Working with a Recommendation Agent: How Recommendation Presentation Influences Users' Perceptions and Behaviors

Emilie Bigras
Marc-Antoine Jutras
Sylvain Sénécal
Pierre-Majorique Léger
Marc Fredette
HEC Montréal
Montréal, QC H3T 2A7, Canada
emilie.bigras@hec.ca
marc-antoine.jutras@hec.ca
sylvain.senecal@hec.ca
pierre-majorique.leger@hec.ca
marc.fredette@hec.ca

Chrystel Black
Nicolas Robitaille
Karine Grande
Christian Hudon
JDA Labs
Montréal, QC H2W 2R2, Canada
Chrystel.Black@jda.com
Nicolas.Robitaille@jda.com
Karine.Grande@jda.com
Christian.Hudon@jda.com

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188639

Abstract

Artificial intelligence (AI) based recommendation agents (RA) can help managers make better decisions by processing a large quantity of decision relevant information. Research on user-RA interactions show that users benefit from RA, but that there are some challenges to their adoption. For instance, RA adoption can only happen if users trust the RA. Thus, this study investigates how the richness of the information provided by an RA and the effort necessary to reach this information influence users' perceptions and usage. A within-subject lab experiment was conducted with 20 participants. Results suggest that perceptions toward the RA (trust, credibility, and satisfaction) are influenced by the RA information richness, but not by the effort needed to reach this information. In addition to contributing to HCI literature, the findings have implications for the design of better AI-based RA systems.

Author Keywords

Recommendation agent; artificial intelligence; trust; perception; behavior; eye tracking.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces; I.2.1. Artificial Intelligence: Applications and Expert Systems.

Introduction

As artificial intelligence (AI) based recommendation agents (RA) become more common in the workplace, investigating how to best present decision recommendations to users becomes important for their adoption and continuous usage. Most research on RA adoption has focused on consumer adoption, not employee adoption [15][19]. Thus, a better understanding of how professionals use and perceive RAs and more importantly which RA characteristics promote their adoption will enrich theory development. In addition, it will inform the human interaction community in the design of better RA interactions.

This study investigates how the way recommendations are presented to professionals influences their usage behavior and perception of the RA. It aims at answering the following questions: Which recommendation representation will planners trust and use the most? Are they likely to prefer a simple recommendation represented as a product score (low information richness) or a product score which provides information on the different variables included in its calculation (high information richness). Also, will users prefer to get less information faster on the product score (low effort) or more information which may be longer to access (high effort). To answer these questions, the recommendation representations were manipulated with two factors in the study: effort required to access information (low or high) and information richness (low or high). The effort is associated to the number of steps required to get to the information (e.g., number of screens) and the information richness is associated with the amount of information provided by the RA to assist assortment planners through their decision-making process. The latter has been demonstrated to influence the RA perceived credibility [5]. We suggest that these two factors will impact users' RA perceptions.

The context of the study is retail assortment decisions. In order to create an optimal assortment of products, assortment planners need to take into consideration qualitative and quantitative criteria while making a decision [3]. These criteria include an important number of variables (e.g., past sales, retail trends, inventory, sales forecast, customers' needs) which creates a certain level of uncertainty. AI based recommender systems can now be used as an aid by assortment planners through their decision-making process [17]. Research shows that RAs reduce the volume of information to process [18], enhance decision-making process [1], and improve user satisfaction [4]. But, some challenges affect their adoption. For instance, one challenge resides in building RA credibility in order for users to trust recommendations and adopt the RA [6][9][20].

Method

A within-subject laboratory experiment was conducted using the experimental RA prototype for assortment planning developed by JDA Labs (Montreal, Canada). Twenty logistics and marketing professionals (M_{age} = 26, SD = 3.92; 9 women) participated in the study. Participants had to make assortment decisions for two fictitious scenarios and for each scenario they were exposed to 3 conditions in a counterbalanced order: Task 1 represented a low effort & low richness condition

Figure 1: Experimental set-up.

Figure 2: Each product was presented with an image, its name including its brand, and its product score (i.e., RA's recommendation).

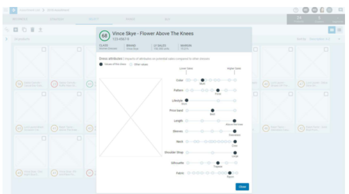

Figure 3: For T2, each product had a modal window presenting additional information on the variables included in its product score calculation.

Figure 4: For T3, each product had a new page to access the additional information which required more effort.

(T1), Task 2 reflected a low effort & high richness condition (T2), and the third task was a high effort & high richness condition (T3). Due to time constraints, the high effort & low richness condition was not included in the experiment. Each scenario started with a practice task with no RA to familiarize participants with the assortment planning software. They each received a $30 gift card as a compensation. This experiment was approved by the IRB of our institution.

For each task of each scenario, 24 different products were presented to participants. Through their decision-making process, participants needed to make an assortment decision by selecting, from the 24 products displayed, the optimal assortment of products. Each scenario specified the total number of products (ranging from 6 to 7) that needed to be selected by the participants for their optimal assortment for each condition. Figure 2 represents the elements that were displayed for each product and in each condition to the participants. The product score, i.e., the RA's recommendation, has a value between 0 to 100 and provides the AI-based forecast based on the RA. For T1, the product score represented in Figure 2 was the only source of information made available to the participants. For T2 and T3, the product score was also made available to the participants, but, by clicking on the products, participants had also access to more information. This additional information included various product characteristics (e.g., attributes, past sales, margin, comparative products). However, the effort required to reach this additional information varied between T2 and T3. For T2, the information was made available through a modal window (Figure 3). As for T3, the information was accessed with additional navigation through a new page (Figure 4).

Apparatus and Measures

The experimental prototype for assortment planning by JDA Labs (Montreal, Canada) was presented through Axure RP 8. Task stimuli were presented on a monitor with a 1680 x 1050 resolution. All statistical analyses were performed with SAS 9.4.

Psychometric Measures

After each task, participants completed a questionnaire. It measured users' perceptions toward the RA in terms of trust [8], credibility [11], satisfaction [16] and type of future usage (i.e., RA used as a decision aid or as a delegated agent [8]).

Behavioral measures

In addition to self-reported measures, exploratory observational measures were also used. A Smart Eye Pro system (Gothenburg, Sweden) was used to track users' visual attention in each task (at a 60Hz sampling rate). A 9-point (3 x 3) calibration grid was used. Calibration was repeated for each participant until sufficient accuracy was reached (± 2 degrees of accuracy). The eye tracking data was analysed with the MAPPS 2016.1 software. Areas of interest (AOIs) were created for each RA product score presented in Figure 2 (1 AOI x 24 products x 6 tasks) and for each modal window or a new page with additional information that was consulted by participants. For each AOI, the number and duration of ocular fixations were measured. Based on Rayner [13], the fixation threshold set at 200 milliseconds.

Results

Users' Perceptions

A linear regression with random intercept was used to test the difference between the means of users' perceptions (Table 1). First, no difference between low effort & high richness (T2) and high effort & high richness (T3) was found, indicating that the effort required to access the information does not have an impact on user perceptions. Second, results suggest that information richness plays a key role in users' positive perceptions since RA trust, credibility, satisfaction and intention to adopt are all greater in high information conditions (T2 and T3) than in the low richness condition (T1). In addition, a Wilcoxon signed rank test with a two-tailed level of significance was also performed to compare the difference between the intention to adopt the RA as a delegated agent and the intention to adopt the RA as a decision aid for each condition. Results suggest that participants are more willing to adopt the RA as a decision aid than as a delegated agent for all three conditions (T1: $p \leq .0001$; T2: $p \leq .0001$; T3 $p \leq .0001$).

	Result	Estimate	p-value[1]
Trust	T2 > T1	0.8748	***
	T3 > T1	0.7070	***
The Intention to Adopt the RA as a Delegated Agent	T2 > T1	0.8503	***
	T3 > T1	0.5647	*
The Intention to Adopt the RA as a Decision Aid	T2 > T1	1.0705	***
	T3 > T1	0.9681	***
Credibility	T2 > T1	0.7055	***
	T3 > T1	0.6139	***
Satisfaction	T2 > T1	0.6068	**
	T3 > T1	0.5103	*

[1] Two-tailed level of significance: * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Table 1: User Perception Results

Users' Behaviors

First, a linear regression with a mixed model adjusted for multiple comparisons with a two-tailed level of significance was used to test the difference between the least square means of the duration for the AOIs of the different conditions. The difference between the least square means of the number of fixations for the AOIs of the different conditions was also tested with a Poisson regression with a mixed model adjusted for multiple comparisons. When additional information is available (T2 and T3), participants spend less time on product scores, but consult them more frequently (T2 and T3 greater than T1, respectively 1.1642, $p = .0003$ and -1.0554, $p \leq .0001$; 1.3147, $p \leq .0001$ and -1.3978, $p \leq .0001$).

Second, a Wilcoxon signed rank test with a two-tailed level of significance was performed to compare the difference between the first 25% and the last 25% of the task duration and the number of fixations for each AOI of each condition. Results show no difference in the time and the frequency that the product score was consulted by the participants through time. However, it revealed that when in the low effort & high richness condition (T2), participants consulted the additional information more frequently and for a longer period of time at the beginning of the task ($p = .0274$ and $p = .0342$, respectively). It also showed that in the high effort & high richness condition (T3), participants consulted the additional information for a longer period of time, but not more frequently at the beginning of the task ($p = .0559$ and $p = .1104$, respectively). A similar test was also performed to compare the difference between the first 40% and the last 40% of the task duration and the number of fixations for each AOI of each condition. Results were in line with the above results.

Discussion and Concluding Comments

Our results show that in a professional context, the way recommendations from RA are presented influences users' perceptions and behaviors. First, a RA providing rich information is perceived as more trustworthy, credible, and satisfactory. In addition, users are more willing to adopt the RA as a decision aid than as a delegated agent and this perception is increased when the RA provides rich information. That being said, users seem to prefer a product score that is enhanced with additional information on the variables included in its calculation. The effort to reach this information does not seem to impact users' satisfaction. Second, results suggest that when assortment planners have access to richer information, they consult for a longer period of time, but less frequently, this information. Furthermore, users are referring more to this additional information at the beginning of their decision-making process, compared to the product score that is consulted consistently through their decision-making process.

As AI based RA provide recommendations based on a tremendous amount of data, the way the information needed by the users must be presented is a formidable challenge. Although a condensed visual representation needs to be created by designers of such tools in order to manage this amount of data [10], our findings suggest that users still need to open the black box and access additional information for them to trust and be satisfied with the RA. As AI will eventually become common in the workplace [2], it is crucial to understand how to best present AI based recommendations in order to have employees trust these recommendations. Best practices in UX design for AI can then be generated based on such insights.

Obviously, more research is needed to better understand how professional users interact with RAs and provide more guidelines to UX designers. The effort/accuracy decision-making framework [12] could be useful to investigate the tradeoff between decision effort and accuracy users are willing to make in their professional context. As user behavior changes over time [7][14], longitudinal studies of RA usage will also contribute to understanding how the user-RA relationship evolves in terms of perceptions and behaviors.

Acknowledgements

We are thankful for the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. ALJUKHADAR, M., SENECAL, S., and DAOUST, C.-E., 2012. Using recommendation agents to cope with information overload. *International Journal of Electronic Commerce 17*, 2, 41-70.
2. ANDREWS, W., SAU, M., DEKATE, C., MULLEN, A., BRANT, K. F., REVANG, M., and PLUMMER, D. C., 2017. Predicts 2018: Artificial Intelligence. *Gartner* (13 November). DOI= http://dx.doi.org/G00343423.
3. BRIJS, T., SWINNEN, G., VANHOOF, K., and WETS, G., 1999. Using association rules for product assortment decisions: A case study. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* ACM, 254-260.

4. CEARLEY, D., BURKE, B., SEARLE, S., WALKER, M., 2017 Top 10 Strategic Technology Trends for 2018. *Gartner*(October 3). DOI= http://dx.doi.org/G00327329.
5. HEESACKER, M., PETTY, R.E., and CACIOPPO, J.T., 1983. Field dependence and attitude change: Source credibility can alter persuasion by affecting message-relevant thinking. *Journal of Personality 51*, 4, 653-666.
6. HENGSTLER, M., ENKEL, E., and DUELLI, S., 2016. Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change 105*, 105-120. DOI= http://dx.doi.org/10.1016/j.techfore.2015.12.014.
7. KNIJNENBURG, B.P., WILLEMSEN, M.C., GANTNER, Z., SONCU, H., and NEWELL, C., 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction 22*, 4-5, 441-504.
8. KOMIAK, S.Y. and BENBASAT, I., 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS quarterly*, 941-960.
9. LEMOINE, J.-F. and CHERIF, E., 2012. Comment générer de la confiance envers un agent virtuel à l'aide de ses caractéristiques? Une étude exploratoire. *Management & Avenir*, 8, 169-188.
10. NILSSON, N.J., 2014. *Principles of artificial intelligence*. Morgan Kaufmann.
11. OHANIAN, R., 1990. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of advertising 19*, 3, 39-52.
12. PAYNE, J.W., BETTMAN, J.R., and JOHNSON, E.J., 1993. *The adaptive decision maker*. Cambridge University Press.
13. RAYNER, K., 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin 124*, 3, 372.
14. SÉNÉCAL, S., FREDETTE, M., LÉGER, P.-M., COURTEMANCHE, F., and RIEDL, R., 2015. Consumers' cognitive lock-in on websites: evidence from a neurophysiological study. *Journal of Internet Commerce 14*, 3, 277-293.
15. SENECAL, S. and NANTEL, J., 2004. The influence of online product recommendations on consumers' online choices. *Journal of Retailing 80*, 2, 159-169. DOI= http://dx.doi.org/10.1016/j.jretai.2004.04.001.
16. SIRDESHMUKH, D., SINGH, J., and SABOL, B., 2002. Consumer trust, value, and loyalty in relational exchanges. *Journal of marketing 66*, 1, 15-37.
17. WANG, L., ZENG, X., KOEHL, L., and CHEN, Y., 2015. Intelligent fashion recommender system: Fuzzy logic in personalized garment design. *IEEE Transactions on Human-Machine Systems 45*, 1, 95-109.
18. WANG, W., and BENBASAT, I., 2005. Trust in and adoption of online recommendation agents. *Journal of the association for information systems 6*, 3, 71-101.
19. WANG, W. and BENBASAT, I., 2009. Interactive decision aids for consumer decision making in e-commerce: The influence of perceived strategy restrictiveness. *MIS quarterly*, 293-320.
20. XIAO, S. and BENBASAT, I., 2003. The formation of trust and distrust in recommendation agents in repeated interactions: a process-tracing analysis. In *Proceedings of the 5th international conference on Electronic commerce* ACM, 287-293.