The Effect of Message Framing and Timing on the Acceptance of Artificial Intelligence's Suggestion

Abstract

AI helps us make decisions in various domains such as healthcare, finance or entertainment (e.g. Netflix, IBM Watson and etc.). However, people's trust and acceptance of AI are highly susceptible to when and how the suggestion is presented. This study examined the role of the message framing and timing on acceptance when the performance of AI is stated. The study employed a 2 (message timing: before vs. after decision) x 3 (message framing: no information vs. negative framing vs. positive framing) between-subjects experiment where participants were told to solve the specific problem with AI in different conditions. The results showed that participants perceived the suggestion of AI more reasonable and accepted it more when the performance is not stated than any information is provided and they perceived the suggestion of AI more reasonable when the message is presented before the decision is made. The theoretical and practical implications are discussed.

Author Keywords

Artificial Intelligence; Acceptance; Framing Effect; Message Timing

CCS Concepts

•**Human-centered computing** → **Empirical studies in HCI;** *Empirical studies in interaction design;* Laboratory experiments;

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.3383038

Introduction

Artificial Intelligence (AI) with machine learning and deep learning algorithms are now increasingly employed in various domains such as healthcare [10], finance [11] and entertainment [7], which help us make decisions on a daily basis. For instance, IBM's Watson has been adopted in healthcare and contributed to clinical decision support since 2013 [24]. It presents several treatment options with some explanations for supporting doctors to diagnose. Also, Netflix, one of the leading online media-services provider company also actively utilizes the recommendation algorithm to help the viewers to decide what content to watch [7].

Whether it is a recommender system or a decision support system for experts, the output of the system needs to be explained and provided well in order to gain trust [8]. The process of explanation in an AI system is, however, constrained to the human cognitive processes and biases [20]. Thus, it is crucial to provide the explanation and suggestion in a way that humans think and communicate to make people trust and accept the suggestion of the AI.

For instance, the way people think and behave is largely influenced by how the message is framed [22], and it is even true for the experts [18, 5]. Thus, when the AI system helps users, it is important to present the suggestion of AI in a way that does not harm the trust of the system.

In addition, message presentation timing might also be important for the AI system to gain trust. When people feel ownership of their decisions, they feel reluctant to change them since it is considered as loss [12, 13]. Thus, it is also crucial for AI decision support to intervene right moment of the time for users to accept the suggestion more.

Therefore, it is important to find a way to present the suggestion in an appropriate way and provide it at an appropriate time for the suggestion of AI to work well. This study examined the role of the framing effect [22] and the message presentation timing in order to find the optimal way of providing the suggestion of the AI decision support system.

To investigate this, the between-subjects experiment was conducted with different wordings of framing and timing. The results showed that people perceive the suggestion of AI more reasonable and accepted it more when no information about the performance of AI was stated than any wording about the performance information was stated. The results also showed that people consider the suggestion of AI more reasonable when the message is presented before the participants made the decision than it is presented after they made the decision. This study makes some contributions by providing some guidance about the way the message should be presented for making AI more acceptable and the theoretical background behind it.

Backgrounds

Transparency of the AI

AI and Human-Computer Interaction (HCI) researchers emphasize the importance of explainability and transparency of the technology for AI to gain trust [8]. However, lots of scholars question the feasibility and effectiveness of making AI transparent and explainable [3, 1, 19] since the process of achieving the output is hard to interpret even for the experts [9]. In addition, even though the engineers achieved to explain how the algorithms work, laypeople might not understand such a complex explanation [1]. Hence, it might be more informative and practical for users to provide solely the performance of an algorithm when they make decisions supported by the AI system. There are various manners of stating the performance of AI. However, human cognitive processes and biases are highly influential to the process of explanation in an AI system [20], which implies even the small change in the wordings might affect how people think and behave [22].

The Framing Effect
The framing effect is a cognitive bias where people decide on options based on whether the options are presented in a positive or negative manner [22]. For instance, the attribute of ground beef can either be framed as positive (e.g. 75% lean of ground beef) or negative (e.g. 25% fat of ground beef) ways [15], which is called attribute framing [16]. It is argued that the evaluation of favorability of accepting an object or events is higher when the same attribute is framed positively than it is framed negatively [4, 6, 17] since the information of positive labeling highlights favorable association in the memory whereas that of the negative evoke unfavorable association in the memory [15]. Therefore, people would show a more favorable attitude and behavior toward AI and its suggestion when the performance is stated positively than negatively.

H1: Participants will show more reasonableness and acceptance when positively-framed explanation about the stated performance is presented than negatively-framed explanation is presented

The Message Timing
The timing of presenting messages can also affect acceptance. When a persuasion message is presented after people already made the decision, they might be resistant to change their opinion. People consider what they already possess more valuable than what they do not [12]. It is also true for the decisions they made [2]. When people feel ownership of their own choices, any change from them might be perceived as loss [12, 13], which, in turn, would lead to low acceptance. Therefore, if the suggestion of AI is presented after people already made the decision, they would be reluctant to change their own decision correspond to the suggestion of AI since it would be perceived as a loss.

H2: Participants will show more reasonableness and acceptance when the suggestion of AI is presented before the decision is made than it is presented after the decision is made.

Method

Participants
128 participants took part in the experiment at Sungkyunkwan University. They were mainly university students aged from 18 to 51 years (M = 22.72, SD = 4), and 75 of them (59%) were female. Eight participants were excluded since they indicated potential doubts in deception or failed to follow the experiment instructions. Thus, 120 participants (Age: M = 22.72, SD =4.08; Gender: 70 female [58%]) were included for the data analysis. They were compensated with 3000 won (approximately 3 US dollars) and then debriefed after the experiment.

Materials
The Desert Survival Problem [14] was employed for the experiment. The problem was implemented by using Python [25] and its GUI toolkit library PyQt5 [23]. Participants were given the scenario that they were stranded in the desert and have to rank the set of 15 items [1] depending on the importance of each item in surviving in the desert. They were given two chances to work on their answers: before and after the suggestion of AI. They were informed and believed that the suggestion was created by AI. Yet, the rank-

[1] The 15 items were: cosmetic mirror, plastic raincoat, a quart of water per person, flashlight, pocket knife, red and white parachute, .45 caliber pistol, a pair of sunglasses, compress kit with gauze, compass, sectional air map of the area, a book entitled *Edible Animals of the Desert*, 2 quarts of 180 proof Vodka, 100 salt tablets, and 1 quart of rubbing alcohol (in order of importance on which the suggestion of AI is based).

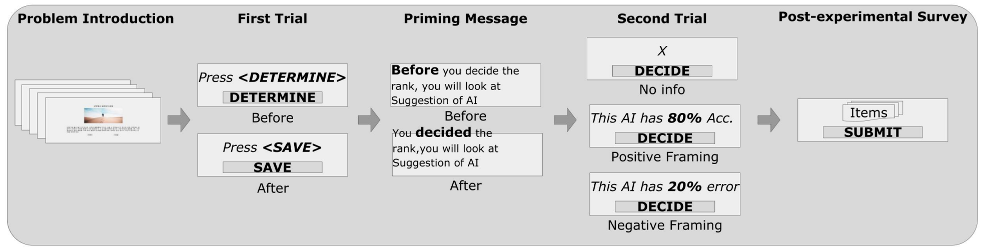

Figure 1: Flow chart of the experimental procedure. Participants solved the same problem twice, one before and one after the suggestion of AI. The message timing was manipulated by instructing the participants to *"Save"* in the Before condition and *"Determine"* in the after condition at first trial. Also, different priming messages were shown between the trials. The message framing was manipulated by providing different information about the performance of AI to participants. *No message* was provided in no information condition, they were told that *"The AI has 80% Accuracy"* in positive framing condition, and *"The AI has 20% error rate"* in negative framing condition.

ing of the items in suggestion was not created by any algorithms but, pre-determined based on the ranking assessed by survival experts [14]. The suggestion was presented identically across all conditions.

Design

The experiment was conducted in a 2x3 between-subjects design. The independent variables were the timing of the presenting message (message timing: before vs. after the participants made a decision) and the framing of stated performance (message framing: no information vs. negative framing vs. positive framing about the accuracy of the AI system). The dependent variables were the acceptance and the perceived reasonableness of the suggestion of AI.

The acceptance was defined as the extent to which the participants change their own decisions before and after they were provided with the suggestion of AI. To score the answers of the first and second trials, the Root Mean Squared Deviation (RMSD) between the participant's answer and the AI's suggestion was calculated for each trial. The smaller RMSD signified that the participant's answer was closer to the AI's suggestion. Then the acceptance was calculated by subtracting RMSD of the second trial from the first one. The larger acceptance score indicated that the participants changed their opinion more in accordance with the AI's suggestion.

The perceived reasonableness (α=0.75) was measured by asking questions including "The suggestion of AI seemed to be reasonable", and "My answer seemed to be more correct." The control variables were the perceived accuracy of AI. It was measured by directly asking the extent to which the participants think that the AI system is accurate. The perceived accuracy was controlled because it hugely variates among people based on the prior familiarity with AI, and the prior level of expertise about AI.

Procedure

When the participants arrived in the laboratory, they were seated in front of the computer and instructed about the experiment. They were introduced with the Desert Survival Problem [14] and solved it using a computer program.

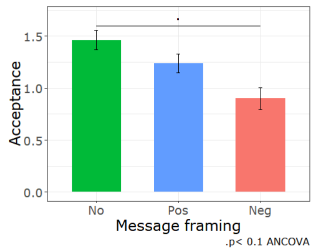

Figure 2: Acceptance between message framing conditions. Planned contrasts result showed a trend of higher acceptance in no information condition compared to Framing conditions when perceived accuracy was controlled.

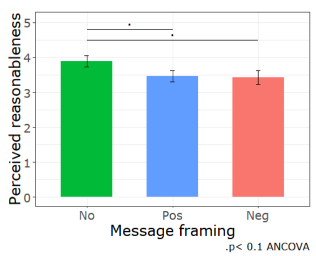

Figure 3: The perceived reasonableness of AI between message framing conditions. Planned contrasts result demonstrated that the perceived reasonableness was significantly higher in no information condition than framing conditions when perceived accuracy was controlled.

In the task, participants were given two chances to work on their answers. The wordings of the instruction in the experiment were different depending on the timing condition that those who were in the before condition were told that they need to press the "Save" button, whereas the participants in the after condition were instructed to press the "Determine" button after making their decisions. After the first trial, participants were presented with a message that either says that “Before you decide the rank of the items, you will now look at what AI system has suggested about this problem” (i.e. before condition) or “You decided the rank of the items. You will now look at what the AI system has suggested this problem” (i.e. after condition) based on the timing condition.

In the second trial, participants were told that they have to finalize the decision with the AI (i.e. before condition) or based on the suggestion of AI (i.e. after condition). In addition, participants were shown with the different message about the accuracy of the AI according to the condition they were in: No information (No message presented), Negative framing (“This AI has 20% error rate”), and Positive Framing (“This AI has 80% accuracy”). After finishing the second trial, the participants were instructed to fill out the survey (Figure 1).

Results

The paired t-test of RMSD between the first and second trials was conducted to see how people considered the suggestion of AI. The result showed a significant difference between the RMSD of first (M=5.87, SD=0.85) and second trial(M=4.67, SD=1.61), $t(119)=10.374$, $p<.001$. Thus, the result demonstrates that in general, the participants did change their initial decision reflecting the AI’s suggestion.

Multivariate Analysis of Covariance (MANCOVA) was conducted to see the main effect of message timing (before vs. after the participants made a decision) and message framing (no information vs. negative framing vs. positive framing about the stated performance of the AI system). The covariates were perceived accuracy (M=56.52, SD=22.4). Multivariate outliers were removed when the maximum Mahalanobis distance value was greater than the critical χ^2 value. In this analysis, none of the data was identified as outliers.

The result of MANCOVA showed a marginally significant main effect of message framing ($F[2,113]=2.54$, $p=.084$) in acceptance after controlling for perceived accuracy (Figure 2). The planed contrasts showed a marginally significant difference between no framing (M=1.46, SD=1.40) and framing conditions (M=1.06, SD=1.18) , $F(1,113)=3.26$, $p=.074$. The main effect of message framing ($F[2,113]=3.35$, $p=.039$) was also found in perceived reasonableness after controlling for perceived accuracy (Figure 3). In message framing, the planed contrasts showed a significant difference between no framing (M=3.89, SD=1.47) and framing conditions (M=3.44, SD=1.35) , $F(1,113)=6.67$, $p=.011$. The result showed that people were less likely to accept the suggestion of AI and perceive that AI is reasonable when the information about AI’s stated performance – either positive or negative- was given compared to when no information was provided.

A significant main effect of the message timing on perceived reasonableness ($F[1,113]=13.71$, $p<.001$) was shown after controlling for perceived accuracy (Figure 4). The post hoc analysis of Tukey’s HSD showed a significant difference between before (M=3.89, SD=1.46) and after (M=3.29, SD=1.29) condition, $p=.003$. However, the main effect of message timing on acceptance was not sig-

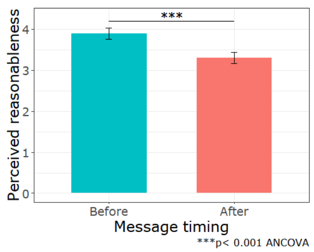

Figure 4: The perceived reasonableness between timing conditions. Tukey's post hoc presented that the perceived reasonableness was significantly higher when the suggestion of AI was provided before the decision is made than after the decision is made when perceived accuracy was controlled.

nificant (F[2,113]=0.90, p=.345). The result implies that people think that the suggestion of AI is more reasonable when the message is presented before they made decisions than after they do so.

Discussion

When designing the AI decision support system, it is important to understand that the different ways of providing information result in different consequences. This study investigated whether providing only the performance of AI is enough for people to accept the suggestion of AI or not. Contrary to the expectation of H1, the results show that participants who were given no information about the performance of AI perceived the suggestion of AI is more reasonable and were more likely to accept the suggestion compared to when information is provided, regardless of the framing.

Although it is hard to conclude that stating only the performance without explanation hampers the acceptance and understanding of AI, a recent study showed that not explaining the algorithms allows more positive perceived trust and understanding of AI than explaining the algorithm regardless of whether the fidelity of the explanation about the algorithm is high or low [21]. These findings demonstrate the possible side effect of proving a performance or an explanation of algorithms. Thus, scholars and designers should be very careful when stating performance or providing an explanation of AI. It also suggests that it would be necessary to test if providing AI's performance information is always better compared to not providing even when AI's performance is perceived to be high.

This study also investigated when would be the better timing for AI to provide a suggestion. Although there was no difference in acceptance between two conditions, the study shows that people perceive the suggestion of AI more reasonable when the message is present before they make a decision than they already made a decision, confirming H2. It implies that when making an AI decision support system, it is important to make the AI that makes a decision with humans, not the AI that just provides feedback or evaluation of the decision that has already been made by humans.

Timing of providing AI's suggestion is an interesting issue since it may have different consequences on different contexts, yet there are limited studies about it. For example, when the uncertainty is rather low such as educational context or when the complexity and stakes are high such as medical situations, an optimized message providing timing may vary. Thus, future studies should explore further on it. In addition, one important thing to note is the significance of the way in which AI communicates, as the small difference of wording in timing manipulation created a big difference in this study.

As the AI decision support system is becoming more and more common, it is important to provide the output of AI in more adequate ways. The current study provides some guidance about how and when the message should be presented for making AI more acceptable. A major concern about this study is that the material used in the experiment was not realistic. Future research should investigate whether the results in the current study can be applied to more naturalistic contexts.

Acknowledgement

This work was supported by the BK21 Plus Program(Interaction Science for Developing Global Human-Informatics leaders focusing on UI/UX, 10Z20130000013) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea(NRF).

REFERENCES

[1] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* 20, 3 (2018), 973–989.

[2] Ziv Carmon, Klaus Wertenbroch, and Marcel Zeelenberg. 2003. Option attachment: When deliberating makes choosing feel like losing. *Journal of Consumer research* 30, 1 (2003), 15–29.

[3] P Daniel Chen and R Michael Haynes. 2016. Transparency for whom? Impacts of accountability movements for institutional researchers and beyond. *New Directions for Institutional Research* 2015, 166 (2016), 11–21.

[4] Mark A Davis and Philip Bobko. 1986. Contextual effects on escalation processes in public sector decision making. *Organizational Behavior and Human Decision Processes* 37, 1 (1986), 121–138.

[5] Dennis Duchon, Kenneth J Dunegan, and Sidney L Barton. 1989. Framing the problem and making decisions: The facts are not enough. *IEEE transactions on engineering management* 36, 1 (1989), 25–27.

[6] Kenneth J Dunegan. 1993. Framing, cognitive modes, and image theory: Toward an understanding of a glass half full. *Journal of Applied Psychology* 78, 3 (1993), 491.

[7] Carlos A Gomez-Uribe and Neil Hunt. 2016. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)* 6, 4 (2016), 13.

[8] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* 2 (2017).

[9] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An introduction to statistical learning*. Vol. 112. Springer.

[10] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology* 2, 4 (2017), 230–243.

[11] Dominik Jung, Verena Dorner, Florian Glaser, and Stefan Morana. 2018. Robo-advisory. *Business & Information Systems Engineering* 60, 1 (2018), 81–86.

[12] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. 1990. Experimental tests of the endowment effect and the Coase theorem. *Journal of political Economy* 98, 6 (1990), 1325–1348.

[13] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives* 5, 1 (1991), 193–206.

[14] J.C. Lafferty, P.M. Eady, A.W. Pond, and Human Synergistics. 1974. *The Desert Survival Problem: A Group Decision Making Experience for Examining and Increasing Individual and Team Effectiveness: Manual.* Experimental Learning Methods. https://books.google.co.kr/books?id=X37-GwAACAAJ

[15] Irwin P Levin and Gary J Gaeth. 1988. How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of consumer research* 15, 3 (1988), 374–378.

[16] Irwin P Levin, Sandra L Schneider, and Gary J Gaeth. 1998. All frames are not created equal: A typology and critical analysis of framing effects. *Organizational behavior and human decision processes* 76, 2 (1998), 149–188.

[17] Irwin P Levin, Sara K Schnittjer, and Shannon L Thee. 1988. Information framing effects in social and personal decisions. *Journal of Experimental Social Psychology* 24, 6 (1988), 520 – 529. DOI: http://dx.doi.org/https://doi.org/10.1016/0022-1031(88)90050-9

[18] Theresa M Marteau. 1989. Framing of information: Its influence upon decisions of doctors and patients. *British Journal of Social Psychology* 28, 1 (1989), 89–94.

[19] Michael Mason. 2008. Transparency for whom? Information disclosure and power in global environmental governance. *Global Environmental Politics* 8, 2 (2008), 8–13.

[20] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38.

[21] Andrea Papenmeier, Gwenn Englebienne, and Christin Seifert. 2019. How model accuracy and explanation fidelity influence user trust. *arXiv preprint arXiv:1907.12652* (2019).

[22] Scott Plous. 1993. *The psychology of judgment and decision making.* Mcgraw-Hill Book Company.

[23] PyQT. 2012. PyQt Reference Guide. (2012). http://www.riverbankcomputing.com/static/Docs/PyQt4/html/index.html

[24] Bruce Upbin. 2013. IBM's Watson Gets Its First Piece Of Business In Healthcare. (Feb 2013). https://www.forbes.com/sites/bruceupbin/2013/02/08/ibms-watson-gets-its-first-piece-of-business-in-healthcare/#227b6c265402

[25] Guido Van Rossum and Fred L Drake Jr. 1995. *Python tutorial.* Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.