethics of automated vehicles. *Science and Engineering Ethics*, *26*, 293–307. https://doi.org/10.1007/s11948-019-00096-1

Kim, S., Chang, J. J. E., Park, H. H., Song, S. U., Cha, C. B., Kim, J. W., & Kang, N. (2020). Autonomous taxi service design and user experience. *International Journal of Human–Computer Interaction*, *36*(5), 429–448. https://doi.org/10.1080/10447318.2019.1653556

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. https://doi.org/10.1016/j.trf.2015.04.014

Lang, N., Rüssmann, M., Mei-Pochtler, A., Dauner, T., Komiya, S., Mosquet, X., & Doubara, X. (2016). *Self-driving vehicles, robo-taxis, and the urban mobility revolution*. The Boston Consulting Group and World Economic Forum.

Li, S., Zhang, J., Li, P., Wang, Y., & Wang, Q. (2019). Influencing factors of driving decision-making under the moral dilemma. *IEEE Access*, *7*, 104132–104142. https://doi.org/10.1109/ACCESS.2019.2932043

Lin, P. (2016). Why ethics matters for autonomous cars. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomous Driving: Technical, Legal and Social Aspects* (pp. 69–85). Springer. https://doi.org/10.1007/978-3-662-48847-8_4

Liu, H., Yang, R., Wang, L., & Liu, P. (2019a). Evaluating initial public acceptance of highly and fully autonomous vehicles. *International Journal of Human–Computer Interaction*, *35*(11), 919–931. https://doi.org/10.1080/10447318.2018.1561791

Liu, P. (2020). Positive, negative, ambivalent, or indifferent? Exploring the structure of public attitudes toward self-driving vehicles on public roads. *Transportation Research Part A: Policy and Practice*, *142*, 27–38. https://doi.org/10.1016/j.tra.2020.10.008

Liu, P., Guo, Q., Ren, F., Wang, L., & Xu, Z. (2019b). Willingness to pay for self-driving vehicles: Influences of demographic and psychological factors. *Transportation Research Part C: Emerging Technologies*, *100*, *100*, 306–317. https://doi.org/10.1016/j.trc.2019.01.022

Liu, P., Wang, L., & Vincent, C. (2020). Self-driving vehicles against human drivers: Equal safety is far from enough. *Journal of Experimental Psychology: Applied*, *26*(4), 692–704. https://doi.org/10.1037/xap0000267

Liu, P., & Xu, Z. (2020). Public attitude toward self-driving vehicles on public roads: Direct experience changed ambivalent people to be more positive. *Technological Forecasting and Social Change*, *151*, 119827. https://doi.org/10.1016/j.techfore.2019.119827

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

Lu, X., & White, H. (2014). Robustness checks and robustness tests in applied economics. *Journal of Econometrics*, *178*(Part 1), 194–206. https://doi.org/10.1016/j.jeconom.2013.08.016

Luetge, C. (2017). The German ethics code for automated and connected driving. *Philosophy & Technology*, *30*(4), 547–558. https://doi.org/10.1007/s13347-017-0284-0

Mayo, M. A., & Marks, L. (1990). An empirical investigation of a General Theory of Marketing Ethics. *Journal of the Academy of Marketing Science*, *18*(2), 163–171. https://doi.org/10.1007/bf02726432

Moody, J., Bailey, N., & Zhao, J. (2020). Public perceptions of autonomous vehicle safety: An international comparison. *Safety Science*, *121*, 634–650. https://doi.org/10.1016/j.ssci.2019.07.022

Mordue, G., Yeung, A., & Wu, F. (2020). The looming challenges of regulating high level autonomous vehicles. *Transportation Research Part A: Policy and Practice*, *132*, 174–187. https://doi.org/10.1016/j.tra.2019.11.007

Motamedi, S., Wang, P., Zhang, T., & Chan, C.-Y. (2020). Acceptance of full driving automation: Personally owned and shared-use concepts.

*Human Factors*, *62*(2), 288–309. https://doi.org/10.1177/0018720819870658

National Highway Traffic Safety Administration. (2013). *Preliminary statement of policy concerning automated vehicles*.

National Highway Traffic Safety Administration. (2016). *Federal automated vehicles policy: Accelerating the next revolution in roadway safety*.

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

Nordhoff, S., De Winter, J., Kyriakidis, M., van Arem, B., & Happee, R., 1. (2018). Acceptance of driverless vehicles: Results from a large cross-national questionnaire study. *Journal of Advanced Transportation*, *2018*, 5382192. https://doi.org/10.1155/2018/5382192

Nordhoff, S., Kyriakidis, M., van Arem, B., & Happee, R. (2019). A multi-level model on automated vehicle acceptance (MAVA): A review-based study. *Theoretical Issues in Ergonomics Science*, *20*(6), 682–710. https://doi.org/10.1080/1463922X.2019.1621406

Nyholm, S. (2018). The ethics of crashes with self-driving cars: A roadmap, I. *Philosophy Compass*, *13*(7), e12507. https://doi.org/10.1111/phc3.12507

Nyholm, S., & Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem? *Ethical Theory and Moral Practice*, *19*(5), 1275–1289. https://doi.org/10.1007/s10677-016-9745-2

Orlove, R. (2016). *Now Mercedes says its driverless cars won't run over pedestrians, that would be illegal*. Retrieved November 10 2019, from https://jalopnik.com/now-mercedes-says-its-driverless-cars-wont-run-over-ped-1787890432

Panagiotopoulos, I., & Dimitrakopoulos, G. (2018). An empirical investigation on consumers' intentions towards autonomous driving. *Transportation Research Part C: Emerging Technologies*, *95*, 773–784. https://doi.org/10.1016/j.trc.2018.08.013

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2019). *Package 'MASS'*. Retrieved November 13 2019 https://cran.r-project.org/web/packages/MASS/index.html

Schlegel, B., & Steenbergen, M. (2018). *Package 'brant'*. Retrieved November 13 2019, from https://benjaminschlegel.ch/r/brant/

Schoettle, B., & Sivak, M. (2014). *Public opinion about self-driving vehicles in China, India, Japan, the U.S., the U.K., and Australia*. UMTRI-2014-30. Transportation Research Institute, University of Michigan.

Shapiro, A., & Ten Berge, J. M. F. (2002). Statistical inference of minimum rank factor analysis. *Psychometrika*, *67*(1), 79–94. https://doi.org/10.1007/bf02294710

Society of Automotive Engineers. (2018). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems*. SAE J 3016-2018. SAE International.

Taylor, M. (2016). *Self-driving Mercedes-Benzes will prioritize occupant safety over pedestrians*. Retrieved November 17 2019, from https://www.caranddriver.com/news/a15344706/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395–1415. https://doi.org/10.2307/796133

Tianjin Municipal Bureau of Statistics. (2012). *Tabulation on the 2010 population census of Tianjin Municipality*. China Statistics Press (in Chinese).

Wang, K., & Akar, G. (2019). Factors affecting the adoption of autonomous vehicles for commute trips: An analysis with the 2015 and 2017 Puget Sound Travel Surveys. *Transportation Research Record*, *2673*(2), 13–25. https://doi.org/10.1177/0361198118822293

Wintersberger, P., Frison, A. K., & Riener, A. (2017). The experience of ethics: Evaluation of self harm risks in automated vehicles. In *IEEE intelligent vehicles symposium*, June 11–14, 2017. Los Angeles, CA. https://doi.org/10.1109/IVS.2017.7995749.

Zhang, T., Tao, D., Qu, X., Zhang, X., Zeng, J., Zhu, H., & Zhu, H. (2020). Automated vehicle acceptance in China: Social influence and initial trust are key determinants. *Transportation Research Part C: Emerging Technologies*, *112*, 220–233. https://doi.org/10.1016/j.trc.2020.01.027

## About the Authors

**Peng Liu** is an associate professor in the College of Management and Economics Department in Tianjin University in China. His research is focused on human factors and ergonomics, risk analysis, and risk perception related to complex systems such as autonomous driving and nuclear power plants.

**Jinting Liu** is a graduate student of the College of Management and Economics in Tianjin University in China. He graduated from the Beijing University of Chemical Technology in China. His research is focused on human reliability analysis and human factors in autonomous driving.