Effect of Presentation Methods on User Experiences and Perception in VR Shopping Recommender Systems

Yang Zhan
yang.zhan@dcl.cs.waseda.ac.jp
Waseda University
Tokyo, Japan

Tatsuo Nakajima
tatsuo@dcl.cs.waseda.ac.jp
Waseda University
Tokyo, Japan

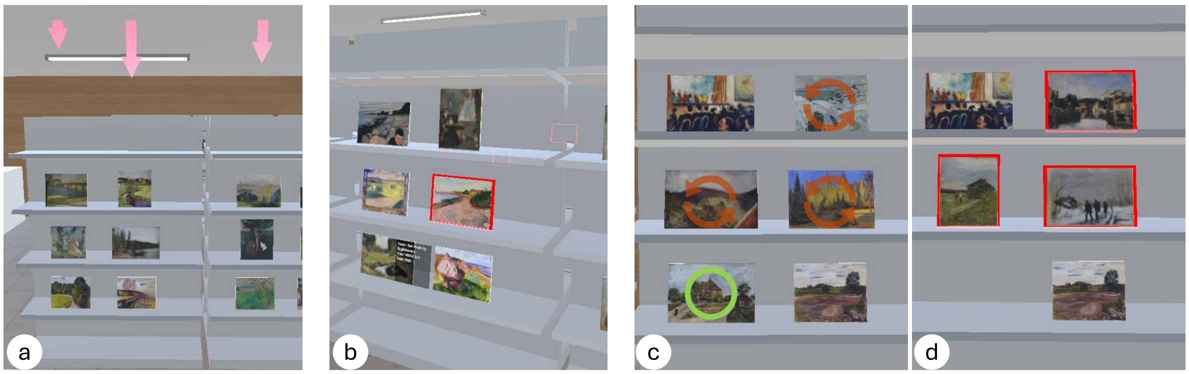

Figure 1: Presentation methods for the Recommender System (RS). (a) Arrows indicating the recommendation results through their position and length. (b) Highlighting the recommended paintings with a red frame. (c) The paintings prior to swapping, where the green circle marks the painting selected by users, and orange swap marks indicate the paintings to be replaced. (d) The recommended paintings swapped to the shelf.

ABSTRACT

Shopping in Virtual Reality (VR) contains numerous advantages, such as detailed product diagnosticity and virtually unlimited store space. It also provides superior hedonic features compared to traditional online shopping. However, Recommender Systems (RSs), which are commonly used to assist users in finding preferred products in online shopping, have not yet been extensively researched in VR shopping. It is crucial to understand how users experience RSs and perceive the recommendation results within VR stores. To address the research gaps, we compared three presentation methods (Arrow, Highlight, Swap) with varying levels of perceptibility for an RS in VR shopping. A within-subject study (N=14) revealed that the methods with higher perceptibility enhanced user experiences, reduced perceived workload, and garnered more preferences. Additionally, we examined the effects of these presentation designs on sense of agency and trust in RS, considering their interaction with users' prior trust. Our study contributes to the design of RS interfaces and the future implementation of Trustworthy Recommender Systems (TRS) in VR shopping.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Human-centered computing** → **Virtual reality**; **User studies**.

KEYWORDS

Virtual Reality, Shopping Experience, Recommender System, Visual Cues, Trustworthy Recommendation

ACM Reference Format:
Yang Zhan and Tatsuo Nakajima. 2024. Effect of Presentation Methods on User Experiences and Perception in VR Shopping Recommender Systems. In *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region (SIGIR-AP '24), December 9–12, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3673791.3698430

1 INTRODUCTION

Virtual reality (VR) enables the creation of immersive virtual environments (VEs) that provide impressive experiences deriving from the real world yet surpassing it. Among the numerous VR applications, VR shopping, which replicates the real-world shopping experience, has gained popularity due to its capacity to show comprehensive goods information and satisfy entertainment needs [15, 33]. The advent of various VR e-commercial applications also exemplifies this trend. However, similar to the experience in offline stores, users often face difficulties in finding desired goods in VR stores, and therefore diminish their overall experiences. We believe that two primary factors contribute to these difficulties in VR shopping. Firstly, VR stores simulate physical three-dimensional spaces, which

SIGIR-AP '24, December 9–12, 2024, Tokyo, Japan.

ACM ISBN 979-8-4007-0724-7/24/12
https://doi.org/10.1145/3673791.3698430

can lead to cognitive overload [41]. Secondly, VR shopping includes not only physical goods commonly encountered in daily life, but also virtual goods merely utilized within VEs, such as in games and applications. Users may lack prior knowledge of these virtual goods when engaging with the new VEs. Previous research [42, 52, 53] has introduced search engines to facilitate goods-seeking; however, these solutions are predominantly effective for target-oriented shopping and may not be suitable for exploratory shopping. Moreover, they presume that users know certain information about the target goods (e.g., name, category), which diminishes their utilitarian values for shopping virtual goods.

Recommender Systems (RS) have proven to be an effective tool for users with exploratory motives to seek desired goods [54], and have been conducted to improve shopping experiences [23]. Recently, Yilma et al. [56] applied an RS based on artificial intelligence (AI) to facilitate visual art discovery, highlighting the potential for RS to be used for specialized products. When considering the user-RS interaction, we believe that the presentation of recommendation results is a key component, as it affects user satisfaction [39, 59] and behavior [57]. Previous research in online shopping has also indicated that the design of the presentation interface affects perceived trustworthiness [5], which is essential for establishing a trustworthy recommender system (TRS) [51]. However, there are still research gaps in understanding how to present the results of recommendation and how the presentation affects users' perception of RS in VR shopping environments, considering the distinct differences between virtual and real-world contexts (e.g., desktop, smartphone) [31]. The simple adaption of website-based interfaces in VEs may decrease immersion and presence, negatively influencing user experiences [47].

In this study, to fill the research gaps, we designed and compared three types of presentation methods that provide users with recommendation results varying in perceptibility. We manipulated the perceptibility of visual cues in three methods (Arrow, Highlight, Swap) illustrated in Figure 1, because presenting information to users in a spatial environment pertains to information awareness and attention guidance processes [25, 28, 34]. Previous research has revealed that visual cues with different perceptibility impact task performance, immersion, and perceived workload [20, 25, 44]. On one hand, higher perceptibility often results in stronger stimuli, enabling users to notice information more swiftly [20]. On the other hand, some highly perceptible cues could detract from immersion experiences in VR [44]. However, these studies primarily focus on guiding attention to specific targets in a single task, rather than assisting users with exploratory goals in a VR shopping context. It is important to investigate how perceptibility affects user experiences when designing the presentation methods in RS. Apart from the previous studies, the sense of agency (SoA) which refers to the degree of users' controllability [12], was considered as a dependent factor influenced by perceptibility in our research. Highly perceptible cues can orient users' attention more easily, potentially leading to a lower SoA which diminishes user experiences [35]. Moreover, perceived trust is recognized as an equally important indicator alongside RS performance [51]. Prior studies demonstrated that the trust in the system was affected by interface designs [22] and controllability [11]. It is reasonable to hypothesize that there is a relationship between the perceptibility of designs and trust in RS within VR environments.

In conclusion, the aims of our study are summarized as two research questions:

- **RQ1**: How do the RS and its presentation method affect user experiences?
- **RQ2**: How does the presentation method influence user perception and trust in the RS?

2 RELATED WORK

2.1 VR shopping experiences

Leveraging the advantages of VE, VR shopping facilitates diverse shopping functions and store designs to enhance the purchase process. Improving utilitarianism and interactivity has been a crucial focus in VR shopping studies. Qin et al. [33] have combined different space relationships, visual presentations, and user interaction to design new presentations of information that help product diagnosticity. Speicher et al. [43] compared isomorphic and non-isomorphic shopping baskets and selections, finding that non-isomorphic designs improved user experiences and reduced error rates. Similarly, Ward et al. [52] examined VR stores with different context alignments and Search Engine Results Pages (SERP), highlighting the importance of contextual information and attention allocation in the information-seeking process. Additionally, numerous studies focused on users' behavior and perception during shopping. Pfeiffer et al. [32] utilized machine learning and eye-tracking to predict users' motivations in shopping and achieved high accuracy. Workload in shopping was assessed across VR, Augmented Reality (AR), and real-world stores, revealing that conducting AR in VR stores decreased the physical demand [55]. Further research has explored the hedonistic values of VR shopping, including immersion [1, 31], social interaction [47], and motivations [15]. Despite the widespread application of RS in online shopping, previous studies have largely overlooked their use in VR shopping, particularly with respect to both utilitarianism and user perception.

2.2 RS Interfaces

The interfaces of RS have been extensively studied to improve the performance of RS [18, 58] and decrease the decision-making effort [45]. Although much research has concentrated on improving the accuracy of RS to better match user preferences, higher accuracy does not necessarily correlate with increased user satisfaction [24]. As a form of persuasion system [10], the presentation results in RS directly influence users' trust and experiences [6], and this trust subsequently affects the adoption rate of the recommendation results [19]. For instance, in the movie RS, Nanou et al. [27] compared various organizational structures and modalities for presenting results, unveiling that structured organizations and results with text and video enhance users' satisfaction and persuasive effectiveness. A study involving 131 college-aged online shoppers explored the optimal components and features of RS presentation interfaces, indicating that users prefer specific information about goods and a limited number of recommendation results [30]. Highlighting in online shopping has been identified as an effective method for orienting attention to the most relevant result [50] or explaining the varying importance of different results [8]. In the context of mobile applications, Ji et al. [17] proposed an implicit presentation method to seamlessly integrate recommendations, thereby reducing users' awareness of the RS and simplifying the interaction complexity.

2.3 Attention orientation in VR

The 360-degree VE provides strong immersion and presence for users but presents challenges in making detailed information in scenes accessible due to limited human attention [4, 21]. Consequently, it is pivotal to direct attention to Regions of Interest (ROI). Based on the theory of attention orientation [49], cues can be classified into two groups: exogenous and endogenous cues. Exogenous cues arouse bottom-up involuntary attention [40], causing individuals to react unconsciously [26]. An example is that people naturally pay attention to a bright light in a dark environment. Conversely, endogenous cues refer to visual signals that guide top-down, voluntary attention [40], such as when individuals direct their focus to a specific location indicated by an arrowhead signal [49]. Furthermore, Soret et al. [40] demonstrated that exogenous visual cues direct attention locally while endogenous cues guide attention holistically. Exogenous cues have been employed to enhance awareness of non-immersed bystanders [25] and to guide attention in cinematic VR [28, 37]. But also there was research [37] suggesting that the taxonomies of visual cues may change when applied to new research fields.

3 COMPARATIVE STUDY

In this study, a within-subject design was employed to investigate the impact of presentation methods on shopping experiences (**RQ1**), perception and trust (**RQ2**). For **RQ1**, given the limited research on the application of RS in VR shopping, it is essential to verify if the RS in VR shopping facilitates the selection of goods. Consequently, we compared user experiences and behaviors with and without RS. We evaluated the utilitarian and hedonic aspects of different presentation methods for recommendation results and assessed user preferences to inform future designs. For **RQ2**, as discussed in Section 1, SoA, which is linked to user experiences [35], was examined to explore how it is affected by perceptibility and its role in RS for VR shopping. Perceived workload was also measured given the time-consuming shopping in VR stores [55]. Since trust in RS is crucial for the adoption of recommended results, we concentrated on the relationships among trust, method perceptibility and prior trust in RS. Totally 14 participants (2 female and 12 male) from the university campus, aged between 21 and 26 ($M = 23.9, SD = 1.61$), participated in the study. Each participant experienced three different presentation methods as well as a baseline condition without the RS. A Latin square design was utilized to randomize the order of the four experimental conditions.

3.1 Experimental Design

We implemented a VR store featuring Edvard Munch's oil paintings to establish a shopping task for the RS. Artwork products, being less familiar than daily necessities, have been utilized to measure shopping workload [55]. The considerable effort required to interpret their aesthetics and make selections makes them suitable for examining how RS and presentation designs assist shopping. We adopted a stock-on-shelves design, widely used in games [46] and VR shopping [31, 42], to enhance the generalizability of our findings to virtual stores with similar designs, regardless of the actual goods being sold. To avoid repetition, a total of 360 oil paintings were randomly divided into four groups, with only one group (90 paintings) being used as goods for selling in one condition. The paintings were distributed across 15 shelves, each holding 6 paintings. Given that vital information about oil paintings can be observed without picking them up, which might simplify user behavior and diminish generality for other goods types, we structured the participants' task into two steps: (1) come across several oil paintings that appealed to them with explorative motivation and collect these oil paintings to a table; (2) choose a favorite painting on the table and throw it into a box to checkout. In step (1), the RS was activated when participants picked up a painting, generating three recommendation results presented using one of the following methods illustrated in Figure 1:

- **Arrow**: Arrows are positioned above the shelves of results with varying lengths (short, middle, long). Their positions indicate the columns of the recommended paintings, while their lengths represent the distances to the paintings. For example, a short arrow signifies that the recommended painting is on the top of the shelf. This method serves as a common endogenous cue, which evokes voluntary attention in previous studies [29, 37, 40], and provides **low** perceptibility for recommendations.
- **Highlight**: This method highlights the frames of the recommended paintings, allowing participants to see the highlights through obstacles, thus aiding in discovery and preventing neglect. However, due to the limited field of view, participants may not notice all highlight targets simultaneously. This method is widely used as an exogenous cue [9, 36] in extended reality and offers **moderate** perceptibility.
- **Swap**: The recommended paintings are swapped with the paintings on the shelf of the painting picked up. In mobile applications, presenting recommendation results by reordering products has been implemented and validated [17]. This method replicates this process in VE. To inform participants of the results, the recommended paintings marked by red frames were preferentially set on the upper shelves but close to the selected painting. This method provides exogenous cues and **high** perceptibility.

We answered the research questions by comparing these three presentation methods and a baseline condition without RS (**NoRS**) in the VR store.

3.2 Procedure and Measures

Upon arrival at the laboratory, participants were informed of the research objectives, provided their consent, and answered a demographics questionnaire. Subsequently, we introduced the experiment task, each presentation method, and the operation of VR controllers. Participants were encouraged to raise any questions to clarify any potential misconceptions. Once they felt comfortable, participants engaged in a two-minute tutorial within the VR store devoid of actual paintings to familiarize themselves with the operation and VE. Participants then proceeded to complete the task using one of the presentation methods or without the RS. When notified

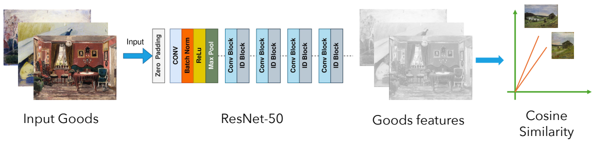

Figure 2: The pipeline of conducted RS. Oil paintings are first fed into the ResNet-50 to extract their features, and then a cosine similarity matrix is calculated for further recommendation. When the RS is triggered, it recommends the most similar paintings to the selected painting.

that the task was completed, they were instructed to remove the VR headset and fill out questionnaires measuring their shopping experience for that particular task. After all four conditions were completed, a semi-structured interview was conducted to gather additional insights.

To investigate user behavior when utilizing the RS and various presentation methods, we recorded objective data in the VR store. For each task, we measured shopping time, the number of pick-up behaviors and new paintings picked up, and the adoption rate of recommendation results. Post-experiment questionnaires included a short User Experience Questionnaire (UEQ-S) [38] to assess overall shopping experiences and a raw NASA Task Load Index (TLX) [14] to evaluate perceived workload. Additionally, we incorporated four questions adapted from [48] to measure SoA, one question for immersion, and two questions to assess trust in the RS and intention to use the presentation method again. The UEQ-S, a condensed version of the UEQ, maintains similar outcomes by measuring pragmatic quality (PQ) and hedonic quality (HQ). The raw NASA-TLX omits the comparison between sub-scales to reduce complexity while retaining the validity of the original questionnaire [13]. The interview explored participants' preferences for the presentation methods and their reasons.

3.3 Apparatus and Implementation

The VR store was developed using Unity 2021.3 and deployed to Meta Quest 2. The selected oil paintings came from the Edvard Munch painting dataset[1], depicting landscapes, still life, and everyday life scenes. The dataset comprises 1769 paintings created by Edvard Munch, accompanied by metadata such as name and year, which was utilized as product information. Regarding the RS in Figure 2, we implemented a ResNet-50 [16], pre-trained on ImageNet [7], as the backbone to extract feature matrices from the paintings. The RS calculates cosine similarities between the features of each painting in the same group. When triggered by user pick-up actions, the RS returns three paintings with the highest similarity scores to the selected artwork. Prior research [56] revealed that purely image-based RS can achieve satisfactory usability.

4 RESULTS

For the ordinal data, a Friedman non-parametric test was conducted, followed by a Wilcoxon Signed-Ranks posthoc test when overall the statistically significant differences among four conditions were identified. For the continuous data, a one-way ANOVA was utilized when the assumption of normal distribution was met. Following [22], we employed a generalized linear-mixed model (GLMM) to analyze the main effects and interaction effects of prior trust and recommendation methods on users' trust in RS. All participants declared that they were not familiar with the paintings of Edvard Munch.

4.1 Objective Measures

Among the four conditions, a borderline difference was observed in the amount of new paintings picked up ($\chi^2(3) = 7.49$, $p = 0.058$). Pairwise comparisons indicated that the NoRS ($p = 0.043$) conditions resulted in more new paintings being explored compared to the Arrow condition. No significant differences were detected in other objective data.

4.2 Overall Experiences

Figure 3 presents the results from the post-experiment questionnaires. Across those mean data, the analysis revealed that there were significant differences in user experiences, as measured by pragmatic quality (PQ) ($\chi^2(3) = 25.43$, $p < 0.001$) and hedonic quality (HQ) ($\chi^2(2) = 16.04$, $p < 0.001$). According to the posthoc tests, participants experienced better PQ in Swap ($p_{Arrow} < 0.001$, $p_{NoRS} < 0.001$) and Highlight ($p_{Arrow} = 0.002$, $p_{NoRS} = 0.003$) conditions than the other conditions. Moreover, a significant difference ($p = 0.043$) was confirmed between Arrow and NoRS, showing the conduction of RS improved the utilitarian perspective of shopping. Next, the posthoc tests indicated significantly higher hedonic values in Swap when compared to Arrow ($p = 0.003$) and Highlight ($p < 0.001$). For the intention to use the method again (IT), Swap ($p = 0.002$) and Highlight ($p = 0.015$) were rated more favorably than Arrow. No significant difference in immersion (IM) was found among the conditions.

4.3 Perception

In Figure 3, the significant difference in SoA was found by a Friedman test ($\chi^2(3) = 15.7$, $p = 0.001$). Specifically, participants reported less agency and controllability when using Arrow ($p = 0.012$), Highlight ($p < 0.001$), and Swap ($p = 0.002$) compared to the NoRS condition. Figure 4 shows the results from the subscales of the raw NASA-TLX questionnaire. A Friedman test identified a significant overall effect ($\chi^2(3) = 11.94$, $p = 0.008$). Participants perceived a lower workload in Swap in comparison to NoRS ($p = 0.014$) and Arrow ($p = 0.030$). There was also a significant difference ($\chi^2(3) = 9.68$, $p = 0.022$) in the mental demand subscale, with mental demand being lower when Swap was used compared to NoRS ($p = 0.020$) and Arrow ($p = 0.014$). Additionally, a significant difference ($\chi^2(3) = 11.57$, $p = 0.009$) was uncovered in the effort subscale, with participants reporting less effort required in Swap compared to NoRS ($p = 0.007$), Arrow ($p = 0.023$), and Highlight ($p = 0.025$).

4.4 Prior Trust and Trust

Results from the GLMM indicated that there was a significant main effect of prior trust ($\beta = -0.026, S.E = 0.013, p = 0.041$) and recommendation methods (Swap vs. Arrow, $\beta = -0.146, S.E = 0.071, p =$

[1]https://www.kaggle.com/datasets/isaienkov/edvard-munch-paintings

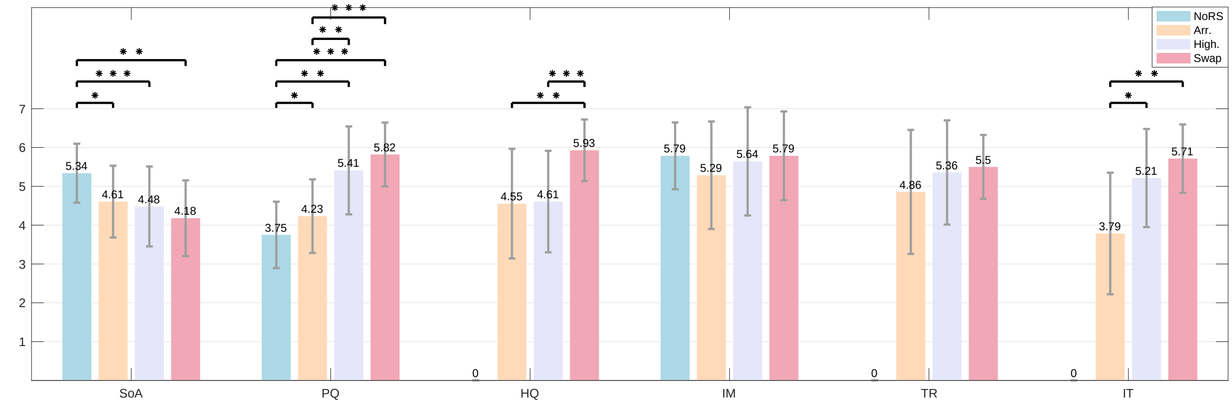

Figure 3: The means and standard deviations of the post-experimental questionnaires, including sense of agency (SoA), pragmatic quality (PQ), hedonic quality (HQ), immersion (IM), trust in RS (TR), and intention to use the method again (IT). As the baseline condition, HQ, TR, and IT are unavailable for NoRS. Statistical significance levels from Wilcoxon Signed-Ranks posthoc tests are denoted by $*$ ($p < 0.05$), $**$ ($p < 0.01$), $***$ ($p < 0.001$).

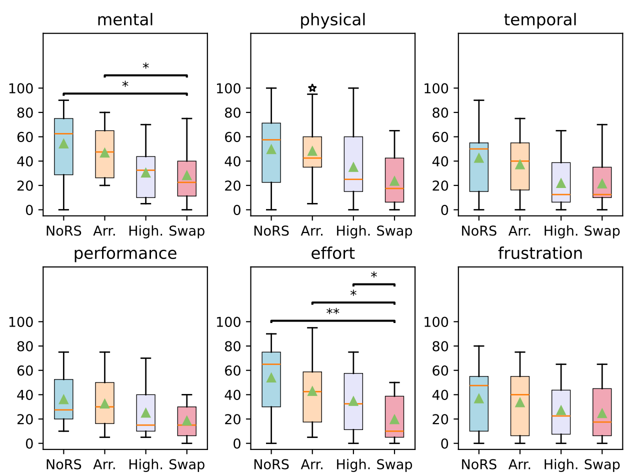

Figure 4: NASA-TLX questionnaire subscale results. NoRS, Arr., and High. are abbreviations for the conditions without RS, Arrow, and Highlight, respectively. The orange line indicates the median value, while the green triangle denotes the mean value. Outliers are represented by star marks. Statistical significance levels are denoted by $*$ ($p < 0.05$) and $**$ ($p < 0.01$)

0.039) on trust in RS. The result found a borderline interaction effect($\beta = 0.023, S.E = 0.013, p = 0.074$) between prior trust and recommendation methods (Swap vs. Arrow) on perceived trust in the RS. No significant main or interaction effects were identified for other conditions.

4.5 Interview

In the semi-structured interview, we utilized the thematic analysis [3] to analyze the recordings, resulting in the identification of two primary themes.

Demands for consistent shopping experiences. A total of 64.3% of participants preferred the Swap method due to its efficiency, which eliminated the need for additional teleportation to view the recommendations. Some participants mentioned that highlighted frames in the Highlight method drew too much attention and disrupted their browsing order, leading to disorientation. Conversely, all the rest of the participants favored the Highlight method and mainly appreciated that it allowed them to facilitate continuous exploration of new paintings by moving in the store. In these participants, however, the Swap method was criticized for messing up the order of paintings, causing confusion and anxiety. The different preferences indicate that RS and presentation methods should avoid interrupting the intended shopping flow of users.

Diminishing workload in exploratory shopping. Certain workload was perceived in the shopping task and many participants expressed a desire for presentation methods to decrease the workload. The high perceptibility of the Swap method was favored for its ability to provide convenient comparisons. Also, the clear cues in both exogenous methods help locate recommended paintings directly. In contrast, most participants believed the Arrow method was unnatural and tiring since they needed to focus on the marks far distant from the goods on shelves, necessitating head movement and redirection.

5 DISCUSSION

Our study revealed that RS and its presentation methods with higher perceptibility can improve user experiences and reduce workload. We confirmed that the RS with presentation methods decreased the SoA and influenced participants' behavior. The divergent preferences between Swap and Highlight provided insights for future designs of user-RS interaction. Additionally, the observed effects of prior trust and presentation methods on trust in the implemented RS underscore their significance in developing TRS for VR shopping.

RQ1: User experiences and behavior in VR shopping with RS. By comparing the PQ with and without the RS, our study demonstrates that integrating the RS into VR shopping improved the user experiences from utilitarian perspectives and decreased the complexity of exploring the goods while maintaining an immersive experience.

Among the presentation methods, as shown in Section 4.2, the Swap and Highlight method achieved better performance on PQ, showing their utilitarian value. This superior performance was attributed to their higher perceptibility, allowing participants to identify recommended results at a glance. Furthermore, the swapping behavior may have contributed to a sense of novelty, thus improving its HQ. Conversely, when the RS was conducted with the Arrow presentation, participants tended to explore fewer paintings compared to the NoRS, suggesting potentially lower shopping efficiency. Interviews revealed that participants required clear and direct cues for presenting recommended results, indicating that endogenous and low perceptibility cues may not be suitable for RS in VR stores. Future designs could benefit from integrating the advantages of both Swap and Highlight methods to provide convenient comparison, high perceptibility, and automation, thereby enhancing user experiences.

The interviews also highlighted that the consistency of shopping experiences influenced preferences for presentation methods and overall user experiences. Participants who favored Swap valued its efficiency, while those who preferred Highlight appreciated its support for exploration within the store space. This suggests that presentation methods should align with the expectations and motivation of participants to maintain their shopping pace. Further exploration may be warranted to explore how users' motivations and personalities affect their experiences when considering the presentation of recommended products.

RQ2: User perception and trust in RS during shopping. In the study, SoA, perceived workload, and trust in RS were measured. A trend of decreasing SoA with higher perceptibility was observed, although no significant differences were confirmed among the three presentation methods. However, some participants noted that the Highlight attracted excessive attention, compelling them to check recommended paintings and diminishing their experience. In addition, previous research [2] suggested that subjective measures of SoA may be compensated by other factors such as workload and usability, potentially moderating differences between methods. Consequently, we propose that SoA remains relevant to user experiences and behavior in VR RS, particularly from an attention guidance perspective. The result of the NASA-TLX questionnaire in Section 4.3 revealed that the Swap method resulted in the least perceived workload, with the least effort requirement across all conditions and less mental demand compared to NoRS and Arrow. Its high perceptibility allowed participants to handle the recommended results comfortably, whereas the rest methods require additional movements and interpretation, leading to increased effort and mental demand.

In terms of trust in RS, results from Section 4.4 showed that participants expressed greater trust in the Swap method compared to the Arrow, and prior trust positively influenced final trust in RS. The interaction effect unveiled that trust in RS was less impacted by the prior trust when using Swap compared to Arrow. We posit that when participants were engaged less with the RS because of the low perceptibility in Arrow, the prior trust played a primary role in the final trust in RS. From a system performance perspective, Lu et al. [22] suggested that prior trust exerted more impact on final trust when system performance was suboptimal. Thus, the experience disparity between Swap and Arrow may also contribute to the observed interaction effect. These findings underscore the importance of considering users' prior trust during initial engagement with RS to build trust effectively.

5.1 Limitations

This study has several limitations. The goods in VR stores merely contained oil paintings, which may not accurately represent the variety found in commercial VR stores. A similar limitation applies to the quantity of goods available. However, the workload reported in the NoRS condition suggests that even a limited selection of similar paintings posed challenges for participants in terms of selection and purchase. The taxonomy of presentation methods may lack precision; for instance, Swap alters the order of paintings in the store. However, this is necessary to enhance the perceptibility without introducing multiple visual cues that could overwhelm the field of view, particularly when numerous recommendation results are present.

6 CONCLUSION

In this paper, to fill the gaps in assisting exploratory shopping, we implemented an RS in a VR store and compared three presentation methods to understand their effects on user experiences, perception, and trust in RS. The Highlight and Swap methods, with their high perceptibility and efficiency, outperformed the Arrow and were preferred by participants. Participants reported the least workload when using Swap for presentation, emphasizing its lower mental demand and effort to find goods. Interview analysis suggested that the SoA could influence overall user experiences and future designs may benefit from focusing on consistent experiences. An adaptive design of presentation that considers users' personalities, motivations, and prior trust is essential for RS in VR shopping. Our study contributes to the future deployment of RS and the development of Trustworthy Recommender Systems (TRS) in VR.

REFERENCES

[1] Ayman Alzayat and Seung Hwan Mark Lee. 2021. Virtual products as an extension of my body: Exploring hedonic and utilitarian shopping value in a virtual reality retail environment. *Journal of Business Research* 130 (2021), 348–363.

[2] Joanna Bergström, Jarrod Knibbe, Henning Pohl, and Kasper Hornbæk. 2022. Sense of agency and user experience: Is there a link? *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 4 (2022), 1–22.

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[4] Donald Eric Broadbent. 1957. A mechanical model for human attention and immediate memory. *Psychological review* 64, 3 (1957), 205.

[5] Li Chen and Pearl Pu. 2005. Trust building in recommender agents. In *Proceedings of the Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces at the 2nd International Conference on E-Business and Telecommunication Networks*. 135–145.

[6] Dan Cosley, Shyong K Lam, Istvan Albert, Joseph A Konstan, and John Riedl. 2003. Is seeing believing? How recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 585–592.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

[8] Cecilia di Sciascio, Vedran Sabol, and Eduardo E Veas. 2015. uRank: Exploring Document Recommendations through an Interactive User-Driven Approach.. In *IntRS@ RecSys*. 29–36.

[9] Kody R Dillman, Terrance Tin Hoi Mok, Anthony Tang, Lora Oehlberg, and Alex Mitchell. 2018. A visual interaction cue framework from video game environments for augmented reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[10] Brian J Fogg. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002, December (2002), 2.

[11] Yingqiang Ge, Shuchang Liu, Zuohui Fu, Juntao Tan, Zelong Li, Shuyuan Xu, Yunqi Li, Yikun Xian, and Yongfeng Zhang. 2022. A survey on trustworthy recommender systems. *ACM Transactions on Recommender Systems* (2022).

[12] Patrick Haggard and Valerian Chambon. 2012. Sense of agency. *Current biology* 22, 10 (2012), R390–R392.

[13] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[14] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[15] Diana Hassouneh and Malaika Brengman. 2011. Shopping in virtual worlds: Perception, Motivation, and Behavior. (2011).

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[17] Chen Ji, Yacheng Li, Rui Li, Fei Jiang, Xiang Li, Wei Lin, Chenglong Zhang, Wei Wang, and Shuyang Wang. 2023. Interactive Recommendation System for Meituan Waimai. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3245–3249.

[18] Yucheng Jin, Nava Tintarev, Nyi Nyi Htun, and Katrien Verbert. 2020. Effects of personal characteristics in control-oriented user interfaces for music recommender systems. *User Modeling and User-Adapted Interaction* 30, 2 (2020), 199–249.

[19] Sherrie YX Komiak and Izak Benbasat. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS quarterly* (2006), 941–960.

[20] Daniel Lange, Tim Claudius Stratmann, Uwe Gruenefeld, and Susanne Boll. 2020. Hivefive: Immersion preserving attention guidance in virtual reality. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[21] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell me where to look: Investigating ways for assisting focus in 360 video. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2535–2545.

[22] Feiyu Lu, Yan Xu, Xuhai Xu, Brennan Jones, and Laird Malamed. 2023. Exploring the Impact of User and System Factors on Human-AI Interactions in Head-Worn Displays. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 109–118.

[23] Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang. 2015. Recommender system application developments: a survey. *Decision support systems* 74 (2015), 12–32.

[24] Sean M McNee, John Riedl, and Joseph A Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*. 1097–1101.

[25] Daniel Medeiros, Rafael Dos Anjos, Nadia Pantidi, Kun Huang, Maurício Sousa, Craig Anslow, and Joaquim Jorge. 2021. Promoting reality awareness in virtual reality through proxemics. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 21–30.

[26] Manon Mulckhuyse and Jan Theeuwes. 2010. Unconscious attentional orienting to exogenous cues: A review of the literature. *Acta psychologica* 134, 3 (2010), 299–309.

[27] Theodora Nanou, George Lekakos, and Konstantinos Fouskas. 2010. The effects of recommendations' presentation on persuasion and satisfaction in a movie recommender system. *Multimedia systems* 16 (2010), 219–230.

[28] Lasse T Nielsen, Matias B Møller, Sune D Hartmeyer, Troels CM Ljung, Niels C Nilsson, Rolf Nordahl, and Stefania Serafin. 2016. Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*. 229–232.

[29] Nahal Norouzi, Gerd Bruder, Austin Erickson, Kangsoo Kim, Jeremy Bailenson, Pamela Wisniewski, Charlie Hughes, and Greg Welch. 2021. Virtual Animals as Diegetic Attention Guidance Mechanisms in 360-Degree Experiences. *IEEE Transactions on Visualization and Computer Graphics* 27, 11 (2021), 4321–4331. https://doi.org/10.1109/TVCG.2021.3106490

[30] A Ant Ozok, Quyin Fan, and Anthony F Norcio. 2010. Design guidelines for effective recommender system interfaces based on a usability criteria conceptual model: results from a college student population. *Behaviour & Information Technology* 29, 1 (2010), 57–83.

[31] Christian Peukert, Jella Pfeiffer, Martin Meißner, Thies Pfeiffer, and Christof Weinhardt. 2019. Shopping in virtual reality stores: the influence of immersion on system adoption. *Journal of Management Information Systems* 36, 3 (2019), 755–788.

[32] Jella Pfeiffer, Thies Pfeiffer, Martin Meißner, and Elisa Weiß. 2020. Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. *Information Systems Research* 31, 3 (2020), 675–691.

[33] Kaitong Qin, Yankun Zhen, Tianshu Dong, Liuqing Chen, Ting Zhang, Yumou Zhang, and Tingting Zhou. 2023. Design and User Experience Evaluation of 3D Product Information in XR Shopping Application. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 87295. American Society of Mechanical Engineers, V002T02A060.

[34] Sylvia Rothe, Daniel Buschek, and Heinrich Hußmann. 2019. Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction* 3, 1 (2019), 19.

[35] Quentin Roy, Futian Zhang, and Daniel Vogel. 2019. Automation accuracy is good, but high controllability may be better. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.

[36] Michelle L Rusch, Mark C Schall Jr, Patrick Gavin, John D Lee, Jeffrey D Dawson, Shaun Vecera, and Matthew Rizzo. 2013. Directing driver attention with augmented reality cues. *Transportation research part F: traffic psychology and behaviour* 16 (2013), 127–137.

[37] Anastasia Schmitz, Andrew MacQuarrie, Simon Julier, Nicola Binetti, and Anthony Steed. 2020. Directing versus attracting attention: Exploring the effectiveness of central and peripheral cues in panoramic videos. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 63–72.

[38] Martin Schrepp, Jörg Thomaschewski, and Andreas Hinderks. 2017. Construction of a benchmark for the user experience questionnaire (UEQ). (2017).

[39] Rashmi R Sinha, Kirsten Swearingen, et al. 2001. Comparing recommendations made by online systems and friends. *DELOS* 106, 1 (2001), 1–6.

[40] Rébaï Soret, Pom Charras, Christophe Hurter, and Vsevolod Peysakhovich. 2019. Attentional orienting in virtual reality using endogenous and exogenous cues in auditory and visual modalities. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–8.

[41] Alexis D Souchet, Stéphanie Philippe, Domitile Lourdeaux, and Laure Leroy. 2022. Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. *International Journal of Human–Computer Interaction* 38, 9 (2022), 801–824.

[42] Marco Speicher, Sebastian Cucerca, and Antonio Krüger. 2017. VRShop: a mobile interactive virtual reality shopping environment combining the benefits of on- and offline shopping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–31.

[43] Marco Speicher, Philip Hell, Florian Daiber, Adalberto Simeone, and Antonio Krüger. 2018. A virtual reality shopping experience using the apartment metaphor. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. 1–9.

[44] Marco Speicher, Christoph Rosenberg, Donald Degraen, Florian Daiber, and Antonio Krúger. 2019. Exploring visual guidance in 360-degree videos. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. 1–12.

[45] Alain Starke, Martijn Willemsen, and Chris Snijders. 2017. Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system. In *Proceedings of the eleventh ACM conference on recommender systems*. 65–73.

[46] Bethesda Game Studios. 2023. *Starfield*. Game [Computer software]. Bethesda Softworks, Rockville, Maryland, United States.

[47] Yiming Sun, Yang Zhan, and Tatsuo Nakajima. 2024. Investigating User Experience in Virtual Goods Shopping through a VR Diegetic In-Game Store. In *Proceedings of the 2024 International Conference on Advanced Visual Interfaces*. 1–5.

[48] Adam Tapal, Ela Oren, Reuven Dar, and Baruch Eitam. 2017. The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in psychology* 8 (2017), 1552.

[49] Jan Theeuwes. 1991. Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & psychophysics* 49, 1 (1991), 83–90.

[50] Wesley Waldner and Julita Vassileva. 2014. Emphasize, don't filter! displaying recommendations in twitter timelines. In *Proceedings of the 8th ACM Conference on Recommender systems*. 313–316.

[51] Shoujin Wang, Xiuzhen Zhang, Yan Wang, and Francesco Ricci. 2022. Trustworthy recommender systems. *ACM Transactions on Intelligent Systems and Technology* (2022).

[52] Austin Ward, Sandeep Avula, Hao-Fei Cheng, Sheikh Muhammad Sarwar, Vanessa Murdock, and Eugene Agichtein. 2023. Searching for products in virtual reality: Understanding the impact of context and result presentation on user experience. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2359–2363.

[53] Austin R Ward and Rob Capra. 2020. Immersive search: Using virtual reality to examine how a third dimension impacts the searching process. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1621–1624.

[54] I-Chin Wu and Hsin-Kai Yu. 2020. Sequential analysis and clustering to investigate users' online shopping behaviors based on need-states. *Information Processing & Management* 57, 6 (2020), 102323.

[55] Nannan Xi, Juan Chen, Filipe Gama, Marc Riar, and Juho Hamari. 2023. The challenges of entering the metaverse: An experiment on the effect of extended reality on workload. *Information Systems Frontiers* 25, 2 (2023), 659–680.

[56] Bereket A Yilma and Luis A Leiva. 2023. The elements of visual art recommendation: Learning latent semantic representations of paintings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[57] Kyung-Hyan Yoo, Ulrike Gretzel, and Markus Zanker. 2015. Source factors in recommender system credibility evaluation. *Recommender Systems Handbook* (2015), 689–714.

[58] Jiyong Zhang, Nicolas Jones, and Pearl Pu. 2008. A visual interface for critiquing-based recommender systems. In *Proceedings of the 9th ACM conference on Electronic commerce*. 230–239.

[59] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. 22–32.