



# How virtue signalling makes us better: moral preferences with respect to autonomous vehicle type choices

Robin Kopecky<sup>1,2</sup> · Michaela Jirout Košová<sup>1,2</sup> · Daniel D. Novotný<sup>1,3</sup> · Jaroslav Flegr<sup>1,2</sup> · David Černý<sup>1,4</sup>

Received: 30 July 2021 / Accepted: 13 April 2022

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

## Abstract

One of the moral questions concerning autonomous vehicles (henceforth AVs) is the choice between types that differ in their built-in algorithms for dealing with rare situations of unavoidable lethal collision. It does not appear to be possible to avoid questions about how these algorithms should be designed. We present the results of our study of moral preferences ( $N=2769$ ) with respect to three types of AVs: (1) selfish, which protects the lives of passenger(s) over any number of bystanders; (2) altruistic, which minimizes the number of casualties, even if this leads to death of passenger(s); and (3) conservative, which abstains from interfering in such situations. We differentiate between scenarios in which participants are to make their decisions privately or publicly, and for themselves or for their offspring. We aim to answer two research questions: (1) whether the public visibility of the choice of an AV type choice make this choice more altruistic and (2) which type of situation makes it more difficult to choose altruistically: when choosing for society as a whole, when choosing only for oneself, or when choosing only for one's offspring. Our results show that respondents exhibit a preference for an altruistic strategy for AVs and that it is reinforced when signaled to others. The altruistic preference is strongest when applies to everybody else, weaker when it reflects a solely personal choice, and weakest when choosing for one's own child. We conclude that a public choice is considerably more likely to pressure consumers into accepting a more socially beneficial solution.

**Keywords** Moral dilemmas · Autonomous vehicles · Theory of signalization · AI ethics

## 1 Introduction

Autonomous vehicles are expected to significantly benefit our transportation systems, their safety, efficiency, and impact on the environment (Talebpour and Mahmasani 2016). Autonomous vehicles (henceforth AVs) also called self-driving cars or driverless cars are in this context defined as autonomous motor vehicles not operated on rails but with rubber tires for use on highways, with examples, such as cars, trucks and buses. However, many technical, social, legal, and ethical questions and challenges concerning AVs and their introduction to the mass market still remain (Shariff et al. 2017; de Sio 2017; Hulse et al. 2018; Epting 2018). Countries with developed automotive industries, such as the Czech Republic, need to face these questions to implement the right legal measures with respect to significant social changes that the introduction of AVs is likely to bring. One of the issues concerns the choice between different types of AVs that vary in their built-in algorithms for dealing with situations of unavoidable lethal collision (Contissa et al. 2017; Goodall 2014;

---

✉ Robin Kopecky  
robin.kopecky@natur.cuni.cz

Michaela Jirout Košová  
michaela.kosa@gmail.com

Daniel D. Novotný  
novotnyd@tf.jcu.cz

Jaroslav Flegr  
flegr@cesnet.cz

David Černý  
david.cerny@ilaw.cas.cz

<sup>1</sup> The Karel Čapek Center for Values in Science and Technology, Prague, Czechia

<sup>2</sup> Faculty of Science, Charles University, Vinicna 7, 128 00 Prague 2, Czechia

<sup>3</sup> Faculty of Theology, University of South Bohemia, České Budějovice, Czechia

<sup>4</sup> Institute of State and Law, Czech Academy of Sciences, Prague, Czechia

Haboucha et al. 2017). The reason why this particular issue is somewhat pressing has to do not so much with the fact that these situations will be happening on a regular basis. Rather, the issue is pressing from a psychological point of view: a non-negligible number of people still tend to be sceptical towards the reliability of new technologies and their safety, of which AVs are a clear example (Nielsen and Haustein 2018; Moody et al. 2020). Thus, the manufacturers will have to provide potential consumers with sensible solutions to collision situations, regardless of how common they will turn out to be in practice (Gogoll and Müller 2016; Gerdes and Thornton 2016).

Some of these situations will involve moral dilemmas inspired by trolley problems (Foot 2002; Thomson 1985; Kagan 1991; Lin 2015). In many ways, these dilemmas differ from the classical trolley problems: in the typical trolley cases, the person who decides about the solution of the situation is usually not the one to be harmed; in contrast, the AV cases concern the responsibilities of the person as a driver/traffic participant. Furthermore, AVs will not be limited to set tracks and will have to operate with probability. Perhaps other differences could be found but the question of the similarities and differences of the new AV dilemmas from old trolley problems is not directly relevant to our concerns. What needs to be highlighted is that moral intuitions that people have about these dilemmas do have bearing on the preference for a concrete AV algorithm and for willingness to purchase it. (Note also that we are not primarily concerned here with “risks or the legal and moral responsibilities we face in traffic”, unlike e.g. Nyholm and Smids (2016) who even criticize the relevancy of trolley cases for the AV literature).

There are at least three ways AVs can be set to handle the critical situations: (1) selfish AVs protect the lives of passenger(s) over any number of bystanders; (2) altruistic AVs minimize the number of casualties, even if this leads to death of passenger(s); and (3) conservative AVs abstain from interfering in such situations, even if it leads to the death of a higher number of subjects or death of passenger(s). (The latter two are both non-selfish strategies). In this paper, we present the results of our study of moral preferences with respect to these types of AVs. We furthermore differentiate between scenarios in which participants are to make their decisions privately or publicly, and for themselves or for their offspring. We disregard gender, age, health, biological species and other characteristics of (potential) casualties that can affect the preferences and decisions of respondents in our scenarios (Noothigattu et al. 2018).

Our study is based on the sample of 2769 mostly Czech volunteers (1799 women, 970 men; age IQR 25–32). The data come from our web-based questionnaire that was accessible from May 2017 to December 2017. We aim to answer the following two research questions: (1) Whether the public

visibility of an AV type choice makes this choice more altruistic and (2) which type of situation is more problematic with regard to the altruistic choice: opting for society as a whole, for oneself, or for one’s offspring.

Showing that the private or public character of AV type selection could affect the level of altruism has potentially important implications for policies concerning AV introduction. Findings concerning situations in which people choose most selfishly could be similarly significant. This could inform legislators where to proceed with caution to prevent moral panic which could lead to a strong rejection of AVs or severely delay the introduction of this life-saving technology.

The first research question is inspired by a biological theory of signalization (Laidre and Johnstone 2013; Dawkins and Krebs 1978; Zahavi 1975). Willingness to use AVs that would sacrifice the owner to save others could be perceived as a costly *honest signal*. This form of altruism is a signal of willingness to help others at one’s own cost (Hardy and Van Vugt 2006; Fehrler and Przepiorka 2013). This altruistic concern is related to the concern for one’s own reputation and it will pay off in social standing in the future in repeated social interactions (Nowak and Sigmund 1998), even with respect to strangers (Semmann et al. 2005). Participating in road traffic with people, one does not personally know and yet showing altruism by costly signalling, is analogous to altruism towards strangers (Bereczkei et al. 2010).

The owner of a virtue signalling AV would be exposed to social sanction (either positive or negative) by neighbours able to see which type of vehicle he or she has. Lee et al. (2018) experimentally investigated whether social observation has an impact on trolley problem-like decisions involving physical harm. The study showed that experimentally induced reputation concern reduces sacrificial decisions (called ‘utilitarianism’ in these contexts, since no self-sacrificing altruist was present in the situations). Brown and Sacco (2017) showed that people who prefer utilitarian decision in trolley problem cases seem less trustworthy to participants as long-term partners. Based on these findings, one could easily expect that reputation concern in a public choice will decrease choice of strategy that minimizes the number of casualties in an AV choice if the user is not present in their vehicle (which was not amongst the questions we asked our participants). However, in case the user is present in their vehicle the opposite expectation comes forward—the strategy that minimizes the number of casualties will be increased if a participant is allowing their vehicle to put their own life at stake and thus exhibit the aforementioned altruism towards strangers.

The second research question aims to compare our findings with those of Bonnefon et al. (2016) in different socio-cultural contexts. A part of Bonnefon’s study focused on the AV moral collision software type choice for a user when travelling alone or for a user when travelling with a work

colleague or when travelling with family. The participant was still hypothetically present in the AV type that was to be selected. In our study, we focused solely on the choice of AV type for the participant's beloved child to reduce the possible influence of their own presence and to compare directly between how participants weigh the safety of their child and their personal safety in a secret personal choice and the safety of everybody in a secret voting in Parliament.

## 2 Methods

The study is a part of the project “Factors Influencing Altruism” that was approved by the institutional review board at the Faculty of Science, Charles University (No. 2017/14).

### 2.1 Participants

We devised a web-based study running between May 2017 and December 2017 in which 2769 volunteers participated (1799 women, 970 men; age IQR 25–32). Most participants were college graduates and students [“Master and Doctor”: 1060 (38.3%), “Bachelor”: 349 (12.6%), “bachelor student”: 375 (13.5%)]. All participants understood the Czech language of the questionnaire. The majority of respondents were of Czech nationality [2436, (88%)], the second most represented nationality were Slovaks [267, (9.6%)].

### 2.2 Procedure

Participants were recruited either through our web page [www.PokusniKralici.cz](http://www.PokusniKralici.cz) or our Facebook fan page of the same name (Pokusní králíci) that has a frequent coverage by the media. The internet questionnaire was distributed as a Qualtrics survey. Participants were not incentivized by monetary reward. Rather, they were provided with results and comparison of sub-questionnaires directly after finishing the whole survey. After providing an informed consent, the volunteers filled out a comprehensive combined questionnaire composed of three sub-questionnaires focusing on various research topics, such as the aesthetic appreciation of plants and experimental philosophy of personal identity. It took around 90 min to complete the survey. There were also two attention-checking questions, one before and one after the AV section. Participants who failed to answer either of the questions correctly were excluded from data analyses (there were 45 failures on the first and 26 failures on the second checkpoint). By the end of 2017, the drop rate of the whole survey reached 38% of the 4443 participants who started this questionnaire. Only the participants who passed both attention checks were included in this study ( $N=2769$ ).

The AV section, which is the subject matter of the present study, was opened by the following brief information on AV:

“According to experts, AVs can prevent 95% of deadly car accidents, which means that the lives of roughly 600 people and 5.6 billion CZK (220 million EUR) may be saved every year in the Czech Republic alone.” Then participants were asked to evaluate nine trolley problem visualizations (Fig. 1), choosing between different software-conveyed solutions to the given lethal collision situation.

The reason why these nine visualizations were included, was to make the participants familiar with different trolley-like cases. This may differ with respect to the number of people inside the AV, the number of people on the road, and with respect to the position of the wreck of the van.

The last part of the 90 min questionnaire concerned AVs. We introduced three types of AVs in it: so-called “Tank” (with preference for passengers in the AV, here we refer to it as “selfish” AV), so-called “Calculator” (with preference for the highest number of saved lives, here referred to as “altruistic” AV), and so-called “Knight” (which will not actively change direction if that would lead to killing a pedestrian or a passenger, even though changing direction would save more lives in total, here referred to as “conservative” AV). Our paper draws solely on this last part of the questionnaire, in which the participants were asked the following five questions (paraphrase): (1) what type of AV would you choose as your own car if nobody was able to find out about your choice (“secret/self”)? (2) What type of AV would you choose as your own car if your choice was visible at first glance (“public/self”)? (3) What type of AV would you choose as the car of your beloved child if nobody was able

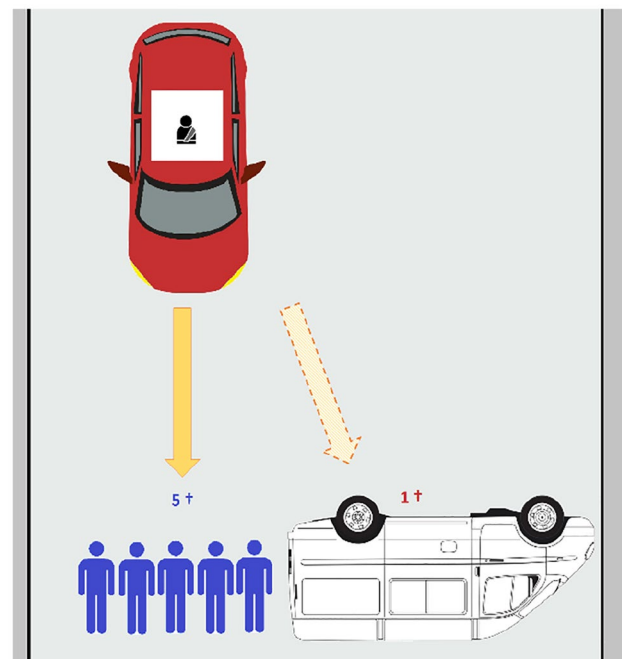


Fig. 1 An example of trolley problem visualization

to find out (“secret/child”)? (4) What type of AV would you vote for in a secret parliamentary vote if it was to become the only legal type of AV (“secret/parliament”)? (5) What type of software would you vote for in a public parliamentary vote if it was to become the only legal type of AV (“public/parliament”)? These five questions were consecutively asked on the same page of the questionnaire, so that participants were able to see them at a single glance.

### 2.3 Statistics

Statistical software IBM SPSS version 20 was used for the statistical tests, descriptive statistics, and producing graphs. Testing the association between two binary variables was done by contingency table using chi-square test. The statistical significance was characterized as *p* value and the effect size as risk ratio (RR). The associations with *p* value lower than 0.05 were considered statistically significant.

## 3 Results

### 3.1 Descriptive statistics

Summary descriptive statistics are displayed in Table 1 for following five questions: (1) What type of AV would you choose as your own car if nobody was able to find out about your choice (“secret/self”)? (2) What type of AV would you choose as your own car if your choice was visible at first glance (“public/self”)? (3) What type of AV would you choose as the car of your beloved child if nobody was able to find out (“secret/child”)? (4) What type of AV would you vote for in a secret parliamentary vote if it was to become the only legal type of AV (“secret/parliament”)? (5) What type of software would you vote for in a public parliamentary vote if it was to become the only legal type of AV (“public/parliament”)?

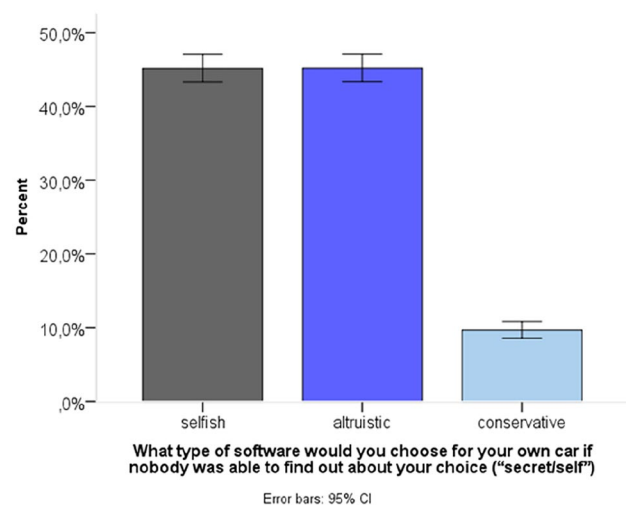
**Table 1** AV choices in different scenarios

|                    | Selfish      | Altruistic   | Conservative |
|--------------------|--------------|--------------|--------------|
| Secret, self       | 1219 (45.2%) | 1220 (45.2%) | 260 (9.6%)   |
| Public, self       | 809 (30%)    | 1565 (58.1%) | 319 (11.8%)  |
| Secret, child      | 1794 (66.6%) | 751 (27.9%)  | 150 (5.6%)   |
| Secret, parliament | 554 (20.6%)  | 1795 (66.9%) | 335 (12.5%)  |
| Public, parliament | 460 (17.1%)  | 1879 (69.9%) | 348 (13%)    |

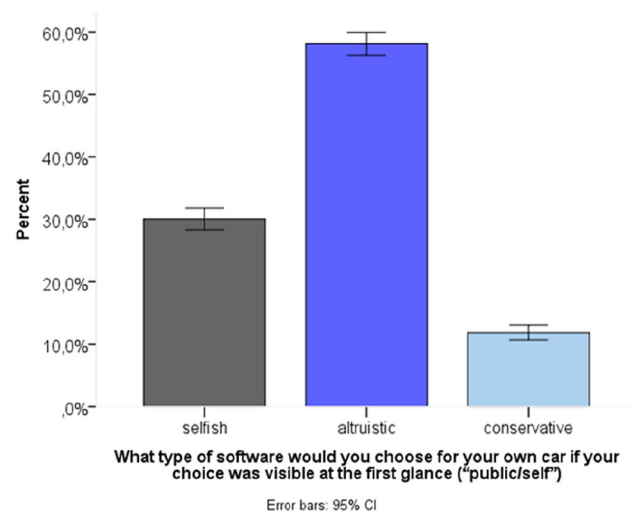
The rows represent different scenarios, while the columns represent choices. (We intended to include the scenario with a child only in the “secret” version as we aimed to answer the second research question in “secret” setting.)

### 3.2 Hypotheses 1.1

Via the first tested hypothesis “H1.1: secret/self > public/self” we tested whether participants are less likely to choose the selfish AV for themselves if the choice is indicated on the car and clearly visible. The distribution of choices is illustrated in Figs. 2 and 3. A chi-square test of independence was performed to examine the relation between the type of choice (secret; public) and the type of AV (selfish or non-selfish AVs, where the latter includes both altruistic and conservative AVs). The relation was significant, since



**Fig. 2** Secret/self. The y-axis shows the fraction of subjects selecting a particular variant of strategy. The error bars show the 95% confidence interval



**Fig. 3** Public/self. Y-axis shows the fraction of subjects selecting a particular variant of the strategy. The error bars show the 95% confidence interval

**Table 2** Change of opinion between secret and public choice

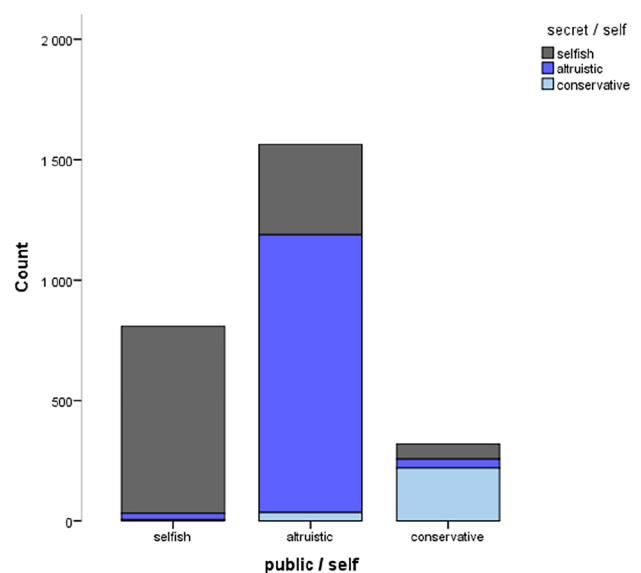
| public / self | N    | ← | secret / self | N    | %    |
|---------------|------|---|---------------|------|------|
| selfish       | 809  | } | selfish       | 777  | 96.0 |
|               |      |   | altruistic    | 27   | 3.3  |
|               |      |   | conservative  | 5    | 0.6  |
| altruistic    | 1564 | } | selfish       | 376  | 24.0 |
|               |      |   | altruistic    | 1153 | 73.7 |
|               |      |   | conservative  | 35   | 2.2  |
| conservative  | 319  | } | selfish       | 62   | 19.4 |
|               |      |   | altruistic    | 27   | 8.5  |
|               |      |   | conservative  | 220  | 69.0 |

On the left side there are participants divided into groups depending on their choice in the public/self-question. The right side describes their other decision in the secret/self-question so the change illustrated in the following Fig. 4 is completed with exact numbers

respondents expressed lower preferences for a selfish strategy in a public choice than in a secret one [ $\chi^2(1) = 131.40$ ,  $p < 0.00001$ ,  $RR = 0.665$ ,  $CI(95\%) = 0.620-0.714$ ]. Table 2 and the stacked bar chart in Fig. 4 shows the change in preference of participants between secret and public choice.

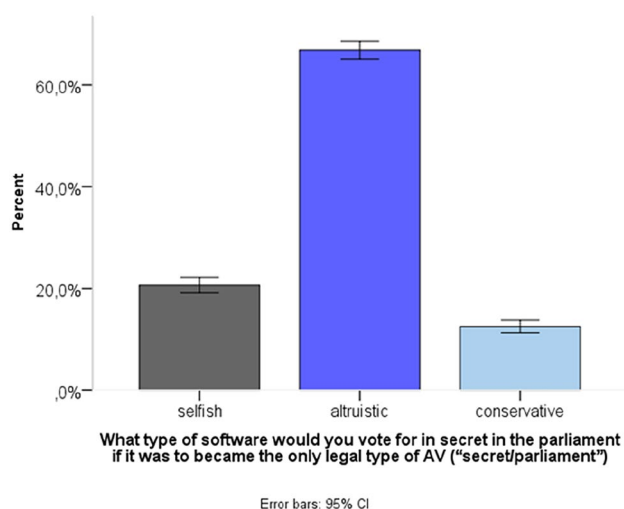
### 3.3 Hypothesis 1.2

Via the second hypothesis “H1.2: secret/parliament > public/parliament” we tested whether participants are less likely to vote for selfish software in Parliament if the voting is public. The distribution of choices is illustrated in Figs. 5 and 6. A chi-square test of independence was performed to examine the relation between the type of choice (secret/parliament; public/parliament) and the type of AV (selfish AV or non-selfish AV, where the latter includes altruistic and conservative AVs). The relation was significant, respondents expressed lower preferences for voting for a selfish strategy in a public vote than in a secret vote [ $\chi^2(1) = 11.519$ ,  $p = 0.0007$ ,  $RR = 0.825$ ,  $CI(95\%) = 0.738-0.922$ ]. Table 3 and the stacked bar chart in Fig. 7 show the change in preference of participants between secret and public voting.

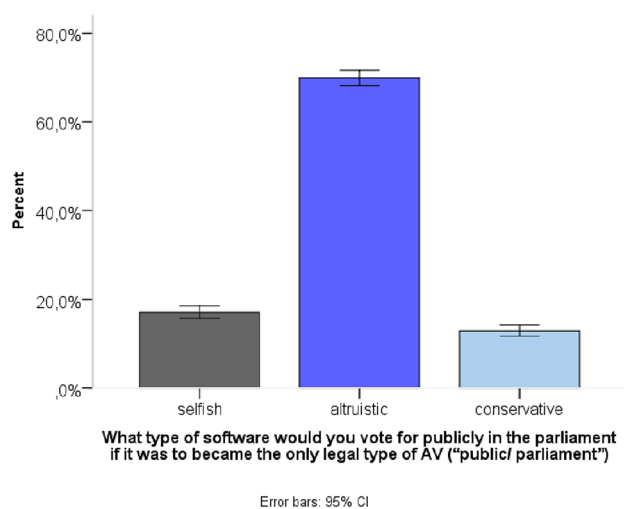


**Fig. 4** The change in preference of three strategies in a public choice by subjects who choose between various strategies in a secret choice. Columns show the preference in public/self-question and colours indicate what the same person has chosen in secret/self-scenario. E.g. the second column shows all people who preferred altruistic choice in public/self-scenario and charcoal-grey colour signifies a subgroup of them that preferred selfish strategy in secret/self-scenario





**Fig. 5** Secret/parliament. Y-axis shows the fraction of subjects selecting a particular variant of the strategy. The error bars show the 95% confidence interval



**Fig. 6** Public/parliament. Y-axis shows the fraction of subjects selecting a particular variant of the strategy. The error bars show the 95% confidence interval

### 3.4 Hypothesis 2

Via the last hypothesis “H2: secret choices: child > self; child > parliament” we tested whether participants prefer the most selfish strategy when deciding for their child, less selfish strategy when deciding for themselves and the least selfish strategy when deciding for everybody. The distribution of choices for their own child is illustrated in Fig. 8. A chi-square test of independence was performed to examine the relation between the type of choice (child; secret/self) and the type of AV (selfish and non-selfish). The relation

was significant when choosing a car for one's own child: people were more likely to choose selfish software than when choosing for themselves [ $\chi^2(1) = 250.53, p < 0.00001, RR = 0.678, CI(95\%) = 0.646-0.713$ ].

Similarly, when choosing a car in secret for one's own child people were more likely to choose selfish software than while voting in secret in parliament [ $\chi^2(1) = 1153.17, p < 0.00001, RR = 0.31, CI(95\%) = 0.2866-0.3355$ ].

## 4 Discussion

### 4.1 Empirical considerations

Our results showed that the majority of the respondents of the anonymous internet survey strongly preferred the altruistic strategy based on minimizing the number of people harmed even at the cost of self-sacrifice or based on an active sacrifice of a lesser number of other people. This preference was even stronger in cases of public choice when other people are able to see what type of strategy the AV owner has chosen, which shows that the type of choice indeed has an effect on altruism. The results also showed that the respondents reported much lower preferences for an altruistic strategy when choosing an AV for their child than when choosing for themselves. This result—the personal preference of an altruistic AV for themselves but not for their child—is consistent with the findings of Bonnefon et al. (2016).

Participants chose selfish software more often in cases when the choice of the strategy was private, i.e., not visible at first glance. This suggests that people regard selfish strategies as socially undesirable or morally dubious. Our results do not make it possible to decide on what grounds the group of participants shifted their choice from being selfish in secret to being altruistic in public. Nevertheless, it can be expected that building one's reputation could play a role. Results of experimental games indicate that reputation is the key factor to maintain altruism in both reciprocal and indirect forms (Kraft-Todd et al. 2018; Semmann et al. 2005; Olivella and Siciliani 2017). Players usually act more altruistically towards altruistic individuals regardless of whether they have helped them directly or not. Signalling an altruistic strategy could also be advantageous in *intra-sexual competition*, i.e. biological concept in which two members of the same sex are competing against each other over a potential mate (Zahavi 1975; Buunk and Massar 2012; Puska et al. 2016).

If reputation building is the crucial motive for signalling the use of an altruistic strategy, then this signal can potentially be falsified, i.e., some AV users, the cheaters, will be signalling an altruistic strategy but in fact they will be using

**Table 3** Changed opinion between secret and public vote

| public / parliament | N    | ← | secret / parliament | N    | %    |
|---------------------|------|---|---------------------|------|------|
| selfish             | 460  | } | selfish             | 448  | 97.4 |
|                     |      |   | altruistic          | 11   | 2.4  |
|                     |      |   | conservative        | 1    | 0.2  |
| altruistic          | 1874 | } | selfish             | 95   | 5.1  |
|                     |      |   | altruistic          | 1750 | 93.4 |
|                     |      |   | conservative        | 29   | 1.5  |
| conservative        | 346  | } | selfish             | 11   | 3.2  |
|                     |      |   | altruistic          | 30   | 8.7  |
|                     |      |   | conservative        | 305  | 88.2 |

On the left side there are participants divided into groups depending on their vote in the public/parliament question. The right side describes their other decision in the secret/parliament question, so the change illustrated in the following Fig. 7 is completed with exact numbers

selfish AVs. Zahavi’s (1975) theory shows that the only evolutionary stable strategy in this circumstance is to make the signalling costly for the cheaters. In the case of signalling an altruistic strategy, it should be obligatory to adopt such legal and technical measures that will make fake signalling both illegal and as technically difficult as possible.

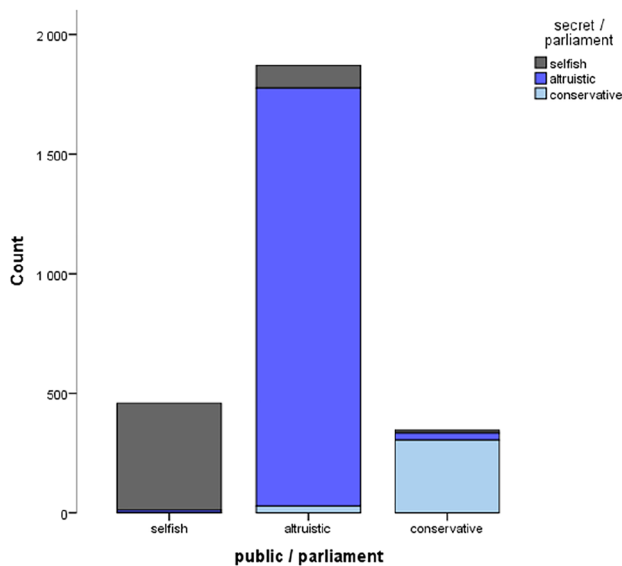
Lee et al. (2018) showed in their study of trolley problems that experimentally induced social observation leads to a drop in utilitarianism, i.e., there is a higher preference for deontological judgement if the person is socially observed. Based on their study, one would expect that participants in a public personal choice would shift towards lower utilitarianism as well. However, in our study, the opposite was the case. A possible explanation has already been suggested in the introduction: having an altruistic (utilitarian) AV is at the same time being altruistic towards strangers—the person is willing to sacrifice his or her own life to save others and thus the reputation concern would work in the opposite direction. The fact that one’s own life is involved in the situation is the game-changer here.

While data were being gathered for this study in the Czech population, the German ethical committee for autonomous driving issued a report (German Ministry of Transportation (2017; philosophically reviewed in Luetge (2017)

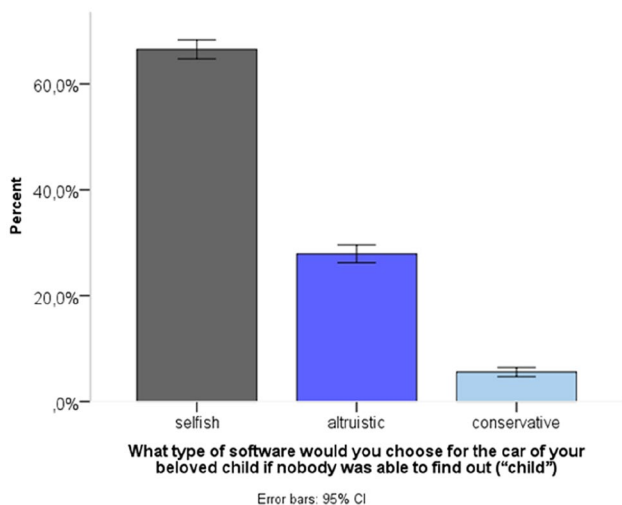
in which it is insisted that no innocent human lives may be taken by AVs in collision situations. It is prohibited to offset victims against one another, even though reducing the number of personal injuries may be justifiable. This means that AVs would not be able to react to trolley-like scenarios. This amounts to opting for conservative AVs in our terminology. This strategy was by far the least popular among our respondents in the Czech Republic in all five scenarios, at maximum accepted by 13% of participants in case of a public vote in Parliament. The striking difference between German ethical and legal experts and the Czech public may indicate that there will be culturally dependent and diverse attitudes in different countries. A recent study (Awad et al. 2018) on a global scale indicates similar findings that the strongest preference is for saving more people in lethal collision as well. Further studies on these cultural influences on moral judgements with respect to AVs are needed.

#### 4.2 Ethical considerations

German declarations (Luetge 2017) state that the responsibility for driving a car should always stay with an (individual) human person. Perhaps surprisingly, this solution could lead to a widespread use of selfish AVs. Moreover,



**Fig. 7** The change in preference of three strategies in a public parliamentary vote by subjects who choose between various strategies in a secret parliamentary vote. Columns show the preference in public/parliament question and colours indicate what the same person has chosen in secret/parliament scenario. E.g. in the second column there are all people who preferred altruistic choice in public/parliament and charcoal-grey colour signifies a subgroup of them that preferred selfish strategy in secret/parliament scenario



**Fig. 8** Secret/child. Y-axis shows the fraction of subjects selecting a particular variant of the strategy. The error bars show the 95% confidence interval

it is unrealistic and unfair to expect individual (non-driving) operators to be able to step in at the last moment in emergency situations and then bear the full responsibility for these exceptional collision situations. Thus, our policy recommendation is to decide about the permissible AV type

at the state level, rather than leave it up to private choice and competition between AV manufacturers. It would be especially undesirable to try to regulate the choice economically: for example, if the selfish cars were made more expensive or fell under higher taxes, this would result in a “selfish” advantage for the wealthy buyers. Alternatively, to provide the public with the possibility of free choice, we recommend that laws require a clear indication of what type of AV one is driving. This too is likely to increase the number of altruistic AVs and hence overall decrease the death rate by car accidents. This kind of peer pressure would appeal to the sense of moral responsibility towards the other members of the AV owner's community, which is comparable to the peer pressure already present e.g. due to discussions about ecology and responsibility towards one's environment.

If it becomes possible for the public to choose freely between various types of driving software in the near future, we strongly advocate the introduction of obligatory measures by means of design indication (e.g. colours, stripes) to publicly display altruism or at least non-selfishness. This measure is necessary for social sanctions and reputational concern to properly accommodate customer choice. We believe that the altruistic algorithm is the right choice given the most commonly held ethical theories, especially consequentialism. Other traditional theories, such as deontology, contractualism or virtue theory, are at least compatible with it as well, or at any rate, it is not straightforward clear how to apply them to decide for or against altruism (Nyholm 2018). At any rate, we do not have the opportunity here to discuss more general issues in ethical theory in detail and the data suggest that the wider public also prefers altruistic strategy, whatever the ultimate ethical theory. The selfish strategy only competes with the altruistic one in the specific “secret choice” situation, which suggests that this type of selfishness is not considered as good and preferable at a more abstract moral level. (For more general considerations how to take avail of such moral intuitions, see Savulescu et al. (2019).

Less clear is the problem of how to deal with our heightened sensibility and “passenger-prioritizing” with respect to our own offspring. Careful approach is needed here to prevent moral panic. Future research should include a study of moral judgments in other samples, especially in non-Western cultures, to test whether our results can be generalized. The decision over which type of AV to choose for one's own child proved to be the most problematic, since in this scenario participants chose mostly “passenger-prioritizing” software referred in paper as “selfish”. This indicates that the future AV communication ought not to be primarily focused on convincing the drivers themselves but on communicating the advantages of AVs to parents regarding the safety of school buses or AV taxis for their children.

Although our study is focused specifically on AV type choice, we need to be aware that it is not a standalone



aspect of AV mobility. Real-life environment, especially cities where most of the human–AI interactions do and will take place turns the ethics of AV into a much more complex issue. First, there is the question of test implementation of AV that is already happening in cities of various sizes and geographical locations, usefulness and effectiveness of their experimental design (Karvonen and van Heur 2013), as well as their impact on local politics as discussed in Dowling and McGuirk (2020). Based on investigation into public stances toward implementation and type of ownership of AVs, a second issue arises as of how will the price, availability, and comfort of personal AV impact the scale of personal transportation in cities and how will it translate into changes in urban planning (Cugurullo et al. 2020) and political regulatory context (Aoyama and Alvarez Leon 2021). New urban technologies, including autonomous transportation are incorporated into modern plans of smart and sustainable cities (Chehri and Mouftah 2019), both newly built and transformed from existing ones, but the actual outcome is often problematic (Cugurullo 2017), which further underlines the need for complex consideration of the issue as suggested also in Cugurullo (2021).

### 4.3 Limitations

Our results are based on self-reports in a broad anonymous internet survey and not on respondents' observable behaviour in real-life situations. Therefore, we do not claim to know their real-life choices in buying AVs or voting in Parliament. On the other hand, it is probable that both observed phenomena (higher preference for altruistic strategies in public choices and lower when our children are involved) will play an important role also in real-life decisions. Anyway, a real preference for an altruistic strategy would be proven only by observing individual behaviour in everyday life or in an experimental setting. Another possible limitation of this study is the self-selection of participants, since the participation was not monetary-incentivized. Last but not least, this broad internet survey although with large sample size cannot be considered a sociological representative sample for the whole Czech population and could be culturally specific.

## 5 Conclusion

Our results show that the respondents exhibited a clear preference in AVs for an altruistic strategy that minimizes the number of casualties, even if this leads to death of passenger(s). This preference was reinforced if the AV signals its strategy to others. The altruistic preference was strongest when people choose software for everybody else, weaker in their personal choice, and the weakest in the

choice for their own child, when people preferred strategy that protects the lives of passenger(s) over any number of bystanders. The moral intuition of Czech participants seems not to correspond with the recommendation of the German ethical committee, as the AV strategy that abstains from the trolley problem was the least popular in all situations for participants in our study. Based on the results, we conclude that, in contrast to a private choice, a public choice is considerably more likely to pressure consumers in their personal choice to accept a non-selfish solution, making it a reasonable and relatively cheap way to shift car owners and users towards higher altruism. Also, a hypothetical voting in Parliament about a single available program would probably be less selfish if the voting did not take place in secret.

**Acknowledgements** RK developed the study concept. RK, MJK, JF were responsible for the study design and data collection. RK performed the data analysis and interpretation under the supervision of JF. RK, DN, DC drafted and all authors revised the manuscript. All authors approved the final version of the manuscript for submission. For comments on earlier drafts of this paper we would like to thank: Světa Hanke Jarošová. We would also like to thank Lenka Příplatová and anonymous reviewers for very useful comments on the revision of this paper.

**Funding** This work was supported by the Grant Agency of the Charles University under grant GAUK number 929216; Charles University Research Program under grant UNCE number 204056; and The Technological Agency of the Czech Republic under ETA grant number TL01000467.

**Data Availability** <https://osf.io/y7t5q>.

### Declarations

**Conflict of interest** No potential conflict of interest was reported by the authors.

## References

- Aoyama Y, Alvarez Leon LF (2021) Urban governance and autonomous vehicles. *Cities* 119:103410. <https://doi.org/10.1016/j.cities.2021.10341>
- Awad E, Dsouza S, Kim R, Schulz J, Henrich J, Shariff A, Bonnefon JF, Rahwan I (2018) The moral machine experiment. *Nature* 563(7729):59–64. <https://doi.org/10.1038/s41586-018-0637-6>
- Bereczkei T, Birkas B, Kerekes Z (2010) Altruism towards strangers in need: costly signaling in an industrial society. *Evol Hum Behav* 31(2):95–103. <https://doi.org/10.1016/j.evolhumbehav.2009.07.004>
- Bonnefon JF, Shariff A, Rahwan I (2016) The social dilemma of autonomous vehicles. *Science* 352(6293):1573–1576. <https://doi.org/10.1126/science.aaf2654>
- Brown M, Sacco DF (2017) Is pulling the lever sexy? Deontology as a downstream cue to long-term mate quality. *J Soc Pers Relat* 36(3):957–976. <https://doi.org/10.1177/0265407517749331>

- Buunk AP, Massar K (2012) Intrasexual competition among males: competitive towards men, prosocial towards women. *Personal Individ Differ* 52(7):818–821. <https://doi.org/10.1016/j.paid.2012.01.010>
- Chehri A, Mouftah HT (2019) Autonomous vehicles in the sustainable cities, the beginning of a green adventure. *Sustain Cities Soc* 51:101751. <https://doi.org/10.1016/j.scs.2019.101751>
- Contissa G, Lagioia F, Sartor G (2017) The Ethical Knob: ethically-customisable automated vehicles and the law. *Artif Intell Law* 25(3):365–378. <https://doi.org/10.1007/s10506-017-9211-z>
- Cugurullo F (2017) Exposing smart cities and eco-cities: frankenstein urbanism and the sustainability challenges of the experimental city. *Environ Plan a: Econ Space* 50(1):73–92. <https://doi.org/10.1177/0308518x17738535>
- Cugurullo F (2021) *Frankenstein urbanism* (Routledge studies in urbanism and the city), 1st edn. Routledge, London
- Cugurullo F, Acheampong RA, Gueriau M, Dusparic I (2020) The transition to autonomous cars, the redesign of cities and the future of urban sustainability. *Urban Geogr* 42(6):833–859. <https://doi.org/10.1080/02723638.2020.1746096>
- Dawkins R, Krebs JR (1978) Animal signals: information or manipulation. *Behav Ecol: Evolut Approach* 2:282–309
- de Sio FS (2017) Killing by autonomous vehicles and the legal doctrine of necessity. *Ethical Theory Moral Pract* 20(2):411–429. <https://doi.org/10.1007/s10677-017-9780-7>
- Dowling R, McGuirok P (2020) Autonomous vehicle experiments and the city. *Urban Geogr*. <https://doi.org/10.1080/02723638.2020.1866392>
- Epting S (2018) Automated vehicles and transportation justice. *Philos Technol* 32(3):389–403. <https://doi.org/10.1007/s13347-018-0307-5>
- Fehrler S, Przepiorka W (2013) Charitable giving as a signal of trustworthiness: disentangling the signaling benefits of altruistic acts. *Evol Hum Behav* 34(2):139–145. <https://doi.org/10.1016/j.evolhumbehav.2012.11.005>
- Foot P (2002) The problem of abortion and the doctrine of the double effect. In: Foot P (ed) *Virtues and vices and other essays in moral philosophy*. Clarendon Press, Oxford, pp 19–33 (**Originally appeared in Oxford Review, 5, 1967**)
- Gerdes JC, Thornton SM (2016) Implementable ethics for autonomous vehicles. In: Mauer M et al (eds) *Autonomous driving. Technical, legal and social aspects*. Springer, Berlin/Heidelberg, pp 87–102
- German ministry of transportation. Report. (2017). [https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?\\_\\_blob=publicationFile](https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile)
- Gogoll J, Müller JF (2016) Autonomous cars: in favor of a mandatory ethics setting. *Sci Eng Ethics* 23(3):681–700. <https://doi.org/10.1007/s11948-016-9806-x>
- Goodall NJ (2014) Ethical decision making during automated vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board* 2424(1):58–65. <https://doi.org/10.3141/2424-07>
- Haboucha CJ, Ishaq R, Shifan Y (2017) User preferences regarding autonomous vehicles. *Transp Res Part C: Emerg Technol* 78:37–49. <https://doi.org/10.1016/j.trc.2017.01.010>
- Hardy CL, van Vugt M (2006) Nice guys finish first: the competitive altruism hypothesis. *Personal Soc Psychol Bull* 32(10):1402–1413. <https://doi.org/10.1177/0146167206291006>
- Hulse LM, Xie H, Galea ER (2018) Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age. *Saf Sci* 102:1–13
- Kagan S (1991) *The limits of morality* (Oxford ethics series), Revised. Clarendon Press, Oxford
- Karvonen A, van Heur B (2013) Urban laboratories: experiments in reworking cities. *Int J Urban Reg Res* 38(2):379–392. <https://doi.org/10.1111/1468-2427.12075>
- Kraft-Todd GT, Bollinger B, Gillingham K, Lamp S, Rand DG (2018) Credibility-enhancing displays promote the provision of non-normative public goods. *Nature* 563(7730):245–248. <https://doi.org/10.1038/s41586-018-0647-4>
- Laidre ME, Johnstone RA (2013) Animal signals. *Curr Biol* 23(18):R829–R833. <https://doi.org/10.1016/j.cub.2013.07.070>
- Lee M, Sul S, Kim H (2018) Social observation increases deontological judgments in moral dilemmas. *Evol Hum Behav* 39(6):611–621. <https://doi.org/10.1016/j.evolhumbehav.2018.06.004>
- Lin P (2015) Why ethics matters for autonomous cars. In: Maurer M et al (eds) *Autonomous driving. Technical, legal and social aspects*. Springer, Berlin/Heidelberg, pp 69–85
- Luetge C (2017) The German ethics code for automated and connected driving. *Philos Technol* 30(4):547–558. <https://doi.org/10.1007/s13347-017-0284-0>
- Maurer M et al (2016) *Autonomous driving. Technical, legal and social aspects*. Springer, Berlin/Heidelberg
- Moody J, Bailey N, Zhao J (2020) Public perceptions of autonomous vehicle safety: an international comparison. *Saf Sci* 121:634–650. <https://doi.org/10.1016/j.ssci.2019.07.022>
- Nielsen TAS, Haustein S (2018) On sceptics and enthusiasts: What are the expectations towards self-driving cars? *Transp Policy* 66:49–55. <https://doi.org/10.1016/j.tranpol.2018.03.004>
- Noothigattu R, Gaikwad S, Awad E, Dsouza S, Rahwan I, Ravikumar P, Procaccia A (2018) A voting-based system for ethical decision making. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1.
- Nowak MA, Sigmund K (1998) Evolution of indirect reciprocity by image scoring. *Nature* 393(6685):573–577. <https://doi.org/10.1038/31225>
- Nyholm S (2018) The ethics of crashes with self-driving cars: a roadmap. *I. Philos 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 Moral Pract* 19(5):1275–1289. <https://doi.org/10.1007/s10677-016-9745-2>
- Olivella P, Siciliani L (2017) Reputational concerns with altruistic providers. *J Health Econ* 55:1–13. <https://doi.org/10.1016/j.jhealeco.2017.05.003>
- Puska P, Kurki S, Lähdesmäki M, Siltaoja M, Luomala H (2016) Male-male status signaling through favoring organic foods: is the signaler perceived and treated as a friend or a foe? *Psychol Mark* 33(10):843–855. <https://doi.org/10.1002/mar.20922>
- Savulescu J, Kahane G, Gyngell C (2019) From public preferences to ethical policy. *Nat Hum Behav* 3(12):1241–1243. <https://doi.org/10.1038/s41562-019-0711-6>
- Semmann D, Krambeck HJ, Milinski M (2005) Reputation is valuable within and outside one's own social group. *Behav Ecol Sociobiol* 57(6):611–616. <https://doi.org/10.1007/s00265-004-0885-3>
- Shariff A, Bonnefon JF, Rahwan I (2017) Psychological roadblocks to the adoption of self-driving vehicles. *Nat Hum Behav* 1(10):694–696. <https://doi.org/10.1038/s41562-017-0202-6>
- Talebpour A, Mahmassani HS (2016) Influence of connected and autonomous vehicles on traffic flow stability and throughput. *Transp Res Part c: Emerg Technol* 71:143–163. <https://doi.org/10.1016/j.trc.2016.07.007>
- Thomson JJ (1985) The trolley problem. *Yale Law J* 94:1395–1415
- Zahavi A (1975) Mate selection—a selection for a handicap. *J Theor Biol* 53(1):205–214. [https://doi.org/10.1016/0022-5193\(75\)90111-3](https://doi.org/10.1016/0022-5193(75)90111-3)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.