

Causes of Social Dilemma: A Case of Autonomous Vehicles

MORITA Tamaki* and MANAGI Shunsuke**

Abstract

Artificial intelligence (AI) equipped in autonomous vehicles (AV) that drives fully automatically may have to face dilemmas to choose between two evils, such as running over pedestrians versus sacrificing themselves and their passengers in order to save them. This dilemma may also cause a social dilemma that if people have different ideas about “moral AVs” versus “wishing AVs” while wishing (or demanding) AVs are immoral, the market will eventually be filled with immoral AVs regardless of peoples' ascribed moralities. With an online survey giving 14,829 effective responses from all over Japan, we find that the social dilemma will occur in Japan as well as in the U.S. We also focused on the relationship between the morality and the relative intention to buy “moral AVs,” likewise exploring the factors of the social dilemma. We find that the more credible AVs are, the higher the likelihood that social dilemmas will occur. This finding implies that there is a possibility that such social dilemmas will not be resolved until the number of car accidents reaches zero.

キーワード：自動運転車、社会的ジレンマ、人工知能 (AI)、道徳観

Key words: Artificial Intelligence (AI), Autonomous Vehicles, Morality, Social Dilemma

I. Introduction

Social dilemma is defined as the situation when “private interests are at odds with collective interests” (Van Lange, 1989). In the case of autonomous vehicles (AVs), if people have different ideas about “moral AVs” and “wishing AVs” and if wishing AVs are immoral, it will cause a social dilemma. The market will eventually be filled with immoral AVs because people will actually purchase what they need regardless of their moral scruples.

An artificial intelligence (AI)-equipped AV that drives fully automatically may end up facing the dilemma of choosing between two evils such as running over pedestrians versus sacrificing the passengers themselves in order to save them.¹⁾ In a survey conducted in the

United States, Bonnefon, et al. (2016) revealed that people would like AVs to be utilitarian, i.e., a moral actor to save as many lives as possible even though the AVs will occasionally sacrifice passengers, whereas people will prefer to purchase AVs that are programmed to save passengers (themselves). They propose two suggestions to overcome this social dilemma in Shariff, et al. (2017) : One is to shift the discussion from the relative risk of injury to the absolute reduction of risk to passengers due to overall accident reduction, and the other is to appeal to consumers' desire for virtue signaling. We feel the need to provide an econometric analysis of the causes of the dilemma further before we make such suggestions.

* 山梨県立大学 国際政策学部 総合政策学科

Department of Policy Management, Faculty of Glocal Policy Management and Communications, Yamanashi Prefectural University

**九州大学 工学研究院 環境社会部門 都市システム工学講座

Department of Urban and Environmental Engineering, Faculty of Engineering, Kyushu University

We follow Bonnefon, et al.'s U.S. survey and conducted an online survey named *Survey on Auto Driving* to about 18,000 Japanese respondents²⁾ and collected 14,829 effective responses from those who revealed their intention to buy an AV in the future regardless of the weakness of their intent. In Chapter II, we compare the morality in Japan and the U.S., Chapter III analyzes the relationship between the morality and purchase, Chapter IV investigates the determinants of purchase intention and consumer heterogeneity, and Chapter V concludes.

II. Comparisons of Morality

First, we compare our results with the Bonnefon, et al. (2016)'s previous research in the U.S. In our research, we explained the situation with both the scenario below and Figure 1.

You and [a coworker or an acquaintance/

a family member] are in the car traveling down a main road on a bridge. Suddenly, 10 pedestrians appear ahead, in the direct path of the car. If the car swerves to the side of the road, it will plunge into the river, killing you and your [coworker or acquaintance/family member] but leaving the pedestrians unharmed. If the car stays on your current path, it will kill the 10 pedestrians, you and your [coworker or acquaintance/family member] will be unharmed.³⁾

One-third of the respondents are set to imagine riding alone and read the text beginning with, “You are in the car traveling down a main road...,” another third imagine riding with a coworker or an acquaintance and reading, “You and a coworker or an acquaintance are in the car traveling down a main road...,” and the last third read, “You and a family member are in the car traveling down a main road...”



Figure 1. The Two Alternatives for AVs to Choose

Note: The left caption means “swerve,” and the right one means “stay”.
Source: *Survey on Auto Driving*, modified from Bonnefon et al. (2016) Figure 1, p. 1574.

Morality is defined by the response to the question: Rate what action you think is the most moral, on a 0-100 slider anchored at “stay, saving you [and your coworker or

your acquaintance/and your family member] but killing the 10 pedestrians” and “swerve, sparing the 10 pedestrians but killing you [and your coworker or your acquaintance/

and your family member].” What follows are the questions regarding how inclined they would be to buy an AV programmed to swerve (“minimize” the number of sacrifices, i.e., sacrifice the passengers), and how inclined they would be to buy an AV programmed to stay the course (i.e., be “self-protective”).

Respondents select the answer on a 0-100 slider anchored by “not at all likely” and “extremely likely” for each question. In Figure 2, we see that respondents' tendencies are alike in both countries, except that the purchase intention of a protective AV when riding with families is lower in Japan.

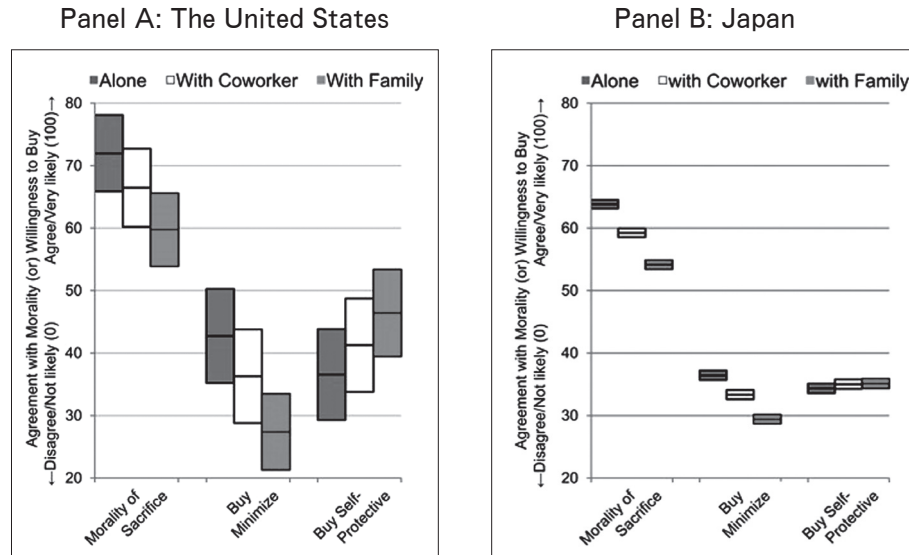


Figure 2. Morality and Purchase Intentions (The United States and Japan)

Source: Bonnefon et al., Figure 3.A, p. 1575, (A); *Surveys on Auto Driving* (B)

Boxes in Figure 2 show the 95% confidence intervals from the mean (the horizontal line in the center of each box is the mean). In both countries, participants expressed moral preferences (over 50) for AVs sacrificing their passengers. However, participants did not express a comparable intention to buy utilitarian AVs that minimize the number of sacrifices especially when they are asked to imagine their family member riding in the car. This tendency is common in both countries. As for self-protective AVs, U.S. respondents showed a higher intention to buy it when they thought of their family members than when riding alone, while the Japanese did not show any significant difference. The Japanese

would not buy moral cars if they usually ride with families, but whether they would buy protective cars does not depend on the existence of fellow passengers.

III. The relationship between Morality and Purchase Intention

We now focus on the relationship between the morality and purchase intention to the two types of the AVs. Figure 3 shows the averages of purchase intention for “minimize” (variable *Buy_Minimize*) and “self-protective” (*Buy_Protective*), and the averages of each respondent's difference between them ($Buy_Minimize - Buy_Protective$, variable *Difference*)⁴ as box plots. The *x*-axis depicts the morality scaled by ten from 0 to 49 and

from 51 to 100. Fifty is solely presented because as much as 17.0% of respondents chose it. If the relationship between *Morality* and *Buy_Minimize* are upward sloping, that between *Morality* and *Buy_Protective* are downward sloping, and that between *Morality* and *Difference* results to be upward sloping, then, the social dilemma will not occur. The more *Moral* people get, the broader the *Difference* and more moral AVs will be purchased.

In Figure 3, the relationship seems to hold,

but the variances are large and unevenly distributed. The distribution of *Buy_Minimize* is upward sloping with almost even but with large variances (Panel A), and that of *Buy_Protective* is downward sloping with large, uneven variances (Panel B). *Difference*, i.e., each respondent's gap between *Buy_Minimize* and *Buy_Protective* (Panel C) is roughly upward sloping, but the variance is also large and uneven. These variances may be one of the reasons for the social dilemma.

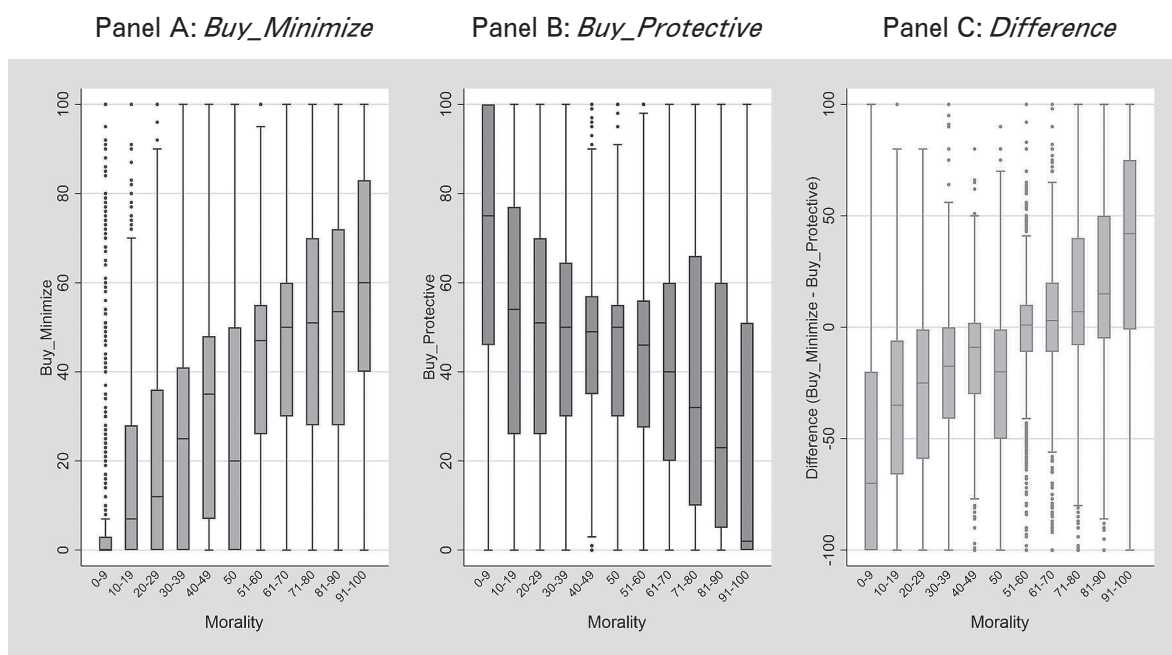


Figure 3. Relationships between Morality and Purchase Intention

Note: Of each box, upper hinge is 75th percentile, lower hinge is 25th percentile, and the centerline depicts the median.

Whiskers show upper and lower adjacent values. Points indicate outside values.

Source: *Survey on Auto Driving*

That being so, let us check the relationship between *Morality* and *Difference* further. Figure 4 is the distribution of the value of *Difference* scattered along with *Morality* scaled by one. The data is the same as Panel C in Figure 3, but those who evaluated both AVs equally ($Buy_Minimize = Buy_Protective$ and Both = 0 in Table 1) are excluded. The upward sloping line is a fitted line with an approximate thickness of the 95% confidence interval. The

fit indicates that, on average, the higher the morality, the more likely respondents are to buy minimizing AVs ($Difference > 0$); and the lower the morality, they are more likely to buy protective AVs ($Difference < 0$). As we noted for Figure 3, if this relationship holds strictly, the social dilemma will not occur. In our samples, however, many marks are unevenly scattered away from the fitted line. Looking at those who prefer to minimize the self-

protected outcome ($Difference > 0$, quadrants one and two in Figure 2), the number of respondents increases with morality, leaving quadrant one ($N = 3,952$) much more crowded than quadrant two ($N = 640$).⁵⁾ Nevertheless, those who prefer protective AVs ($Difference < 0$, quadrants three and four) are not clearly increasing in morality. They almost equally

distribute in quadrant three ($N = 2,300$) and four ($N = 2,489$).⁶⁾ Here, less-moral people ($Morality < 50$) intend to buy protective AVs, but more-moral people ($Morality > 50$) also may buy protective ones. In order to find the reasons for this distribution pattern, we look into the determinants of $Difference$ other than $Morality$.

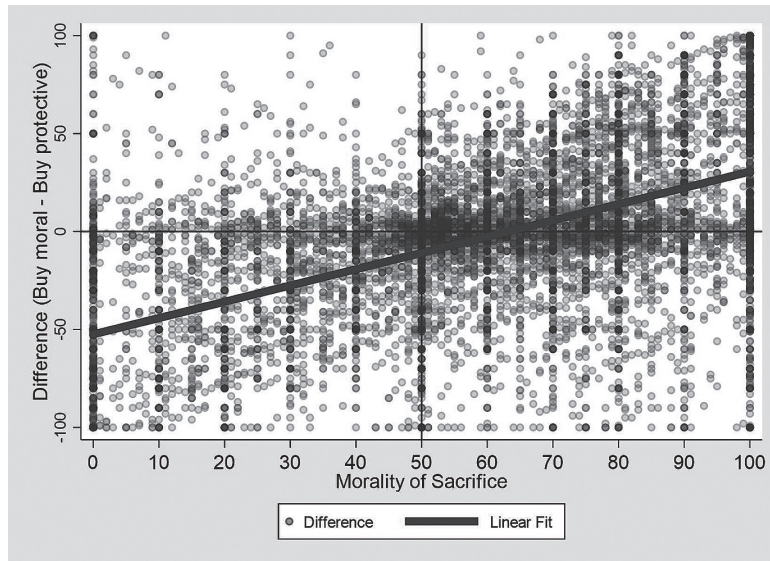


Figure 4. Morality and preferences

Note: Each point depicts each response. The points get thicker where there are multiple responses.
 Source: *Survey on Auto Driving*

Table 1. Frequencies of Responses by Types of Differences

Types of the Difference	Frequency	Percentage
$Buy_Minimize > Buy_Protective$	4,796	32.3
$Buy_Minimize = Buy_Protective$ (except Both = 0)	2,032	13.7
$Buy_Minimize < Buy_Protective$	5,438	36.7
Both = 0	2,563	17.3
Total	14,829	100.0

Note: “Both = 0” means $Buy_Minimize = Buy_Protective = 0$.
 Source: *Survey on Auto Driving*

IV. Determinants of Purchase Intentions

Using simple ordinary least squares (OLS) estimation, we explore the determinants of *Buy_Minimize* (eq. 1 and model 1 in Table 2), *Buy_Protective* (eq. 2 and model 2), *Difference* (eq. 3 and model 3), and *Morality* (eq. 4 and model 4). The equations are as follows:

$$\underline{Buy_Minimize} = \mathbf{Y}^M = \mathbf{X}_p (\boldsymbol{\beta}_p^M) + \mathbf{X}_j (\boldsymbol{\beta}_j^M) + \boldsymbol{\varepsilon}^M \quad (1)$$

$$\underline{Buy_Protective} = \mathbf{Y}^P = \mathbf{X}_p (\boldsymbol{\beta}_p^P) + \mathbf{X}_j (\boldsymbol{\beta}_j^P) + \boldsymbol{\varepsilon}^P \quad (2)$$

$$\begin{aligned} \underline{Difference} &= \underline{Buy_Minimize} - \underline{Buy_Protective} \\ &= \mathbf{Y}^D = \mathbf{X}_p (\boldsymbol{\beta}_p^D) + \mathbf{X}_j (\boldsymbol{\beta}_j^D) + \boldsymbol{\varepsilon}^D \end{aligned} \quad (3)$$

$$\underline{Morality} = \mathbf{Y}^{Mo} = \mathbf{X}_p (\boldsymbol{\beta}_p^{Mo}) + \mathbf{X}_j (\boldsymbol{\beta}_j^{Mo}) + \boldsymbol{\varepsilon}^{Mo}. \quad (4)$$

We set \mathbf{Y}^* as the $n \times 1$ vector of independent variable (underlines denote vectors) where n is the number of respondents, and \mathbf{X}_p as a $n \times 3$ matrix whose columns indicates the passenger type as [*Alone With Coworker With Family*], where one of [0 0 0], [0 1 0], or [0 0 1] is randomly assigned to each respondents. $\boldsymbol{\beta}_p^*$ is an 3×1 vector, \mathbf{X}_j is an $n \times k$ matrix ($j=1, \dots, k$ where k is the number of respondents' attributes,) and $\boldsymbol{\beta}_j^*$ is a $k \times 1$ vector. The descriptive statistics and distributions of the explanatory variables are summarized in Table 3 and Table 4.

Besides, keeping significant explanatory variables, equations 3 and 4 are simultaneously estimated by Zellner's seemingly unrelated regression,⁷⁾ assuming the error terms $\boldsymbol{\varepsilon}^D$ and $\boldsymbol{\varepsilon}^{Mo}$ are correlated (models 5 and 6). Because they are in fact correlated with significant correlation coefficient of 0.4959, models 5 and 6 are more reliable than

models 3 and 4.

In Table 2, if the sign of the coefficient of each independent variable are same in models (5) and (6), then the attribute can be considered to prevent the social dilemma. The more moral the people get, the more they relatively choose to “minimize,” and vice versa. The solution for the dilemma in this case would be to produce AVs that suit consumers' moralities, or educate people to be more moral if the citizens would like to make society as a whole more moral. In our data, factors preventing social dilemma are *Age*, *Altruism*, *Male*, *High Education* and *Impatience*. *Age* and *Altruism* have positive effects on *Buy_Minimize*, and negative ones on *Buy_Protective*. This leads to have them positive effects on *Difference*, while they as well have positive effects on *Morality*. *Male*, *High Education*, and *Impatience* works negatively on both *Difference* and *Morality* so that they will buy less moral AVs according to their morality.

It is important to note that the existence of the fellow passengers will not generate the dilemma. When the respondents are assumed to ride with coworkers or families, coefficient of *Buy_Minimize* is lower and *Buy_Protective* is higher than assuming riding alone, which leads to negative coefficients of *Difference*. Together with negative coefficients of *Morality*, their purchasing behavior will not contradict to their morality. If all dependent variables show this tendency, then both of Shariff, et al. (2017)'s suggestions will be effective.

Table 2. Determinants of Intention to Buy each AVs and Their Difference

Dependent Independent	All Variables				SUR	
	(1) Buy Minimize	(2) Buy Protective	(3) Difference	(4) Morality	(5) Difference	(6) Morality
<i>WithCoworker</i>	-7.908 *** (0.695)	4.295 *** (0.710)	-6.102 *** (0.533)	-10.380 *** (0.677)	-6.112 *** (0.529)	-10.39 *** (0.671)
<i>WithFamily</i>	-3.213 *** (0.685)	2.656 *** (0.705)	-2.935 *** (0.527)	-4.683 *** (0.658)	-2.910 *** (0.524)	-4.642 *** (0.665)
<i>Male</i>	2.297 *** (0.610)	6.869 *** (0.624)	-2.286 *** (0.462)	-5.298 *** (0.580)	-2.156 *** (0.461)	-5.163 *** (0.584)
<i>Age</i>	0.201 *** (0.023)	-0.192 *** (0.024)	0.197 *** (0.018)	0.234 *** (0.022)	0.200 *** (0.017)	0.242 *** (0.022)
<i>HighEducation</i>	-3.217 *** (0.609)	1.981 ** (0.624)	-2.599 *** (0.468)	-3.197 *** (0.587)	-2.739 *** (0.462)	-3.547 *** (0.578)
<i>Income (log)</i>	-1.360 ** (0.427)	1.625 *** (0.440)	-1.493 *** (0.333)	-0.805 + (0.430)	-1.138 *** (0.274)	
<i>DislikeShareCars</i>	-2.491 *** (0.653)	-1.567 * (0.681)	-0.462 (0.511)	0.239 (0.639)		
<i>Pride</i>	3.019 *** (0.689)	3.185 *** (0.709)	-0.0833 (0.517)	-0.991 (0.669)		
<i>NoCar</i>	0.0384 (0.732)	1.564 * (0.744)	-0.763 (0.564)	-1.372 + (0.716)		
<i>FavorDriving</i>	-3.806 *** (0.683)	-3.029 *** (0.712)	-0.388 (0.533)	-0.592 (0.676)		
<i>CausedAccidents</i>	-1.708 ** (0.645)	-3.394 *** (0.667)	0.843 + (0.499)	1.040 + (0.630)		
<i>Credibility</i>	0.0217 (0.015)	0.157 *** (0.015)	-0.0677 *** (0.012)	0.049 ** (0.015)	-0.0675 *** (0.011)	0.0478 *** (0.014)
<i>Altruism</i>	5.282 *** (0.336)	-3.789 *** (0.344)	4.535 *** (0.264)	5.226 *** (0.338)	4.524 *** (0.244)	5.141 *** (0.309)
<i>Impatience</i>	-0.138 (0.140)	0.378 ** (0.145)	-0.258 * (0.110)	-0.277 * (0.140)	-0.265 * (0.104)	-0.307 * (0.132)
<i>Constants</i>	24.410 *** (3.117)	34.830 *** (3.237)	44.790 *** (2.445)	43.780 *** (3.178)	42.270 *** (2.113)	38.540 *** (1.850)
AIC	97619.3	98110.3	92159.1	97028.4	186297.9	
R-squared	0.0587	0.0602	0.075	0.0792	0.0744	0.0781

Note: 1. Standard errors in parenthesis

2. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Source: *Survey on Auto Driving*

Table 3. Descriptive Statistics of Variables

Variable	Description	Mean	S. D.	Minimum	Maximum	Population Mean
<i>BuyMinimize</i>	A number selected from the slider, which ranged from “not at all likely (0)” to “extremely likely (100).”	38.8	29.3	0	100	
<i>BuyProtective</i>	A number selected from the slider, which ranged from “not at all likely (0)” to “extremely likely (100).”	41.9	30.1	0	100	
<i>Difference</i>	Each respondent's reply to <i>BuyMinimize</i> subtract that with <i>BuyProtective</i> . (Linearly transformed to 0-100 scale)	48.4	22.7	0	100	
<i>Morality</i>	An action that respondents think is the most moral, selected from the slider ranging from “stay (0)” and “swerve (100).”	59.4	28.8	0	100	
<i>WithCoworker</i>	Respondents who were asked to imagine themselves in the car together with a coworker or acquaintance = 1; 0, otherwise.	0.334	0.472	0	1	
<i>WithFamily</i>	Respondents who were asked to imagine themselves in the car together with a family member = 1; 0, otherwise.	0.322	0.467	0	1	
<i>Male</i>	Male = 1, Female = 0.	0.592	0.491	0	1	0.502
<i>Age</i>	Age scaled by one.	47.7	12.7	18	69	
<i>HighEducation</i>	Respondents who graduated from a university, graduate school, or equivalent = 1; 0, otherwise.	0.586	0.493	0	1	0.245
<i>Income (log)</i>	Logged value of the respondents' annual household income.	6.355	0.704	3.912	8.161	
<i>DislikeShareCars</i>	From the choice of “Shared cars are totally acceptable” (1) to “Shared cars are totally unacceptable” (10), those who chose a score of 6 or higher = 1; 0, otherwise.	0.290	0.454	0	1	
<i>Pride</i>	Ranging from the choice “Do not have pride in owning cars at all” (0) to “Have pride in owning cars very much” (10), those who chose a score of 6 or higher = 1; 0, otherwise.	0.271	0.445	0	1	
<i>NoCar</i>	Respondents who owns cars = 1; 0, otherwise.	0.244	0.430	0	1	
<i>FavorDriving</i>	Among respondents who have a driver's license, those who chose a score of 8 or higher from the choice “Dislike driving very much” (1) to “Like driving very much” (10) = 1; 0, otherwise.	0.279	0.448	0	1	
<i>CausedAccidents</i>	Respondents who had been responsible for causing a car accident = 1; 0, otherwise.	0.317	0.465	0	1	
<i>Credibility</i>	A number selected from the slider, which ranged from “If all the cars in the county become fully auto-driving, I think car accidents will increase very much (0)” and “... I think car accidents will decrease very much (100).”	67.2	20.4	0	100	
<i>Altruism</i>	Principal component of two responses: (A) opinion of the statement “No matter the circumstances that we are in, we should help those who are in need,” ranging from “Strongly Disagree” (1) to “Strongly Agree” (5); and (B) frequency of donations.	3.491	0.895	0.772	6.402	
<i>Impatience</i>	Respondents' self-evaluated character, which ranged from “easy-going (1)” to “impatient (10).”	5.904	2.080	1	10	

Source: *Survey on Auto Driving*

Table 4. Distributions of Selected Variables

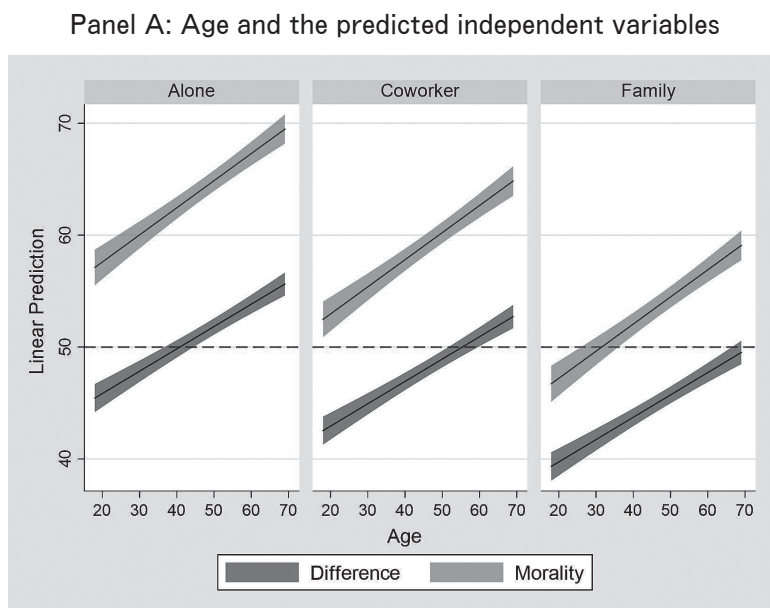
	N	%	Population (%)		N	%
<i>Age</i>				<i>Altruism</i>		
18 - 29	1,134	11.1	15.5	2 > A	351	3.4
30 - 39	1,761	17.2	19.1	3 > A ≥ 2	3,370	32.9
40 - 49	2,479	24.2	23.5	4 > A ≥ 3	3,671	35.9
50 - 59	2,650	25.9	19.1	5 > A ≥ 4	2,464	24.1
60 - 69	2,210	21.6	22.9	A ≥ 5	378	3.7
<i>Income (million Yen)</i>				<i>Impatience</i>		
3 > I	1,327	13.0	34.1	1-2	526	5.1
6 > I ≥ 3	3,563	34.8	31.7	3-4	2,231	21.8
9 > I ≥ 6	2,715	26.5	18.3	5-6	2,917	28.5
12 > I ≥ 9	1,578	15.4	9.1	7-8	3,691	36.1
I ≥ 12	1,051	10.3	7.0	9-10	869	8.5
				<i>Credibility</i>		
				0-24	382	3.8
				25-49	997	9.7
				50	855	8.4
				51-75	4,211	41.2
				76-100	3,789	37.0

Source: *Survey on Auto Driving*

Nevertheless, one factor is found to cause a dilemma. It is *Credibility*: a number selected from the slider, which ranged from “If all the cars in the county become fully auto driving, I think car accidents will increase very much (0)” and “...I think car accidents will decrease very much (100).” As *Credibility* rises, so does *Buy_Protective* (model 1), but people do not show their clear intention to *Buy_Minimize* (model 2). The resulting coefficient of *Credibility* is significantly negative for *Difference*, whereas for *Morality*, it has positive coefficient (models 5 and 6).

Figure 5 exhibits this contrasting effect. The coefficients in Table 2 are average effects, and we can also predict the value of a dependent variable for each value of the independent ones (setting the values of the other independent variables at their average) in Figure 5. Panel A gives the predicted

values (with 95% confidence intervals) of *Difference*⁸⁾ and *Morality*, along with the values of *Age*: one of the variables that may not cause the dilemma. *Difference* and *Morality* go up almost in parallel as *Age* rises. When riding with a family, the *Morality* value of a 20-year-old person is estimated as 46.9 and that of a 69-year-old as 59.0, while *Difference* of a 20-year-old is estimated as 39.5 and a 69-year-old as 49.4. In Panel B, the predicted values are shown according to the values of *Credibility*. Those who think that car accidents will increase very much due to AVs (*Credibility* = 0) have a *Morality* of 50.7 and those who think the opposite (*Credibility* = 100) have that of 55.5. The more people believe in AVs, the more moral they become. Perversely, zero *Credibility* has 49.8 *Difference* and the *Difference* shrinks to 43.1 for those with 100 *Credibility*. It follows that people who think AVs are credible are



Panel B: Respondents' credibility and the predicted independent variables

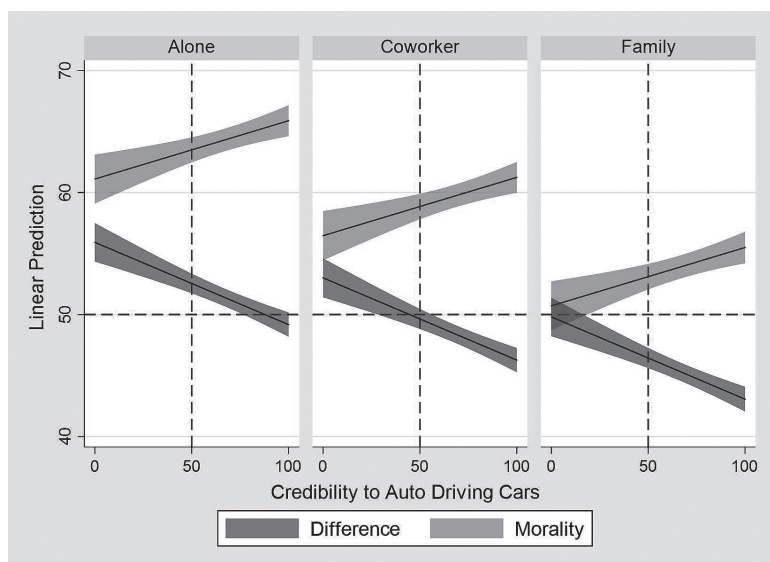


Figure 5. Change in Marginal Effects on Difference and Morality

Source: *Survey on Auto Driving*

relatively less willing to purchase “minimize” AVs, even though they are more moral.

Our results contradict Sharif, et al. (2017)'s first idea that overall accident reduction by AVs may help overcome the social dilemma. When AVs are not credible, Japanese people will avoid deciding both the morality and their preference between two types of AVs (both scoring around 50, especially when riding with families). The reason, we suppose,

is that they believe AVs will increase the number of accidents whatever they think morally for whatever type they choose to buy. As AVs become more credible, they start making judgments that it is more moral to swerve (sparing the 10 pedestrians but killing passengers including themselves), but they would not like to buy AVs that are programmed to swerve.

Interpretation of the behavioral

discrepancies between *Morality* and *Difference* could be twofold. One is that as credibility rises, people (possibly unintentionally) assume the AVs more accurately follow the program, and believe that the “minimize” program will better save pedestrians plus sacrifice themselves. The other interpretation may be that as credibility rises, it will become acceptable for people to purchase self-protecting cars because they can assume the number of such incidents they are to face will decrease thanks to almighty AVs. If these interpretations are right, reducing overall accidents, *ceteris paribus*, will make the social dilemma more serious. The social dilemma may not be solved until any type of AVs will be able to save all pedestrians and all passengers together.

V. Conclusion

AVs that drive fully automatically will have to face the dilemma of choosing between two evils such as running over pedestrians versus sacrificing themselves and their passengers to save pedestrians. How people program AVs to face this dilemma may cause another social dilemma: if people have different ideas between “moral AVs” and “wishing AVs,” even though the latter are immoral, the market will eventually be filled them.

The way to solve this social dilemma is challenging because it is not easy to find the causes. Our study explored them and found credibility to the notion that AVs will eliminate the total number of car accidents — a key factor in Japan. It is thought-provoking that the more credible AVs are, the greater the likelihood that the social dilemma will occur. The social dilemma may not be resolved until the number of car accidents reaches zero, and when neither type of AV requires that

someone be sacrificed.

We are standing at the dawning of the AV society, and fully automotive private vehicles have yet to be brought to the market. However, before the technical barriers to their utilization are finally broken down, we should investigate the factors of the dilemma once again to look into the relationship between our confidence in AVs and our purchasing behavior.

Acknowledgments

We gratefully inform that this research is funded by RIETI research project “Economics of Artificial Intelligence” and partly by JSPS KAKENHI: Grant-in-Aid for Scientific Research (B) [26285057].

References

- Bonnefon, Jean-François, Azim Shariff, and Iyad Rahwan (2016). “The Social Dilemma of Autonomous Vehicles.” *Science* 352 (6293) : 1573-76.
- Foot, Philippa (1967). “The Problem of Abortion and the Doctrine of Double Effect.” *Oxford Review*, 5: 1-5.
- Thomson, Judith Jarvis (1985). “The Trolley Problem.” *Yale Law Journal* 94 (6) : 1395. <https://doi.org/10.1119/1.1976413>.
- Van Lange, Paul A M, Jeff Joireman, Craig D Parks, and Eric Van Dijk (2013). “The Psychology of Social Dilemmas?: A Review.” *Organizational Behavior and Human Decision Processes* 120: 125-141. <https://doi.org/10.1016/j.obhdp.2012.11.003>.
- Shariff, Azim, Jean-François Bonnefon, and Iyad Rahwan (2017). “Psychological Roadblocks to the Adoption of Self-Driving Vehicles.” *Nature Human Behaviour*, 1: 694-696. December. <https://doi.org/10.1038/s41562-017-0202-6>.
- Zellner, A. (1962). “An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias.” *Journal of the American Statistical Association* 57: 348-368.
- _____ (1963). “Estimators for seemingly unrelated

regression equations: Some exact finite sample results.” *Journal of the American Statistical Association* 58: 977-992.

Zellner, A., and D. S. Huang (1962.) “Further properties of efficient estimators for seemingly unrelated regression equations.” *International Economic Review* 3: 300-313.

-
- 1) This is a widely known “Trolley Problem,” on which Foot (1978) originally drew attention, and again discussed by Thomson (1985).
 - 2) The survey was administered in 2016 by Nikkei Research Inc. who offers a representative panel of household in Japan, and funded by the Research Institute of Economy, Trade and Industry, IAA (RIETI).
 - 3) This is a modified scenario used in Study 3 of Bonnefon et al. (2016), and respondents read Japanese translated version.
 - 4) Here, the variable *Difference* ranges from -100 to 100. From the next section on, it will be linearly transformed to take the range from 0 to 100, to make it directly comparable to *Buy_Minimize* and *Buy_Protective*.
 - 5) 240 respondents who chose fifty for *Morality* are not counted here.
 - 6) 649 respondents who chose fifty for *Morality* are not counted here.
 - 7) See Zellner (1962), Zellner and Huang (1962), and Zellner (1963).
 - 8) *Difference* in this section is linearly transformed to take the range from 0 to 100. See Table 3.