

***A SURVEY STUDY MEASURING PEOPLE'S PREFERENCES TOWARDS AUTOMATED
AND NON-AUTOMATED RIDESPLITTING***

Fangda Zhang¹, Shannon Roberts¹, Claudia Goldman²

¹University of Massachusetts Amherst
Amherst, MA, USA

²General Motors Advanced Technical Center Israel
Herzliya Pituach, Israel
Email: fzhang@umass.edu

Summary: Ridesplitting is both common and important as it facilitates daily transportation needs. Alongside an increase in ridesplitting is the introduction of automated driving systems, which together, bring out the possibility of automated ridesplitting. However, previous studies have identified resistance in the acceptance of automated driving systems. In light of past research on automated driving systems, we used a survey to compare people's preferences of automated ridesplitting to non-automated ridesplitting. Statistical and text mining techniques were leveraged to analyze the results. We found similarities in the numeric responses of important factors concerning automated and non-automated ridesplitting whereas there were large differences between automated and non-automated ridesplitting in the text responses. Additionally, people prioritized cost and time in both automated and non-automated ridesplitting. These results can be used in the design of future ridesplitting services, especially with respect to increasing acceptance of and trust in automated ridesplitting services.

INTRODUCTION

Ridesplitting is a service where fares and rides are split among strangers who are traveling in the same direction. This is differentiated from ridesharing, which is the formal or informal sharing of rides between drivers and passengers with similar origin-destination pairings (e.g., carpooling). Dynamic real-time ridesplitting, wherein multiple passengers are matched with one vehicle for a one-time, one-way trip (Amey, Attanucci, & Mishalani, 2011) has increased in popularity as ride-hailing companies like Uber and Lyft offer dynamic real-time ridesplitting to their customers (e.g., Uber Pool or Lyft Line).

Given that ridesplitting services revolve around users, people's attitudes towards them is of interest. Past research has identified which aspects encourage ridesplitting, including costs, personal preference (Thaithatkul, Seo, Kusakabe, & Asakura, 2015), wait time (Deakin, Frick, & Shively, 2010; Hanna, Albert, Chen, & Stone, 2016), safety, and security (Deakin et al., 2010; Thaithatkul et al., 2015). Other researchers (Morales Sarriera et al., 2017) reported that social interactions were important in ridesplitting, as well as riders' prejudice with respect to age, social class, and race. Altogether, these studies have stressed the importance of people's preferences when choosing ridesplitting services.

As time progresses, not only will there be more ridesplits, but ridesplits may take place in automated driving systems. Past research has identified if people would accept the technology

and how user characteristics impact their adoption of the technology (Bansal, Kockelman, & Singh, 2016; Krueger, Rashidi, & Rose, 2016). Although experts and the public are often positive about automated driving systems, they also exhibit essential concerns (Kyriakidis, Happee, & de Winter, 2015). In addition, since automated ridesplitting is still a new concept, and user acceptance is imperative for the successful implementation of this service (Pakusch & Bossauer, 2017), its acceptance by the public is of much importance. Moreover, Level 3+ automated driving systems are a technological innovation that is not yet market-ready and therefore has not yet been tested by users (Pakusch et al., 2017). This unquestionably increases the public's uncertainty about automated ridesplitting, and therefore makes it difficult to have a complete understanding of automation-related concepts, not to mention concrete opinions about the technology. As such, this study was conducted to understand if there were differences in preferences between automated ridesplitting (AR) and non-automated ridesplitting (non-AR).

METHOD

Survey Development and Description

An online survey was developed with 66 questions and took approximately 20 minutes to complete. The survey was developed and administered using Qualtrics. Survey questions concerning people's ridesplitting usage and perception of different transportation modes were formulated based on previous studies (Abraham et al., 2016), (Ciari & Axhausen, 2013), (Dias et al., 2017), (Morales Sarriera et al., 2017), (Thaithatkul et al., 2015). Past research highlighted the importance of time, cost, and gender in choosing a ridesplit (Hanna et al., 2016; Morales Sarriera et al., 2017; Thaithatkul et al., 2015), hence additional survey questions were included to compare these aspects. Since the main goal of this study was to identify differences in people's preferences towards AR and non-AR, a set of important statements (factors) that might affect people's preference were used (Ciari & Axhausen, 2013), and participants were asked to rate them in both AR and non-AR settings. Demographic questions were also included in the survey, however, they are not analyzed here.

Procedure

Participants were recruited from Massachusetts and Connecticut by one of three means: posted flyers, email advertisements, and Facebook advertisements. Participants indicated their interest in the survey via email. A researcher screened them to ensure they were at least 18 years of age. After screening, participants were sent a link to an online consent form and the survey. After survey completion, they were compensated with a \$10 Amazon gift card.

Participants

A total of 117 participants responded to the survey. Most participants (75.2%) were between 18-35 years old and were female (69.2%) versus male (27.4%). Most participants (77.78%) were from urban areas, 21.37% were from suburban areas and only 0.85% were from rural areas. More than half of participants were either employed full time (37.6%) or a student (40.2%). Forty one percent of participants did not have a car whereas 59% of participants did own a car. Most participants had a driver's license (82.9%) and 79.5% thought they embraced a technology-

driven lifestyle. As for ridesplitting usage, 18.8% of the participants said they never used it before whereas 17.2% used ridesplitting at least 1 day per week. An additional 27.3% did not use ridesplitting within the past 30 days and 36.7% used ridesplitting monthly.

Data Analysis

Numeric and Multiple-choice Questions. For numeric and multiple-choice questions, we used the R statistical programming language to generate the results (e.g., percentage of response, average ratings, etc.). Chi-squared tests were used to investigate if participants' preferences towards ridesplitting in the format of multiple-choice questions differed between AR and non-AR. Additionally, the dependent *t*-test (paired sample *t*-test) was employed to compare the difference in participants' numeric responses to important factors that might influence their preferences towards AR and non-AR. We used an alpha level of 0.05 for statistical tests.

Open-ended Text Questions. We used text-mining to explore the open-ended text questions using the R statistical programming language. There were two open-ended questions: *ABQuestion A* was asked after a numeric question that required participants to rate the importance of 29 statements regarding AR and non-AR (e.g., saving on expenses). Hence, the "other" referred to aspects that were not mentioned in the 29 statements. *Question A* was not required, and we only received 73 responses from 117 participants whereas *Question B* was required and we received responses from all 117 participants.

Similar to previous research (Ghazizadeh, McDonald, Lee, & Madison, 2014; Mehrotra & Roberts, 2018; Roberts & Lee, 2014), to begin data preprocessing, text answers were imported as a corpus data structure and punctuations and numbers were removed. Then, all text was converted to lowercase. Next, we defined and removed stop words, such as "a", "the", and "this", which appear more frequently than other words yet do not provide useful information. The last step was to reduce the words to stems, which reduces variance in the data by converting words to their radical. After processing, we extracted the key words for each question to investigate words that appeared most frequently. The results were visualized by word clouds.

As an exploratory measure, we also performed latent semantic analysis (LSA) to examine the textual data. The underlying idea in this technique is that the aggregate of all the word contexts in which a given word does or does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other (Landauer, Foltz, & Laham, 1998). In other words, it helps us analyze relationships among documents and terms. The key insight in LSA is to reduce the dimensionality of the information retrieval problem (Dumais, 2005). Singular-value decomposition (SVD) is often used as a data reduction method and was leveraged in our study to reduce the dimension of the term-document matrix (where the data were stored) and help extract desired relationships. Term similarities were obtained by calculating the cosine value (Ghazizadeh et al., 2014) between terms in the reduced vector space obtained by SVD. We used heat maps to present the results.

RESULTS

Numeric and Multiple-choice Questions

Most participants (AR: 64.1%, non-AR: 74.4%) preferred to ride split with more passengers at a lower cost over less passengers at a higher cost (AR: 22.2%, non-AR: 17.9%) and a small proportion felt indifferent (AR: 13.7%, non-AR: 7.7%). The difference between AR and non-AR was not significant ($\chi^2(2, N = 117) = 2.9093, p = 0.2335$). More than half of participants (AR: 59.0%, non-AR: 52.1%) preferred a shorter ridesplitting trip with dissimilar passengers than a longer ridesplitting trip with passengers who were similar to them (AR: 18%, non-AR: 21.4%). The remaining participants did not have a preference (AR: 23%, non-AR: 26.5%). No significant difference was found ($\chi^2(2, N = 117) = 0.9694, p = 0.6159$) between AR and non-AR. The majority of participants (AR: 62.4%, non-AR: 68.4%) were more likely to wait a shorter amount of time for a ridesplit where the passengers were not similar to them, instead of waiting a longer amount of time for a ridesplit where the passengers were similar to them (AR: 14.6%, non-AR: 11.9%), whereas some participants did not have a preference (AR: 23.0%, non-AR: 19.7%). The difference between AR and non-AR setting was not significant ($\chi^2(2, N = 117) = 0.8054, p = 0.6685$). Last, a longer ridesplitting trip for less money was more preferred (AR: 58.1%, non-AR: 57.3%) than a shorter ridesplitting trip for more money (24.8% for both AR and non-AR). A small percentage of participants did not prefer either option (AR: 17.1%, non-AR: 17.9%). The difference between AR and non-AR in people's preferences was not significant ($\chi^2(2, N = 117) = 0.0238, p = 0.9982$).

Table 1. Average ratings and standard deviations of important factors for both non-automated and automated ridesplitting with dependent t-test p values

Average ratings ^a and Standard Deviation, Dependent <i>t</i> -test result			
Statement	Non-automated	Automated	<i>p</i> -value
Safety risk with respect to the number of passengers	3.6 (1.27)	3.7 (1.27)	0.1753
Being paired with an unpleasant passenger(s)	3.6 (1.27)	3.5 (1.37)	0.1863
Little chances to find somebody matching my trip requests	3.2 (1.21)	3.0 (1.30)	0.0519
Possibility for rating for ride-mate	2.8 (1.32)	2.9 (1.26)	0.2173
Positive social aspect (e.g., riding with other people)	2.8 (1.25)	2.7 (1.26)	0.5436
Intimidation from being in a rideshare with multiple passengers	2.7 (1.37)	2.8 (1.42)	0.7591
Restriction to certain users (i.e., colleagues from work)	2.7 (1.27)	2.7 (1.29)	0.4333
The norms of interaction between passengers are unclear	2.7 (1.23)	2.6 (1.25)	0.5162
Do not want to be in the car with strangers	2.5 (1.26)	2.6 (1.31)	0.1362
Not having enough information about the other passengers	2.5 (1.28)	2.6 (1.34)	0.2399
Having mobile phone number of rider/mate	2.3 (1.34)	2.2 (1.31)	0.1969
Preferences for gender of ride-mate	2.1 (1.35)	2.1 (1.37)	0.8095

^a 1 represents not important at all while 5 means significantly important

Table 1 refers to participants' average ratings of important statements concerning both AR and non-AR. Participants rated the statement "Safety risk with respect to the number of passengers" the highest and "Being paired with an unpleasant passenger(s)" as the second highest. "Preferences of gender of ride-mate" and "Having mobile phone number of rider/mate" seemed to be the least important statements. Dependent *t*-test was performed on each pair of the questions and the resultant *p*-values are shown in Table 1. There was no significant difference between AR and non-AR in any statement.

Open-ended Text Questions

Figure 1 shows the most frequent terms in word cloud format. For *Question A*, “driver”, “safety”, “passenger”, “time”, “route”, and “music” were the most common words. Figure 2 (a heat map showing the most associated words) indicates “map”, “efficiency”, “waste”, “route”, and “time” are closely associated and are therefore clustered into one group, which might be one concept (the context in which those words tend to appear together). The height of the dendrogram represents dissimilarity between clusters: the more distant the line is, the more dissimilar the clusters. For example, “map” and “efficiency” share the highest similarity and are clustered into one group while “route” can be further combined with this cluster to form an even larger group.

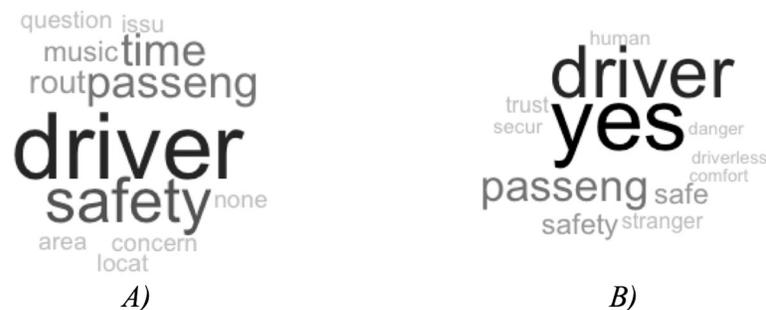


Figure 1. Word cloud displaying frequently mentioned terms in the open-ended text questions; words that are larger and darker represent those with higher frequency

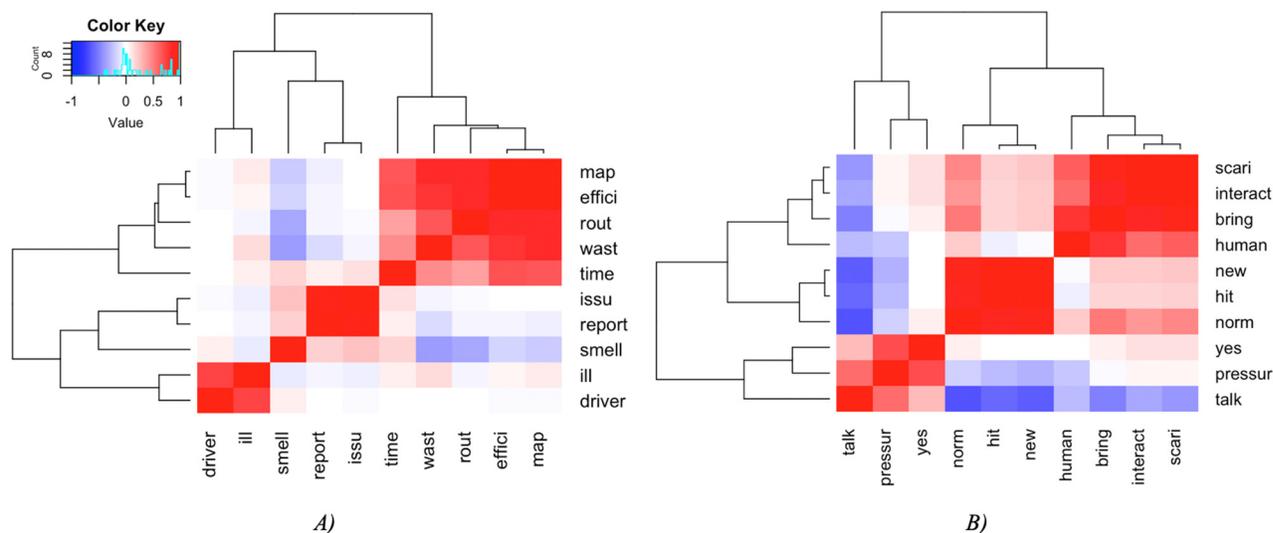


Figure 2. Heat map displaying term associations; red cells indicate similarity between terms and blue means dissimilarity, darker cells represent more (dis)similarity; terms connected by lines are in the same cluster

For *Question B*, the highest frequency term is “yes”, indicating that participants thought there was a difference between AR and non-AR ridesplitting. The word cloud indicates the concern mainly exists with respect to “driver”, “passenger”, “safety”, and “stranger”. The heat map further indicates that people may feel “scared” about the new “interaction” between “humans”.

They perceived AR as a “new hit” and the new “norm”. Lastly, differences between AR and non-AR can be seen with “pressure” to “talk” to each other.

Table 2 presents responses from participants corresponding to some of the most frequently mentioned words and highly associated terms.

Table 2. Example responses corresponding to frequently appearing words and highly associated terms

Question	Frequently Appearing Words	Example Response
A	passengers, driver, safety, concern	<i>An automated ride share is risky I feel for females if the other passengers is male. With no driver there is a big safety concern.</i>
B	yes, driverless	<i>Yes, I would not feel comfortable riding in a driverless car in any circumstances currently.</i>
	Highly Associated Terms	Example Response
B	yes, pressure, talk	<i>Yes, less pressure to talk.</i>
A	route, time	<i>Efficient mapping of route so you don't waste time.</i>

CONCLUSION

This study used a survey to understand if there were differences in people’s preferences between AR and non-AR. In the multiple-choice questions, participants’ ridesplitting priorities (for both AR and non-AR) were centered on cost, wait time, and trip length, not on with whom they split the ride. No significant differences were found in participants’ preferences concerning important factors of ridesplitting between AR and non-AR in the multiple-choice questions.

Driver, passenger, time, route, and music were the most frequently mentioned important factors in ridesplitting for *Question A*. Additionally, map, efficiency, waste, and time were closely related, indicating people’s considerations about time and ride efficiency. When asked if it would make a difference if the car was fully automated when splitting a ride with multiple passengers (*Question B*), the most frequent answer was yes, which suggests participants’ different preferences towards AR and non-AR in the open-ended questions. The main concerns were: the importance of having a driver, how to interact with passengers, safety issues, what the “new norm” would be, the pressure of talking with passengers, potential dangers, and trust.

To conclude, there were small differences in people’s attitudes between AR and non-AR in the numeric responses, but large differences in the text responses. Participants prioritized cost and time when considering ridesplitting, no matter if the vehicle was automated or non-automated. This further supports previous findings (Deakin et al., 2010; Thaithatkul et al., 2015; Hanna et al., 2016). Future studies should be conducted with a larger population from a variety of geographical regions and should use new methods to ensure consistency, accuracy, and validity of participants’ responses. Given the rapid pace at which automated driving systems are being developed, future work can also give participants a clearer picture with respect to how AR will look and feel, thereby allowing participants to make better judgements concerning the difference between this advanced technology and conventional vehicles.

REFERENCES

- Abraham, H., Lee, C., Brady, S., Fitzgerald, C., Reimer, B., & Coughlin, J. F. (2016). Autonomous Vehicles, Trust, and Driving Alternatives: A survey of consumer preferences, (May).

- Amey, A., Attanucci, J., & Mishalani, R. (2011). “Real-Time” Ridesharing – The Opportunities and Challenges of Utilizing Mobile Phone Technology to Improve Rideshare Services, 1–17.
- Bansal, P., Kockelman, K. M., & Singh, A. (2016). Assessing public opinions of and interest in new vehicle technologies: An Austin perspective. *Transportation Research Part C: Emerging Technologies*, 67, 1–14.
- Ciari, F., & Axhausen, K. W. (2013). Carpooling in Switzerland: Public Attitudes and Growth Strategies. *Transportation Research Board 92nd Annual Meeting*, (13–2780).
- Deakin, E. A., Frick, K., & Shively, K. M. (2010). Markets for dynamic ridesharing? Case of Berkeley, California. *Transportation Research Record*, 2187(1), 131-137.
- Dias, F. F., Lavieri, P. S., Garikapati, V. M., Astroza, S., Pendyala, R. M., & Bhat, C. R. (2017). A behavioral choice model of the use of car-sharing and ride-sourcing services. *Transportation*, 44(6), 1307–1323.
- Dumais, S. T. (2005). Latent semantic analysis. *Annual Review of Information Science and Technology*, 38(1), 188–230.
- Ghazizadeh, M., McDonald, A. D., Lee, J. D., & Madison, W. (2014). Text Mining to Decipher Free-Response Consumer Complaints: Insights From the NHTSA Vehicle Owner’s Complaint Database, (2010).
- Hanna, J. P., Albert, M., Chen, D., & Stone, P. (2016). Minimum cost matching for autonomous carsharing. *IFAC-PapersOnLine*, 49(15), 254–259.
- Krueger, R., Rashidi, T. H., & Rose, J. M. (2016). Preferences for shared autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 69, 343–355.
- Kyriakidis, M., Happee, R., & de Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour*, 32, 127–140.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2–3), 259–284.
- Mehrotra, S. K., & Roberts, S. C. (2018). *Identification and Validation of Themes from Vehicle Owner Complaints and Fatality Reports using Text Analysis* (No. 18-05457).
- Morales Sarriera, J., Escovar Álvarez, G., Blynn, K., Alesbury, A., Scully, T., & Zhao, J. (2017). To Share or Not To Share: Investigating the Social Aspects of Dynamic Ridesharing. *Transportation Research Record: Journal of the Transportation Research Board*, (2605), pp 109–117.
- Pakusch, C., & Bossauer, P. (2017). User Acceptance of Fully Autonomous Public Transport. In *ICE-B* (pp. 52-60).
- Roberts, S. C., & Lee, J. D. (2014). Deciphering 140 Characters: Text Mining Tweets On #DriverDistraction, (June 2013), 2195–2199.
- Thaithatkul, P., Seo, T., Kusakabe, T., & Asakura, Y. (2015). A Passengers Matching Problem in Ridesharing Systems by Considering User Preference, *II*, 1416–1432.