Algorithm Aversion and Self-driving Cars

Yuze Kang

University of Toronto
yuze.kang@mail.utoronto.ca

ABSTRACT

Algorithm aversion is the phenomenon that humans tend to resist using algorithms for assisting decision-making, and this aversion has affected the application and spreading of autonomous vehicles, which deploy algorithms. This paper will at first shortly review the researches about algorithm aversion, by mainly referring to the literature review by Burton, Stein and Jensen's in 2020, to address three causes and general solutions for autonomous vehicles industry to alleviate algorithm aversion, the case of Tesla company will occasionally be taken as an example to make detail explanation. Then the article suggests several policies that self-driving car companies can adopt to alleviate influences such as possible sales decline and customers' complaints brought by algorithm aversion, then propose an experiment to verify if these policies are reasonable and feasible, and will finally discuss the limitations of the proposal and this paper and some works that can be done in the future.

Keywords: *Algorithm Aversion, Self-driving car, behavioral economics, experiment proposal.*

1.INTRODUCTION

People are not generally holding an optimistic attitude towards autonomous vehicles now. In China, in the first half of 2021, several incidents reported on Tesla's automobiles have triggered a storm on social media [1], apart from complaints and criticisms towards Tesla's response, many people expressed their worries and antipathies about autonomous vehicles. Likewise, a survey in 2017 has displayed that 78% of Americans feel unsafe to take a ride in a self-driving car [2]. In fact, not only for autonomous vehicles, technological products that use algorithms in their auto-pilot systems, people tend to show an observable repulsion when seeking help from algorithms for decision-making, they tend to value advice from human advisors more, this phenomenon is called algorithm aversion. But algorithms, like those applied in self-driving cars, are actually capable of avoiding dangers and most errors that humans would make [3]. In Kalra and Groves' report in 2017, even introducing autonomous vehicles that are only 10% safer than an average human driver can save a tremendous number of lives [4]. Therefore, either for self-driving car companies or for road traffic safety, if autonomous vehicles become more acceptable to people and thus applications of these cars can be promoted, relieving the influence of people's aversion to the self-driving car is demanded.

2.LITERATURE REVIEW: CAUSES AND SOLUTIONS, AND THEIR MANIFESTATION IN AUTONOMOUS VEHICLES INDUSTRY ON THE SIDE OF CONSUMERS

Algorithm aversion refers to the phenomenon that people, including experts, feel reluctant and hold resistance against applying algorithms in the process of decision-making and forecasting [5]. Such phenomenon can be observed everywhere in this time when algorithms and artificial intelligence based technologies are deployed in nearly every field and group of people, like, in automobiles customers. But studies have shown that algorithms consistently have better performance than humans do, even when it comes to experienced experts with more information input than algorithms [6].

Therefore, to find causes of this algorithm aversion, rather than giving some feasible hypotheses or thoughts, in 2015, Dietvorst, Simmons and Massey performed experiments to discover some empirical evidence [5]. They confirmed their hypothesis from the results of the experiments, that is, seeing algorithms make mistakes can actually reduce people's confidence in algorithms, even people are aware of the fact algorithms perform better than humans [5].

This can be demonstrated more clearly by taking the case of Tesla as an example. In 2021, on April 19, a woman climbed a top of a Tesla car at the Shanghai Auto Show, to protest that a brake failure in her Tesla Model 3 caused an accident and nearly killed 4 members of her family [1]. Reports of this event soon triggered a storm in social media, apart from criticisms towards Tesla's response (the response was rather unsatisfying and controversial) and complaints from other customers who suffered similar issues, obviously, people expressed their worries and antipathies about either taking a ride in a self-driving car or the sharing road with one. Also research said that this incident caused Tesla's new orders to decline by 50% in the following weeks [1]. Although it could not be denied that the fall in confidence for autonomous vehicles and in sales are certainly associated with and amplified by the controversy on social media, this accident itself would have driven away from a part of potential buyers for self-driving cars. And in fact, Tesla has always been posting safety reports of their cars on its official website [7], along with a comparison to automobiles without an auto-pilot system. The accident rate is far below the average indeed, but still, this also does not help increase public's trust in autonomous vehicles and sales after the accident being reported.

Burton, Stein and Jensen in 2020 reviewed systematically on literature between 1950 and 2018 that involve algorithm aversion as a topic, categorized them into 5 themes and concluded 5 problems, or to say 5 causes of why people apply algorithms for decision-making inappropriately (either over-use or under-use) out of these themes with corresponding general solutions [8]. The author will then mainly refer to the structure of this review, demonstrate each cause and solution, and then will explain how they could manifest in autonomous vehicles industry and finally suggest some specific policies for the industry according to the general solutions.

2.1. Cause 1

False expectations, as the very first problem, is associated with human decision maker [8]. The review by Burton et al. concluded that people would generate ideas about an algorithm's capabilities before implying it for decision-making. For example, experts with rich experiences in their domain would probably reject using algorithms as aid for decision [9]. Human decision-makers tend to be influenced by this previously formed expectations when it comes to use algorithms. For self-driving car companies, this indicates they might confront customers that are experienced drivers who are over-confident with their driving safety or driving skills, such customers could pay less interest in the auto-pilot systems [3]. Or, as people tend hold onto a belief that algorithms are likely to be biased, and humans are likely to make random errors and able to attain perfection [8,10], there could be customers who recognize autonomous vehicles as rigid and difficult-to-control machines and prefer normal automobiles. If consumers are unfamiliar with the algorithms deployed in autonomous vehicles, they would be less likely to choose these cars [8,11].

2.2. Solutions

Building human decision makers' algorithmic literacy would be effective in combating false expectations, that means, decision makers (humans) are supposed to be informed with sufficient information about an algorithm and its method of interaction [8]. For potential buyers of self-driving cars, if they can learn to have a general idea of how the auto-pilot system works, and are trained to interact with the system, they would gain more trust in autonomous vehicles [11]. Also, if consumers can be more familiar with basic statistical concepts, there might be a greater chance that they choose autonomous vehicles, as studies have shown that even when algorithms and human advisors have the same rate of error, people tend to choose humans over algorithms, if consumers appreciate algorithms' better performance and become tolerant with unavoidable errors of algorithms (by gaining basic statistical knowledge), they might make their choice differently [8,5]. Studies of Dietvorst et al. in 2015 also found that if people did not see algorithms err, they would be more likely to choose algorithms. But if any incidents ever happened to a self-driving car, people will be able to know very soon as this would more attention-grabbing than a normal car accident. The incident of Tesla certainly proves this. Therefore for companies, trying to prevent customers from knowing errors and accidents are not plausible.

2.3. Cause 2

The second cause that the review of Burton et al. Suggested is lack of decision control [8]. If people lose a sense of control while an auto-pilot system controls the car they are in, then the chance they purchase the car would decrease. In the case mentioned above, the protester claimed that her car suddenly had a brake failure and cannot switch back to manual mode, whether or not she told the truth (the incident remains controversial [1]), it is difficult for people to trust Tesla after hearing this. To trust an algorithm, people need enough sense of control, which could come from enough understanding of algorithms, or, from the modification on algorithms they could make [8,10]. Studies by Dietvorst et al. in 2016 demonstrated that this kind of modification does not have to actually influence the algorithm's operation and people are rather insensitive to the size of the change they could make.

2.4. Solutions

Dietvorst et al. showed in their experiments in 2015 and 2016 that, human decision makers would show reluctance to use an algorithm as soon as they observed errors, but would regain their willingness if given the opportunity to modify this ‘imperfect’ algorithm [5,10]. So it is suggested that decision-making should be ‘Human-in-the-loo’,where human decision-makers are able to supervise and intervene, without affecting normal operation of algorithms (even steps without real response can restore people's willingness and alleviate algorithm aversion)[8].

2.5. Cause 3

The third cause suggested was lack of incentivization [8]. Researchers argued that to combine human decision maker and algorithmic decision, to make an augmented decision making, motivation (incentives) is required [12]. The review concluded two types of incentives, one is economic, and the other is social. The economic incentives indicates motivation caused by money. The social incentives indicates motivation caused by conformity to social norms [8]. In fact, research showed that being aware of other people had adopted an algorithm will increase the probability an individual chooses the algorithm to make decision [13]. The effect of economic incentives is rather complex, which needs further empirical experiments to make sure [8].

2.6. Solutions

Although effects of these incentives are rather unsure and context-based, nudging decision makers by framing contexts or other measures from behavioral economics can effectively promote the use of algorithms. Similarly for autonomous vehicle companies, in their propaganda and work of public relations, if customers could be properly nudged, they may find proper incentives to purchase self-driving cars.

3. POLICY IMPLICATIONS

According to the solutions demonstrated above, more specific policies can be suggested for each cause.

3.1. False expectations

To provide customers with sufficient information, companies could regularly invite some customers and media to factory tours to generally explain how the algorithms work to operate the car, the advantages and limitations of self-driving cars, this happens to be what Tesla has been doing [1]. Other than posting safety reports on official website (another measure Tesla has adopted), which emphasize how automobiles with auto-pilot systems outperforms others in safety, companies should consider to illustrate the fact that algorithms could also err and cause extreme small number of accidents, in greater proportion of propaganda, to increase the public's tolerance.

3.2. Lack of decision control

Autonomous car companies can improve customers' sense of control by adding some external modifiable features to their system, or increasing the sense of ritual whenever the system asks the driver for permission to change settings. Also they could consider perfect the process of switching mode (auto-pilot to manual, mutually), not only does this increase driver's sense of control, but also can prevent horrible accidents when autopilot system unfortunately errs.

3.3. Lack of incentives

On the aspect of social incentives, as people pay much attention to what algorithms others may utilize, companies can invite celebrity with high credibility to stand for their autonomous vehicles. Due to the authority effect, the probability that people would follow to purchase might be greater. Moreover, companies could offer current customers rewards and encourage them to publicize self-driving cars to people around them.

4. EXPERIMENT PROPOSAL

To verify whether these measures can effectively alleviate algorithm aversion for self-driving cars, the author decides to propose an experiment.

4.1. Method

4.1.1. Participants & Design

Participants in the experiments are going to be individuals who have need and willingness to purchase a car, this is to ensure participants have similar mental status towards car-purchasing.

Participants are going to be randomly divided into groups and primed with different causes of algorithm aversion to control variables, because participants may possess innate amount and causes of algorithm aversion. The Independent Variable (IV) should be whether a participant receives treatment of corresponding policy after being primed with a cause. The Dependent Variable (DV) should be the amount of algorithm aversion participants' possess towards autonomous vehicles after treated by corresponding measures. The amount of algorithm aversion would be represented by two ratings, perceived safety and willingness to buy, the higher the two ratings, the lower amount of algorithm aversion people possess.

4.2. Procedure

First of all, participants need to read the consent form and sign for agreement to take part in this experiment. Participants are then randomly assigned to five groups, A,B,C,D,E. All participants need to first fill out a survey (Part I) to report their perceived safety and willingness-to-buy for a self-driving car. Group A participants are only required to do survey Part I. Group B participants are assigned a fake objective report of an accident of a self-driving car (Reading Material (1), RM(1)) . For participants who are assigned to Group C, D, E, they are then equally divided into two subgroups, C1, C2, D1, D2, E1, E2. Apart from RM(1), fake complaints eliciting auto-pilot system’s rigidity and human drivers’ ability to be flexible are assigned to Group C, fake complaints eliciting the lack of control from drivers to autonomous vehicles are assigned to Group D, and fake article eliciting a probable status quo that autonomous vehicles are not so popular across society is assigned to Group E. Additionally for the second subgroups (C2, D2, E2), C2 participants are assigned a brief Tesla safety report with data [2], interpretation of data, and explanation which can raise tolerance of errors made by algorithms. D2 participants are assigned fake news that the system is improved and the switch of control is more flexible, and customers now can make more modifications on system by their personal habits. E2 participants are assigned fake news that an authority recommends people to purchase self-driving cars and researches shows that self-driving cars are in fashion. Participants will then complete survey Part II to report ratings for the same topics asked in Part I. Finally, experimenter should debrief to all participants.

In the following statistical analysis, first overview all data collected from survey part I, then conduct paired-t test with data collected from survey part II on A and B, C1 and C2, D1 and D2, E1 and E2, for perceived safety (PS, the first rating) and willingness to purchase (WP, the second rating).

5. INTERPRETATION OF EXPECTED RESULTS & DISCUSSION

The overall opinion for autonomous vehicles before experiment are supposed to be relatively low, as currently the public's acceptance and utilization of self-driving cars are still low [3]. The assumed results, for group A and group B, should confirm the research of Dietvorst et al., that people's algorithm aversion would rise after seeing algorithms err [5]. Result for E1 and E2 should indicate that the solution of raising social incentives and the measure of nudging is effective, as people tend to conform with other in adopting algorithms [13]. The measures aimed at decision control and false expectations are assumed to have minor or no influence, because firstly, researches have shown that even individuals equipped with algorithm literacy cannot avoid bearing aversion while utilizing algorithms [5,8], and secondly, the sense of control elicited by a piece of reading material could be low and thus result a minor effect on people's opinion.

There certainly would be several limitations and improvements can make for this experiment. At first, the participants of the experiment might not be representative. As mentioned above, before the experiment, participants will have uneven mental statuses, some of them may be advocates of self-driving cars, others may previously suffered from issues that self-driving cars have. Although the experiment design takes step to smooth these differences, the effect might not be obvious. This might lead to biased results. Secondly, for Cause 2, only be reading materials might not be sufficient to provide people with sense of control. If people are able to experience this in person on a self-driving car's auto-pilot system, their algorithm aversion could be alleviated evidently, as many studies have found [8]. Moreover, the experiment does not design a test for alternating economic incentives, because people's opinions are gathered based on no financial considerations. Also, current researches have not been sure about the proportional relationship between economic incentives and algorithm aversion [8]. Hopefully in the near future studies can figure this issue out and apply the result to alleviate algorithm aversion.

6. CONCLUSION

This paper focused on studying three of five causes of algorithm aversion concluded by the review of Burton, Stein and Jensen to develop some practical solutions, because the review studied the causes of algorithm aversion for all human decision-makers, but this paper hoped to help companies to relieve customers' algorithm aversion where the author thought as a narrower domain. Probably, by studying the other two causes, combating intuition and conflicting concepts of rationality [8], might help to develop more policies that autonomous vehicle companies can take a try. Nevertheless, algorithm aversion is not the only ‘roadblocks' to the application of self-driving cars [11,14]. Risk heuristics and ethic issues are also influencing the public's acceptance and utilization of these cars [11,14]. And, by the time of writing this paper, due to the pandemic of COVID-19 and lack of sources it is still difficult to conduct the experiment proposed in this paper, to verify if the policies suggested are effective. Hopefully further studies can be enlightened and look into more aspects to find a more comprehensive answer for the widespread of autonomous vehicles.

REFERENCES

[1] Campbell, M. (2021). Tesla's Fall From Grace in China Shows Perils of Betting on Beijing. Bloomberg Businessweek. https://www.bloomberg.com/news/features/2021-07-05/tesla-s-fall-from-grace-in-china-shows-perils-of-betting-on-beijing.

[2] Edmonds, E. (2017). Americans Feel Unsafe Sharing the Road with Fully Self-Driving Cars. Newsroom.

[3] Shariff, A., Bonnefon, JF., Rahwan, I. (2021). How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars, Transportation Research Part C: Emerging Technologies, Volume 126, 103069. https://doi.org/10.1016/j.trc.2021.103069.

[4] Kalra, N., Groves, D.G. (2017). The enemy of good: Estimating the cost of waiting for nearly perfect automated vehicles. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR2100/RR2150/RAND_RR2150.pdf.

[5] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114-126. https://doi.org/10.1037/xge0000033.https://newsroom.aaa.com/2017/03/americans-feel-unsafe-sharing-road-fully-self-driving-cars/

[6] Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical Versus Clinical Data Combination in Selection and Admissions Decisions: A Meta-Analysis. Journal of Applied Psychology, 98(6), 1060-1072. https://doi.org/10.1037/a0034156.

[7] Tesla Vehicle Safety Report. https://www.tesla.com/VehicleSafetyReport.

[8] Burton, J. W., Stein, MK., Jensen, T. B.(2020). A systematic review of algorithm aversion in augmented decision making. J Behav Dec Making. 2020: 33: 220-239. https://doi.org/10.1002/bdm.2155.

[9] Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. Organizational Behavior and Human Decision Processes, 37, 93-110. https://doi.org/10.1016/0749-5978(86)90046-4.

[10] Dietvorst, B.J., Simmons, J. P., Massey, C. (2016) Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. Management Science 64 (3), 1155-1170, https://doi.org/10.1287/mnsc.2016.2643.

[11] Shariff, A., Bonnefon, JF. & Rahwan, I. (2017) Psychological roadblocks to the adoption of self-driving vehicles. Nat Hum Behav 1, 694–696. https://doi.org/10.1038/s41562-017-0202-6.

[12] Rex V. Brown, Decision science as a by-product of decision-aiding: A practitioner's perspective, Journal of Applied Research in Memory and Cognition, Volume 4, Issue 3, 2015, Pages 212-220, ISSN 2211-3681, https://doi.org/10.1016/j.jarmac.2015.07.005.

[13] Alexander, V., Blinder, C., Zak, P. J. (2018). Why trust an algorithm? Performance, cognition, and neurophysiology, Computers in Human Behavior, Volume 89, 279-288. https://doi.org/10.1016/j.chb.2018.07.026.

[14] Kahl, B. (2021). Human Biases Preventing The Widespread Adoption Of Self-Driving Cars. Cornell University. https://arxiv.org/abs/2104.01022.