Trust in autonomous cars: The role of value similarity and capacity for sympathy

Ryosuke YOKOI1),2),*, Kazuya NAKAYACHI3)

1) Graduate School of Psychology, Doshisha University
2) Japan Society for the Promotion of Science Research Fellow
3) Faculty of Psychology, Doshisha University

Abstract

This study tests the effect of moral value similarity on trust in autonomous cars (ACs). We adopt two moral values: utilitarianism (promoting a greater good) and deontology (condemning deliberate harm). Previous research found that utilitarianism similarity had a significant effect on trust in ACs, whereas deontology similarity did not. The research also revealed that when participants preferred a deontological action, ACs were less trusted than a human driver, even when the ACs performed the same action as the participants. We investigated the replicability of these findings and whether distrust in ACs arises from the ACs' inability to sympathize with potential victims. Our online experiment (N=609) found that both utilitarian and deontology similarities positively influenced trust in ACs. Mediation analysis also indicated that when the driver was an AC, the participants recognized that the AC lacked the capacity to feel sympathy, thereby decreasing trust in it. The paper also discusses the theoretical implications of our findings.

Key Words: autonomous car, trust, value similarity, moral dilemma, sympathy

Introduction

There has been an increasing focus on trust in autonomous cars (ACs) due to the relationship between human trust in ACs and their acceptance (Buckley, Kaye, & Pradhan, 2018; Xu et al., 2018). Research on risk perception has produced robust findings that perceived value similarity determines trust in risk managers (Siegrist, 2021); that is, when people perceive that their desired value is similar to the agent’s value, their trust in the agent is established.

Yokoi & Nakayachi (2020), who focused on the manipulation of perceived value similarity, examined the effect of moral value similarity on trust in ACs. They adopted utilitarianism and deontology as moral values because ACs may face future dilemmas that require them to make either utilitarian or deontological judgments (Greene, 2016). Utilitarianism asserts that judgment should be based on the maximization of well-being, that is, saving as many people as possible, even by endorsing deliberate harm (Bentham, 1789/1967), whereas deontology refuses such behavior, forbidding deliberate harm, regardless of the consequences (Kant, 1785/1976). Using a moral dilemma scenario, Yokoi & Nakayachi (2020) asked participants to make either a utilitarian or a deontological judgment (Figure 1a). In their
scenario, the utilitarian judgment was to save five pedestrians, while the deontological judgment was to save a woman and a baby. To allow the researchers to manipulate perceived value similarity, the participants then read a scenario in which the AC made a decision that was either consistent or inconsistent with their values (Figures 1b and 1c). The results demonstrated that similarity with respect to making utilitarian judgments had a significant effect on trust in ACs, while similarity in making deontological judgments did not. Yokoi & Nakayachi (2020) also found that when participants desired deontology, ACs were less trusted than human drivers even if the AC performed the same action as the participants.

Yokoi & Nakayachi (2020) interpreted that the low levels of trust in ACs are attributed to the ACs' inability to feel sympathy. Participants who preferred deontology might resist trusting ACs that cannot feel sympathy because they might sympathize with the vulnerability of a woman and a baby. One of our objectives is to empirically investigate the validity of their post-hoc interpretation. We expect that when people prefer deontology and ACs make a deontological judgment, the ACs' inability to feel sympathy explains why they are less trusted than a human driver. That is, the perception of the capacity for sympathy would mediate the influence of the type of driver (AC or human) on trust in drivers. Research on mind perception has examined whether people recognize that a robot and machine have a mind of their own (Shank, Graves, Gott, Gamez, & Rodriguez, 2019; Shank & Gott, 2019). For example, people perceived the mind in artificial intelligence when they were given information that artificial intelligence could behave as a human (Shank & Gott, 2019). People may also think that ACs can feel emotions in some cases, as shown by the research on the mind perception of the machines. This study also examines the replicability of the findings of Yokoi & Nakayachi (2020). Our method is identical to theirs, except for measuring participants' perceived capacity for sympathy with regard to ACs.

**Method**

**Participants and experimental design**

The ideal sample size was determined using G*power3.1 with an $\alpha$-error level of .05, a desired power of .80, and an $\eta^2_p$ of .032. The power analysis suggested a minimum sample size of 240 participants. A total of 609 adults (304 males and 305 females) volunteered to participate in our experiment.

![Figure 1](image.png)

Figure 1 The traffic dilemma situations used in Yokoi & Nakayachi (2020) and the current experiment. (a) Participants were asked to make a decision about whether to run over five pedestrians by staying in the current lane or a woman and a baby by changing lanes. (b) Driver (AC or human) stays in the current lane, running over the five pedestrians (deontology). (c) Driver changes lanes, running over the woman and the baby (utilitarianism). AC=autonomous car.

2) The effect size ($\eta^2_p$) was based on an interaction effect reported in past research showing that only utilitarianism similarity influenced trust in ACs (Yokoi & Nakayachi, 2020).
research for compensation through the Japanese survey company Cross Marketing (https://www.cross-m.co.jp/).

The number of recruited participants far exceeded the number the power analysis indicated would be required because it was expected that many satisficing responses would be found in the online survey. The respondents’ ages ranged from 22 to 59 years ($M=40.29$ years, $SD=±10.66$ years). Electronic informed consent was obtained from all participants through Cross Marketing. The experiment employed a 2 (moral value similarity: similar vs. dissimilar) × 2 (driver: AC vs. human) × 2 (participants’ preferred moral value: utilitarianism vs. deontology) between-subjects factorial design. Each participant self-selected their preferred value, rather than accomplishing this through experimental manipulation. The dependent variable was trust in drivers.

Moral dilemma scenarios

Our experiment was conducted using the online survey tool Qualtrics. Participants were randomly assigned to one of the four conditions as part of a 2 (moral value similarity: similar vs. dissimilar) × 2 (the driver: AC vs. human) design. Following Yokoi & Nakayachi (2020), we used two scenarios (Scenarios 1 and 2) to create a traffic-related moral dilemma (see the online supplemental file for details about the scenarios). Scenario 1 asked the participants to make an ethical judgment to either run over five pedestrians by staying in the current lane or run over a woman and a baby by changing to another lane (Figure 1a). They then read Scenario 2, where the driver (an AC or a human) ran over the five pedestrians (or the woman and the baby) in the same situation described in Scenario 1 (Figures 1b and 1c). The driver’s decision depended on the experimental condition and each participant’s moral judgment. If the participant in a similar condition chose the utilitarian action, for example, the participant read that the driver (an AC or a human) made a utilitarian decision. If the participant in a dissimilar condition made the utilitarian judgment, the driver made a deontological choice. The human driver in Scenario 2 was given the name Mr. A.

Running over the five pedestrians to save the woman and baby might be regarded as a utilitarian judgment because the life expectancy of the baby was longer than that of the other pedestrians. However, we described the five pedestrians as middle-aged (see the online supplemental file). This description could prevent participants from perceiving the five pedestrians as elderly. Therefore, the total life expectancy of the five pedestrians was not shorter than that of the woman and the baby.

Measures

After reading Scenario 2, the participants were asked to answer a questionnaire to measure trust in the driver as the dependent variable ($a=.93$), perceived value similarity as a manipulation check ($a=.95$), and perceived capacity for sympathy as the mediating variable ($a=.96$). A five-point Likert scale was used for all items in the questionnaire with a scale ranging from “1=do not agree at all” to “5=strongly agree.” This study also developed three comprehension questions to check whether the participants responded to the questions appropriately and actually read the scenarios. These comprehension checks were used to identify satisficing responders. These items can be found in the online supplemental file, along with the dilemma scenarios.

Results

Respondents who failed to correctly answer the comprehension questions were eliminated, resulting in a final sample of 328 respondents for analysis. In Scenario 1, 94 participants made a utilitarian decision, whereas 234 made a deontological decision. This proportion is in line with the findings that females and children are more likely to be saved in traffic moral dilemmas than males and adults (Awad et al., 2018; Faulhaber et al., 2019).

Manipulation check

We conducted a 2 (moral value similarity) × 2 (driver) × 2 (participants’ preferred value) between-subjects ANOVA on perceived value similarity for the manipulation check. The analysis detected the significant main effects of moral value similarity ($F(1, 320)=59.67, p<.001, \eta^2_p=.157$) and type of driver ($F(1, 320)=12.24, p=.001, \eta^2_p=.037$). The results also indicated a marginally significant effect of participants’ preferred value ($F(1, 320)=3.63, p=.058, \eta^2_p=.011$). There was no significant effect for any interaction (see the online supplemental file). These results demonstrated that the participants in a similar condition rated perceived value similarity higher than those in a dissimilar condition, meaning that our manipulation of moral value similarity was successful. Descriptive statistics for perceived value similarity are available in the online supplemental file.
Trust in driver

The ANOVA was performed with trust as the dependent variable (Figure 2). The results showed a significant effect of moral value similarity ($F(1, 320)=6.13, p=.014, \eta^2_p=.019$) and drivers ($F(1, 320)=6.86, p=.009, \eta^2_p=.021$). Participants’ preferred value ($F(1, 320)=.53, p=.466, \eta^2_p=.002$), two-way interactions (between value similarity and driver ($F(1, 320)=.45, p=.505, \eta^2_p=.001$); between driver and participants’ preferred value ($F(1, 320)=1.10, p=.295, \eta^2_p=.003$); between value similarity and participants’ preferred value ($F(1, 320)=.01, p=.918, \eta^2_p<.001$), and three-way interaction ($F(1, 320)=.06, p=.801, \eta^2_p<.001$) did not have significant effects on trust. The results suggested that drivers were trusted more when their moral values were similar to the participants’ values than when they were not. Additionally, it was found that ACs were less trusted than human drivers.

Mediation analysis

Bootstrapped mediation analyses were then conducted to examine whether perceived capacity for sympathy could mediate the effect of the type of driver on trust when both the participants and drivers selected a deontological judgment (Figure 3). The analysis found a significant mediation effect ($B=-.47, 95\% CI [-.77, -.23]$). This result suggested that when the driver was an AC, the participants recognized that the AC lacked the capacity to feel sympathy, which decreased their trust in it. We also tested whether

![Figure 2](image_url)

**Figure 2** Means of trust by participants preferring either utilitarianism or deontology. Error bars indicate 95\% confidence intervals. Values in brackets are frequency. AC=autonomous car.

![Figure 3](image_url)

**Figure 3** The results of the mediation analysis in a case where both the participant and the driver made a deontological judgment. The values are unstandardized coefficients. $N=121$. AC=autonomous car. **$p<.01$.  

3) The resampling number was 5,000.
the mediation model was established when both the participants and drivers made a utilitarian judgment. Although the analysis indicated a significant mediation effect (B=−.44, 95% CI [−.88, −.16]), the type of driver had no significant direct effect on trust in the drivers (see online supplemental file Figure S2). That is, there was no significant difference in the trust scores between ACs and humans in the first place. The mediation effect of perceived capacity for sympathy was investigated when the participants took a deontological action and drivers took a utilitarian one. A significant mediation effect was detected (B=−.51, 95% CI [−.78, −.31]). The direct effect of the type of driver was also significant (see online supplemental file Figure S3). These results suggested that when the participants preferred deontology, irrespective of moral value similarity, the perceived ability to feel sympathy explained why an AC was less trusted than a human driver. Descriptive statistics for perceived capacity for sympathy are available in the online supplemental file (see online supplemental file Figure S4).

Discussion

The current research investigates the replicability of the findings of Yokoi & Nakayachi (2020) and the factor that explains why levels of trust in ACs are lower than those for human drivers when both people and ACs select a deontological action.

Our experiment revealed that both utilitarianism and deontology similarities positively influenced trust in ACs. This result is supported by risk perception research, which found that perceived value similarity increases trust in agents. However, this is not in line with Yokoi & Nakayachi (2020)’s result that only utilitarianism similarity has a significant effect. Previous research did not detect the effect of deontology similarity. The discrepancy between our results and theirs might be due to a difference in sample size. Yokoi & Nakayachi (2020)’s sample (N=196) was smaller than ours. Hence, in our study, the detection power may have been enhanced, resulting in the finding that participants with a preference for deontological judgment trusted ACs when they took the same action more than when they did not. Although Yokoi & Nakayachi (2020) did not find a statistically significant effect of deontology similarity on trust in ACs, the mean score of trust was higher when participants selected deontology and ACs selected utilitarianism than when both made a deontological decision. Their results are not in line with our findings or with the risk perception research finding that perceived value similarity promotes trust in agents. Future studies should further examine the relationship between moral value similarity and trust in ACs.

Mediation analysis confirmed that ACs’ inability to sympathize with potential victims contributed to people’s distrust in ACs, even if the ACs made a deontological judgment over a utilitarian one in line with participants’ preference. This result supports Yokoi & Nakayachi (2020)’s interpretation that people who want to save a woman and a baby feel strong sympathy for them and may be reluctant to trust ACs because they lack the capacity for sympathy. Our analysis found the mediation model, on the other hand, even when participants preferred the deontological judgment and the drivers (AC or human) selected the utilitarian one. Perceived capacity for sympathy may explain the difference in trust between ACs and human drivers, regardless of value similarity. Bigman & Gray (2018) also found that other machines, such as medical diagnosis systems and military drones, are not likely to be entrusted with moral decisions. Their study clarified that people’s aversion to such machines is attributable to the machines’ inability to experience emotions, including pain and compassion. The difference arising from the capacity to feel sympathy results in lower trust in ACs than human drivers and may be more robust.

Why did perceived capacity for sympathy lead to low trust in ACs when people preferred deontological judgment? The ad-hoc interpretation might be as follows. Cummins & Cummins (2012) claimed that as a baby and a female are highly vulnerable, people strongly feel sympathy for them in a moral dilemma situation. Our participants, who saved the woman and the baby, might feel more powerful sympathy for them than the five pedestrians in the left lane. Therefore, perceived capacity for sympathy might be a critical factor in trust. ACs cannot feel emotion, of course. Thus, they were trusted less than human drivers when the participants selected deontology. We suggest that future research measure people’s felt sympathy for potential victims to test the validity of this interpretation.

Some limitations of this study should be mentioned. The scenario method was adopted in the experiment to allow participants to make a reflective decision; however, in a real traffic situation, drivers may be required to make a judgment quickly. Hence, a driving simulator should be
used in the future to enhance ecological validity. Deontological judgment, in our scenario, was identified as saving a woman and a baby. They might feel greater sympathy for the woman and baby than the five pedestrians and therefore prefer deontological judgment. Our research cannot confirm whether they took the deontological action based on deontological ethics or sympathy for the woman and baby. To correctly test the effect of deontology similarity, participants must make a decision, not based on sympathy but on deontological ethics. Future studies should create a scenario in which pedestrians’ attributes are the same in both lanes. This improvement would enable the removal of the extraneous variable of participants’ felt sympathy.

References

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... Rahwan, I. (2018). The moral machine experiment. *Nature*, 563, 59–64.

Bentham, J. (1967). *An introduction to the principles of morals and legislation* (Y. Seki, Ed.; S. Yamashita, Trans.). Tokyo: Chuo Koron. (Original work published 1789).

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, 181, 21–34.

Buckley, L., Kaye, S. A., & Pradhan, A. K. (2018). Psychosocial factors associated with intended use of automated vehicles: A simulated driving study. *Accident Analysis & Prevention*, 115, 202–208.

Cummins, D. D., & Cummins, R. C. (2012). Emotion and deliberative reasoning in moral judgment. *Frontiers in Psychology*, 3, 328.

Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., ... 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, 399–418.

Greene, J. D. (2016). ETHICS. Our driverless dilemma. *Science*, 352, 1514–1515.

Kant, I. (1976). *Grundlegung zur metaphysik der sitten* (H. Shinoda, Trans.). Tokyo: Iwanami Shoten. (Original work published 1785).

Shank, D. B., & Gott, A. (2019). People’s self-reported encounters of Perceiving Mind in Artificial Intelligence. *Data in Brief*, 25, 104220.

Shank, D. B., Graves, C., Gott, A., Gamez, P., & Rodriguez, S. (2019). Feeling our way to machine minds: People’s emotions when perceiving mind in artificial intelligence. *Computers in Human Behavior*, 98, 256–266.

Siegrist, M. (2021). Trust and risk perception: A critical review of the literature. *Risk Analysis*, 41, 480–490.

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

Yokoi, R., & Nakayachi, K. (2020). Trust in autonomous cars: Exploring the role of shared moral values, reasoning, and emotion in safety-critical decisions. *Human Factors*. https://doi.org/10.1177/0018720820933041