Order Effects and Contextual Factors in Relevance Judgment for Text Information Retrieval (Postprint)

Authors: Wang Zhihong, Cao Shujin

Date: 2023-04-01T00:00:00+00:00

Abstract

Purpose/Significance: Ranking search results in a linear list format based on relevance scores is the predominant approach employed by most information retrieval systems. Although existing research has theoretically criticized the limitations of this approach, pointing out that sequentially examined and read documents mutually influence one another, no empirical study has further investigated the contextual factors underlying order effects in relevance judgment. Method/Process: To address this gap, a 2$×2×$2 between-subjects experiment was conducted, recruiting eighty participants who each completed experimental tasks on four topics. Data were collected via questionnaires and other means before and after the experimental tasks, and multiple analytical methods including chi-square tests, violation of the total probability formula, and the QQ equation were employed to investigate order effects in relevance judgment and the influence of contextual factors such as inter-document relationships and document presentation formats. Results/Conclusion: The results demonstrate that document presentation order affects relevance judgment, and this effect varies depending on inter-document relationships and document presentation formats. Specifically, documents with mutually exclusive relationships are more susceptible to order effects in relevance judgment than those with complementary relationships; additionally, document snippets are more susceptible to such order effects than full-text documents. Although this study has certain limitations, its conclusions provide implications for the theory, practice, and research methodology of text information retrieval.

Full Text

A Study on Order Effects in Relevance Judgment and Contextual Factors in Textual Information Retrieval

Wang Zhihong[1],[2], **Cao Shujin**[3]

[1]Department of Computer Science and Technology, Tsinghua University, Beijing 100084
[2]Beijing National Research Center for Information Science and Technology, Beijing 100084
[3]School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract: [Purpose/Significance] Ranking search results linearly based on relevance scores is the primary approach adopted by most current information retrieval systems. Although existing research has theoretically criticized the limitations of this method, pointing out that examined and read documents mutually influence each other, no empirical studies have further investigated the contextual factors underlying order effects in relevance judgment. [Method/Process] To address this gap, this study employed a 2$×2×$2 between-subjects experimental design, recruiting 80 participants who each completed experimental tasks across 4 topics. Data were collected through questionnaires administered before and after task completion, and analyzed using multiple methods including chi-square tests, violations of the law of total probability, and QQ equality to explore order effects in relevance judgment and the influence of contextual factors such as inter-document relationships and document presentation formats. [Result/Conclusion] The results indicate that document presentation order affects relevance judgment, and this effect varies depending on inter-document relationships and presentation formats. Specifically, the study found that compared to documents with complementary relationships, the presentation order of documents with contradictory relationships more strongly influences relevance judgment. Additionally, compared to full-text documents, the presentation order of document snippets more readily affects relevance judgment. While this study has certain limitations, its conclusions offer theoretical, practical, and methodological insights for textual information retrieval.

Keywords: relevance judgment; order effect; inter-document relationship; document presentation type; quantum cognition

---

1 Literature Review

1.1 Position Factors in Search Result Relevance Judgment Since information needs are inherently subjective and difficult to define clearly, relevance is strongly user-oriented while also being a multi-dimensional and measurable concept. In information retrieval systems that primarily use list-based ranking,

the position or order of documents in search result lists affects users' perception and evaluation of relevance. Existing research has examined how search result ordering influences user behavior such as examination and clicking, which can be considered an indirect and explicit reflection of users' relevance assessment. This study primarily reviews research directly investigating the impact of document presentation order on relevance judgment.

Regarding the influence of document presentation order on user relevance judgment, M. Eisenberg and C. Barry found through user experiments that users overestimated and underestimated document relevance under ascending and descending relevance orderings, respectively. Building on this, M. Huang and H. Wang further investigated order effects and their relationship with document set size, discovering that order significantly affected relevance judgments only for document sets containing 15-60 items, with no significant effect for sets smaller than 15 or larger than 60. Subsequently, Y. Xu and D. Wang conducted a two-stage simulated user experiment where the second-stage document set and order were determined by user behavior in the first stage, finding that order effects existed but were small.

As relevance is a multi-dimensional concept, some studies have explored order effects in relevance judgment from the perspective of judgment dimensions. For example, P. Bruza and V. Chang proposed that each dimension of relevance judgment could be viewed as a subspace in a Hilbert space. If user problem states induced by different documents are projected onto the overall relevance plane through different dimensions, these documents would receive different relevance scores. By collecting data on 5 queries from a crowdsourcing platform and judging each query from two dimensions (topicality, comprehensibility, interest, etc.), they confirmed this hypothesis. S. Upreti and D. Song expanded relevance from 2 to 7 dimensions and used query log data and user experiments to test order effects across relevance dimensions. The results confirmed the incompatibility between relevance dimensions and the existence of order effects, speculating that these effects might relate to other factors. Additionally, T. Damesi et al. investigated how order effects (including descending order and identifier-based ordering) influenced inter-assessor agreement and judgment speed, finding that document presentation order affected both agreement and overall speed, with assessors spending less time and achieving higher consistency when documents were ordered by identifier.

1.2 Influence of Inter-Document Relationships and Presentation Formats on Search Result Ranking Regarding inter-document relationships, although early research questioned the independence assumption in information retrieval, progress remains limited on how inter-document relationships affect user relevance judgment and how to incorporate these relationships into result ranking. For instance, A. Bookstein noted that probabilistic deterministic models overlooked inter-document interaction effects and argued that learning effects from examined documents and their influence on subsequent document

evaluation should be considered in sequential retrieval models. D. R. Swanson proposed several intractable assumptions in information science, one being that documents are not independent but exist in a network of relationships, causing corresponding judgments to exist in a "shift network." Further, M. A. Tiamiyu and I. Y. Ajiferuke theoretically explored how inter-document relationships affect users' perception of individual document relevance, proposing a document interaction model that considers three types of relationships: no relationship, substitutive relationships (such as duplication and redundancy), and complementary relationships (discussing different but closely related aspects of a specific topic). They recommended that future research should more deeply explore inter-document relationships and their impact on relevance perception.

Beyond inter-document relationships, document presentation format is another important factor to consider. Existing research has primarily examined how relevance judgments and applied dimensions differ across different presentation formats of the same document, but has not explored differences in order effects across presentation formats. For example, P. Wang et al. proposed a document selection model describing the decision-making process from bibliographic records to reading and citation, noting that only 51.5% and 4.9% of search results are read and cited, respectively, with different criteria applied at different stages. P. Vakkari and N. Hakala investigated differences in document relevance assessment between metadata surrogates and full-text formats, finding that users' application of relevance criteria varied with presentation format during information seeking. Specifically, topical information was more important in metadata-based judgment than in full-text judgment, while physical characteristics, information type, and user context showed the opposite pattern. I. Xie and E. Benoit examined similarities and differences in criteria usage between result list evaluation and full document evaluation, finding that more criteria and elements were used in full-text evaluation, which also required more time. Y. Wang et al. further used eye-tracking to compare search result list and single document evaluation modes, finding differences in fixation patterns and transition patterns, with full-text documents providing more information and generating stronger learning effects than metadata formats.

1.3 Order Effects in Human Judgment and Cognitive Psychology Explanations In the field of human judgment and decision-making, researchers have deeply investigated and explained the psychological and cognitive mechanisms underlying order effects, proposing corresponding theories. For instance, the renowned psychologists and economists A. Tversky and D. Kahneman proposed anchoring and adjustment theory in 1974 to explain judgment behavior under uncertainty. This theory describes how people use initially received numbers or information as anchors and then adjust based on subsequently encountered information. Subsequent research has extensively investigated this phenomenon. R. M. Hogarth and H. J. Einhorn proposed belief adjustment theory, interpreting order effects as resulting from belief adjustment and updating. Later research noted that belief adjustment theory' s predictive power has

certain limitations in tasks involving extreme evidence.

While theoretical explanations for order effect mechanisms differ, researchers generally acknowledge the phenomenon' s existence. Typically, the two events triggering this effect are considered incompatible or non-commutative. Classical probability theory based on set theory (which disregards event order) has limitations in explaining incompatible events, such as conjunction fallacies or disjunction fallacies that are difficult to explain with traditional methods. Similarly, quantum characteristics like incompatibility have prompted researchers to reconsider classical probability or Bayesian theory, proposing quantum inference models to explain the fuzziness or irrationality of human thought. In 1999, A. Khrennikov began using quantum physics-related theories to explain cognitive or social phenomena, subsequently applying quantum-like models to cognitive decision-making and information processing (such as prisoner' s dilemmas or disjunction fallacies). Generally, quantum theory' s key feature is its ability to model and map using orthogonal subspaces in Hilbert space, providing the capacity to explain incompatible events whose order effects cannot be explained by traditional probability theory that ignores event sequence.

In summary, current information retrieval research on order effects in relevance judgment has primarily focused on how presentation order affects judgment outcomes, consistency, speed, and dimensions. However, few studies have further explored the contexts in which these order effects occur. In particular, although researchers have long recognized the importance of inter-document relationships and presentation formats, no studies have examined these two factors as judgment contexts to explore their potential influence on order effects. Meanwhile, cognitive psychology and other fields have conducted extensive research on order effects, providing a foundation for this study' s exploration of order effects and contextual factors in relevance judgment.

---

2 Research Design and Methods

2.1 Experimental Design and Material Selection This study involves three main variables: document presentation order, inter-document relationship, and document presentation format. The basic hypothesis is that presentation order affects relevance judgment outcomes, and this influence is moderated by inter-document relationships (complementary and contradictory) and presentation formats (full-text and snippets). First, for documents with complementary or contradictory relationships, different presentation orders may differentially affect relevance judgment. Second, compared to presenting only snippets, full-text documents may produce stronger learning effects due to increased information volume, resulting in different order effects. Therefore, this study employs a 2$×2×$2 between-subjects experimental design with eight experimental groups, with participants randomly assigned to these groups.

The experiment requires participants to complete tasks across 4 topics. To

balance completion time and reduce pressure from other participants, each participant receives materials with 2 topics presented in full-text format and 2 in snippet format. Since the complementary and contradictory relationships considered in this study primarily concern viewpoints, controversial topics are most suitable as experimental tasks. After surveying topics used in controversial text comprehension research and considering their appropriateness and popularity for the experimental context, four final topics were selected (see Table 1).

As currently available standard datasets do not consider inter-document relationships, most experimental materials for these four topics were selected from English Wikipedia. The primary reason is that English Wikipedia contains dedicated entries for numerous controversial issues, organized by different perspectives with clear topical structures, making it relatively easy to find subtopics with contradictory and complementary relationships. Each subtopic needed to have a corresponding entry to construct document pairs. Nevertheless, some topics could not meet these conditions (e.g., lacking separate encyclopedia entries), so relevant news articles found through Google searches were used as supplementary materials. To save time, ensure readability, and reduce experimental fatigue, all documents were controlled to 200-300 words in length.

Topic 1: Global Warming (the causes of global warming)
Topic 2: Safety of Genetically Modified Food (the safety of genetically modified food)
Topic 3: Smartphone Radiation (radiation of smart phone)
Topic 4: Effects of Video Games (the effects of video game)

2.2 Experimental Procedure This study recruited participants through convenience sampling. These participants were undergraduates from a "double first-class" university' s information management school enrolled in a particular course, ensuring homogeneity in reading ability, English proficiency, and professional background. With the instructor' s consent, the experiment was conducted after class in November 2018. Of the 90 registered students, 80 participated. After excluding 2 students who did not submit questionnaires, 78 participants (34 male, 44 female) submitted 312 questionnaires (78 participants $\times$ 4 tasks each). Five questionnaires with missing values were excluded, leaving 307 questionnaires for analysis.

Since experimental materials were in English, the questionnaire included questions about English proficiency. Most participants (60, 76.9%) had passed the College English Test Band 4, indicating adequate reading skills. Additionally, relatively difficult vocabulary items were translated into Chinese to avoid experimental bias from language barriers. A small-scale pretest was conducted before the formal experiment to ensure material appropriateness and prevent comprehension issues.

To reduce bias from participants' trust in search engines, the experiment used a paper-and-pencil format. All materials were arranged in sequence and dis-

tributed at planned time points. To ensure compliance, researchers emphasized that participants must follow the given material order, and two teaching assistants monitored the process in real-time to prevent non-compliance.

The questionnaire included a pre-experiment section for collecting personal information (gender, English ability, etc.) and a main survey for collecting relevance judgments and rationales. Since quantum cognition theory' s key assumption holds that cognitive context changes if other questions exist before or between target questions, producing inaccurate measurements, the pre-experiment and main questionnaires were administered separately (before and after a 45-minute class session) to maintain consistent cognitive contexts.

During the main experiment, participants received materials including 4 task descriptions, 8 pre-selected documents (2 per task), and subsequent relevance judgment items using a 4-point scale (irrelevant, not very relevant, partially relevant, relevant). A 4-point scale was chosen because it is commonly used in research, and prior studies found that users typically create 3-5 categories when classifying search results. Two open-ended questions followed each relevance judgment, asking participants to explain their assessments to deepen understanding of their reasoning.

2.3 Data Analysis Methods

2.3.1 Law of Total Probability in Classical Probability Theory In classical probability theory, events are represented as sets, with all elementary events composed of sample spaces using Boolean logical operators. This set-theoretic probability model, established by A. Kolmogorov in 1933, is known as the Kolmogorov model. Its fundamental theorem is the law of total probability, built upon conditional probability—where conditional probability refers to the probability of event A occurring given that event B has occurred. The simplest case occurs when event B has only two outcomes (B_1 and B_2), as shown in Formula (2).

Research shows that when considering quantum probability in quantum mechanics, the law of total probability based on traditional Boolean logical operators can be violated. This means that when quantum-like interference exists between two events (i.e., event B' s occurrence affects event A' s probability, or the two events are non-independent/non-commutative), event probabilities cannot be represented using set theory. In such cases, the left and right sides of the equation are unequal. Order effects, or conjunction fallacies in probability judgment, are typical phenomena of interference between two non-commutative events. Therefore, this study applies the law of total probability to detect whether quantum-like interference effects exist in order effects of relevance judgment.

2.3.2 Quantum Equality Test (QQ Test) To overcome classical probability theory' s limitations in explaining non-commutative events, Z. Wang et al. explored order effects in question pairs from 70 national surveys to reveal

quantum characteristics of human cognition in judgment and decision-making. By examining data where questions with yes/no answers were presented in different orders, they found that the difference in the sum of main diagonal values in contingency tables obtained under two different orders approached zero. Based on this, they proposed the quantum question test (QQ test), with calculations shown in Table 2 and Formula (3).

In Table 2, the A-B order represents event A preceding event B, while B-A represents the opposite. Ar and Bn denote yes and no responses for events A and B, respectively. Each order thus yields four possible combinations. For example, the A-B order produces ArBr, ArBn, AnBr, and AnBn, with probabilities P(ArBr), P(ArBn), P(BrAn), and P(AnBn). The QQ test examines whether cognitive context changes induced by different orders, or whether order effects in the first order can be offset by those in the second order. The q-value measures this offset degree, which should approach zero when order effects exist. The prerequisite assumption is that no factors other than the preceding event affect the subsequent event. Some researchers suggest the QQ test is a sufficient but not necessary condition for lack of contextuality. Overall, the QQ test is a widely used quantum probability method that, unlike the Boolean logic-based law of total probability, considers possible combinations of event values and is more suitable for exploring how contextual factors affect order effects.

---

3 Results

3.1 Chi-Square Test Results

3.1.1 Overall and Topic-Specific Relevance Judgment Distributions

In the experiment, participants used a 4-point scale to evaluate two documents (Document A and Document B) presented in different orders for each topic: Document A→Document B and Document B→Document A. For Document A, the Document A→Document B order constitutes a non-comparison scenario where Document A' s relevance judgment is unaffected by other documents. The Document B→Document A order represents a comparison scenario. The opposite applies to Document B.

Table 3 shows the contingency table for relevance judgments of Documents A and B in comparison and non-comparison scenarios, along with differences between scenarios. Differences primarily manifest in three ways: (1) Compared to non-comparison scenarios, comparison scenarios show increased "relevant" judgments and decreased "partially relevant" judgments (for overall data and Topics T1, T4 for Document A and T3 for Document B), or increased "partially relevant" and decreased "not very relevant" judgments (for Topic T2, Document A)—indicating an enhancement effect; (2) Comparison scenarios show decreased "partially relevant" judgments while both "relevant" and "not very relevant" judgments increase (for overall data and Topics T2, T4 for Document B), with

"not very relevant" showing more pronounced increases—suggesting simultaneous weakening and enhancement effects; (3) Minimal differences between scenarios (for Topic T3, Document A)—indicating no clear enhancement or weakening.

Chi-square tests examined these differences' significance. Overall, Document B' s relevance judgments differed significantly between scenarios ($^2 = 8.377$, $p = 0.039 < 0.05$), while Document A' s differences were not significant ($^2 = 4.568$, $p = 0.206 > 0.05$). By topic, only Topic 2 showed significant differences for Document B ($^2 = 9.411$, $p = 0.024 < 0.05$).

3.1.2 Differences Across Inter-Document Relationships and Presentation Formats To test whether order effects differ across inter-document relationships and presentation formats, separate chi-square tests were conducted (see Tables 4 and 5). Note that while A and B represent document pairs in both relationship conditions, the experimental materials differed, preventing direct comparison.

Overall, order significantly affected Document B' s relevance judgments but not Document A' s. Regarding inter-document relationships, when documents had contradictory relationships, Document B' s chi-square test was significant ($^2 = 8.363$, $p = 0.039 < 0.05$), with Document A approaching marginal significance ($^2 = 7.389$, $p = 0.06$). For complementary relationships, neither document showed significance. Regarding presentation format, snippet-format Document B showed significant differences between scenarios ($p = 0.05$), while other conditions were non-significant.

3.2 Violation of the Law of Total Probability Since relevance judgment score distributions are left-skewed, with scores concentrated in highly relevant intervals, and because this study focuses more on relevance than non-relevance assessment, the 4-point scale was converted to a binary scale by collapsing scores 1-3 into "non-relevant" and score 4 into "relevant" for subsequent analysis.

After conversion, the law of total probability was applied to calculate probabilities and differences between comparison and non-comparison scenarios (see Table 6). p(r) and p(r) represent the probability of Document i or j being judged relevant in non-comparison scenarios. p(r |r) is the conditional probability of Document i being relevant given Document j is relevant, while p(r |¬r) is the conditional probability given Document j is non-relevant. p' (r) is the calculated probability using Formula (2). Groups 1-4 had contradictory relationships, while Groups 5-8 had complementary relationships.

Results show 10 of 16 differences exceeded 0.1, confirming violations of the law of total probability and indicating that document order affects relevance judgment. Among differences >0.1, 4 were from contradictory groups and 6 from complementary groups, with the maximum difference being 0.34. Three negative values, all from complementary groups, indicated reduced relevance probabilities compared to non-comparison scenarios. A t-test analyzing differences between

scenarios yielded p = 0.003 (df = 15) with Pearson correlation coefficient r = 0.809 ($p < 0.001$), confirming that document order significantly affects relevance judgment.

3.3 QQ Equality Test For Document A→Document B or Document B→Document A orders, four events exist: ArBr/BrAr, ArBn/BnAr, AnBr/BrAn, and AnBn/BnAn. After data conversion, probabilities for each event and q-values based on the QQ equality were calculated by topic, inter-document relationship, and presentation format (see Table 7).

Only Topic 3' s q-value was significantly below 0.05, while other topics' q-values exceeded 0.05. Additionally, q-values for contradictory relationships and snippet presentation formats were below 0.05, passing the QQ test. This suggests that under these conditions, order effects in relevance judgment can indeed be modeled using quantum theory. Non-significant results may violate the QQ test' s reciprocity principle assumption—that the preceding question is the sole factor affecting context for the subsequent question. Despite experimental controls, documents are complex objects, and participants' background knowledge, beliefs, or preferences may influence cognitive changes from the first document, reducing its impact on the second document' s judgment. Noise in realistic experimental settings may also diminish observable effects.

---

4 Discussion and Conclusions

4.1 Discussion of Results 4.1.1 Influence of Document Presentation Order on Relevance Judgment

Quantitative analysis demonstrates that presentation order significantly affects relevance judgment. Using chi-square tests, law of total probability violations, and QQ equality, the study found order effects exist even with only two documents, with effects related to document characteristics including topic. This aligns with M. Huang and H. Wang and Y. Xu and D. Wang' s findings, though Huang and Wang suggested order effects only occur within specific document quantity ranges—a question requiring further exploration.

Qualitative analysis of open-ended responses revealed that order effects may relate to comparison and learning effects between documents. Due to limited cognitive capacity and time, users compare different documents to identify more useful information. One participant noted when evaluating the second document: "Overall, more objective than the first." Another stated: "···provides more objective reasons about GMO safety, unlike the subjective examples in the previous article." Order effects may also relate to learning effects, as participants used content from the first document to evaluate the second, potentially increasing relevance assessments.

4.1.2 Contexts for Order Effects in Relevance Judgment

This study investigated not only whether order effects exist but also how they relate to inter-document relationships and presentation formats.

Inter-document relationships: Compared to complementary documents, contradictory documents' presentation order more strongly influences relevance judgment. While M. A. Tiamiyu and I. Y. Ajiferuke theoretically proposed that overall relevance depends on relationships with previously examined documents (including redundancy and complementarity), no empirical studies have tested this influence, making this finding novel but requiring further validation.

Presentation format: Compared to full-text documents, snippet presentation order more readily affects relevance judgment. Since snippets contain less information for judgment, users form less stable beliefs that are more susceptible to influence. While I. Xie and E. Benoit confirmed that criteria and elements differ across presentation formats, this study is the first to demonstrate that these differences moderate order effects, likely because different formats involve different judgment standards and elements.

4.2 Research Conclusions Using convenience sampling, 80 participants completed user experiments with 4 tasks each. By controlling presentation order, inter-document relationships, and presentation formats, this first-of-its-kind study examined how these three factors simultaneously influence user relevance judgment. The study found order effects exist even with only two documents, with inter-document relationships and presentation formats serving as contextual factors. Specifically, order effects are more pronounced for contradictory than complementary documents, and more significant for snippets than full-text documents.

4.3 Implications and Limitations 4.3.1 Research Value and Implications

(1) Theoretical implications: Relevance is a core, human-centered concept in information retrieval. While its dynamic nature is typically explained through information need dynamics and "search as learning" perspectives, few studies link this dynamic nature to interactions between user cognitive context and information objects. Future research should investigate these interaction mechanisms, particularly how different document relationships affect the information interaction process and relevance judgment.

(2) System design implications: Most retrieval systems assume document independence, calculating and ranking relevance individually. This study empirically demonstrates this assumption' s unreasonableness and identifies specific conditions where it fails. For complementary information items, systems should consider aggregation to enhance overall utility through learning or attraction effects. For contradictory items, systems should avoid information narrowing and filter bubbles by diversifying results through interleaved or parallel presentation, facilitating comparison and increasing marginal utility of lower-ranked

documents. For snippets, systems should more effectively describe document content to help users form stable beliefs.

(3) Methodological implications: This study applied quantum cognitive psychology methods to comprehensively test order effects and their contexts, demonstrating the promise of quantum-theoretic statistical methods in information retrieval research. Given information retrieval' s cognitive and contextual orientations involving extensive human cognition and situational issues, quantum theory' s complementarity principle—where measurement order affects outcomes—aligns naturally with relevance judgment order effects.

4.3.2 Limitations

This study' s limitations include: (1) convenience sampling of students from one university, limiting sample scope and size; (2) material selection subjectivity due to lack of standardized datasets considering inter-document relationships, a common challenge in information retrieval experiments. Future research should investigate multimodal retrieval, vertical domains, and different document and user characteristics.

Acknowledgments: We sincerely thank Professor Jaana Kekäläinen and Professor Pertti Vakkari from the University of Tampere' s Information Retrieval Research Group for their guidance and assistance in conducting the experiment and analyzing results.

References: [1] SARACEVIC T. Relevance: a review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance[J]. Journal of the American Society for Information Science and Technology, 2007, 58(13): 1915-1933. [2] EISENBERG M, BARRY C. Order effects: a study of the possible influence of presentation order on user judgments of document relevance[J]. Journal of the American Society for Information Science, 1988, 39(5): 293-300. [3] PANG L, AI Q, XU J. Beyond probability ranking principle: modeling the dependencies among documents[C]//Proceedings of the 14th ACM international conference on Web search and data mining. New York: ACM, 2021: 1137-1140. [4] TIAMIYU M A, AJIFERUKE I Y. A total relevance and document interaction effects model for the evaluation of information retrieval processes[J]. Information processing & management, 1988, 24(4): 391-404. [5] WANG Siqi, GUO Jinyun. Question order effects in public service satisfaction measurement: evidence from an embedded survey experiment[J]. Public Management Review, 2020, 2(1): 92-115. [6] TOURANGEAU R, RASINSKI K A. Cognitive processes underlying context effects in attitude measurement[J]. Psychological bulletin, 1988, 103(3): 299. [7] SUN Yu, LU Chun, SI Jiwei, et al. Age differences in how task presentation mode and order affect adolescents' arithmetic strategy use[J]. Psychological Research, 2017, 10(1): 39-49. [8] XU Xing' an, WANG Lilei, GAO Fengqiang. Effects of tourism advertising and online negative word-of-mouth on destination image—an experimental test of order and interaction effects[J]. Tourism Tribune,

2017, 32(12): 37-48. [9] ERT E, FLEISCHER A. Mere position effect in booking hotels online[J]. Journal of travel research, 2016, 55(3): 311-321. [10] HUO Hangfeng. The effect of item presentation order on judgment of learning[D]. Hangzhou: Zhejiang Normal University, 2017. [11] PURNAWIRAWAN N, DEPELSMAECKER P, DENS N. Balance and sequence in online reviews: how perceived usefulness affects attitudes and intentions[J]. Journal of interactive marketing, 2012, 26(4): 244-255. [12] XIE Mengjiao. The relationship between order effects in additive reviews and consumer purchase intention[D]. Beijing: Beijing University of Posts and Telecommunications, 2018. [13] XU Zhihui. The effect of presentation order of declarative and procedural knowledge in teaching videos of different difficulty on learning[D]. Wuhan: Central China Normal University, 2019. [14] WANG Jiawei, DENG Yongshan, XING Qiang. Effects of learning sequence and material similarity on auditory category learning[J]. Studies of Psychology and Behavior, 2019, 17(6): 750-755. [15] BERGUS G R, LEVIN I P, ELSTEIN A S. Presenting risks and benefits to patients: the effect of information order on decision making[J]. Journal of general internal medicine, 2002, 17(8): 612-617. [16] YANG Wenting. The impact of information presentation on individual investment decisions[D]. Hangzhou: Zhejiang University, 2019. [17] WANG Z, SOLLOWAY T, SHIFFRIN R M, et al. Context effects produced by question orders reveal quantum nature of human judgments[J]. Proceedings of the National Academy of Sciences, 2014, 111(26): 9431-9436. [18] CHU H. Factors affecting relevance judgment: a report from TREC legal track[J]. Journal of documentation, 2011, 67(2): 264-278. [19] XU Y, WANG D. Order effect in relevance judgment[J]. Journal of the American Society for Information Science and Technology, 2008, 59(8): 1264-1275. [20] GOLDMAN E. Search engine bias and the demise of search engine utopia[M]//Web Search. Berlin: Springer, 2008: 121-133. [21] HUANG M, WANG H. The influence of document presentation order and number of documents judged on users' judgments of relevance[C]//Proceedings of the 27th ACM international conference on information and knowledge management. New York: ACM, 2016: 1137-1140. [22] DAMESI E T T, SCHOLER F, JÄRVELIN K, et al. The effect of document order and topic difficulty on assessor agreement[C]//Proceedings of the 2016 ACM international conference on the theory of information retrieval. New York: ACM, 2016: 41-48. [23] BRUZA P, CHANG V. Perceptions of document relevance[J]. Frontiers in psychology, 2014, 5(612): 1-8. [24] ZUCCON G, AZZOPARDI L. Using the quantum probability ranking principle to rank interdependent documents[C]//European conference on information retrieval. Berlin: Springer, 2010: 357-369. [25] LI Jingfei. A quantum interaction model for exploratory information retrieval[D]. Tianjin: Tianjin University, 2017. [26] MA Q, NADAMOTO A, TANAKA K. Complementary information retrieval for cross-media news content[J]. Information systems, 2006, 31(7): 659-678. [27] SCHAMBER L, EISENBERG M B, NILAN M S. A re-examination of relevance: toward a dynamic, situational definition[J]. Information processing & management, 1990, 26(6): 755-776. [28] SCHAMBER L. Relevance and information behavior[J]. Annual review of information science and technology, 1994, 29(1): 3-48. [29] UPRETI S, SONG D. Investigating order effects in multi-

dimensional relevance judgment using query logs[C]//Proceedings of the 2018 ACM SIGIR international conference on theory of information retrieval. New York: ACM, 2018: 191-194. [30] LI J, ZHANG P, SONG D, et al. Understanding an enriched multidimensional user relevance model by analyzing query logs[J]. Journal of the Association for Information Science and Technology, 2017, 68(12): 2743-2754. [31] DAMESI E T T, SCHOLER F, CULPEPPER J S. The influence of topic difficulty, relevance level, and document ordering on relevance judging[C]//Proceedings of the 21st Australasian document computing symposium. New York: ACM, 2016: 1-8. [32] BOOKSTEIN A. Information retrieval: a sequential learning process[J]. Journal of the American Society for information Science, 1983, 34(5): 331-342. [33] SWANSON D R. Historical note: information retrieval and the future of an illusion[J]. Journal of the American Society for Information Science, 1988, 39(2): 92-98. [34] WANG P, SOERGEL D. A cognitive model of document use during a research project. Study I. Document selection[J]. Journal of the American Society for Information Science, 1998, 49(2): 115-133. [35] WANG P, DOMAS WHITE M. A cognitive model of document use during a research project. Study II. Decisions at the reading and citing stages[J]. Journal of the American Society for Information Science, 1999, 50(2): 98-114. [36] VAKKARI P, HAKALA N. Changes in relevance criteria and problem stages in task performance[J]. Journal of documentation, 2000, 56(5): 540-562. [37] XIE I, BENOIT E. Search result list evaluation versus document evaluation: similarities and differences[J]. Journal of documentation, 2013, 69(1): 49-80. [38] WANG Y, XIE I, LEE S. Explore eye movement patterns in search result evaluation and individual document evaluation[J]. Proceedings of the Association for Information Science and Technology, 2015, 52(1): 1-4. [39] TVERSKY A, KAHNEMAN D. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment[J]. Psychological review, 1983, 90(4): 293. [40] KHRENNIKOV A. Quantum-like modeling of cognition[J]. Frontiers in physics, 2015, 3(77): 1-18. [41] TVERSKY A, KAHNEMAN D. Judgment under uncertainty: Heuristics and biases[J]. Science, 1974, 185(4157): 1124-1131. [42] HOGARTH R M, EINHORN H J. Order effects in belief updating: the belief-adjustment model[J]. Cognitive psychology, 1992, 24(1): 1-55. [43] TRUEBLOOD J, BUSEMEYER J. A comparison of the belief-adjustment model and the quantum inference model as explanations of order effects in human inference[C]//Proceedings of the Annual Meeting of the Cognitive Science Society. Seattle: Cognitive Science Society, 2010, 32(32): 1166-1171. [44] TRUEBLOOD J S, BUSEMEYER J R. A quantum probability account of order effects in inference[J]. Cognitive science, 2011, 35(8): 1518-1552. [45] BUSEMEYER J R, WANG Z, KHRENNIKOV A. Applying quantum principles to psychology[J]. Physica scripta, 2014 (T163): 014007. [46] AERTS D, GABORA L, SOZZO S. Concepts and their dynamics: a quantum-theoretic modeling of human thought[J]. Topics in cognitive science, 2013, 5(4): 737-772. [47] KHRENNIKOV A. Quantum-like model of cognitive decision making and information processing[J]. Biosystems, 2009, 95(3): 179-187. [48] ZHITOMIRSKY-GEFFET M, BAR-ILAN J, LEVENE M. Categorical relevance judgment[J]. Journal of the Association for Information Science and Technology, 2018, 69(9):

1084-1094. [49] GREISDORF H. Relevance: an interdisciplinary and information science perspective[J]. Informing science, 2000, 3(2): 67-72. [50] JANES J W. Relevance judgments and the incremental presentation of document representations[J]. Information processing & management, 1991, 27(6): 629-646. [52] WANG Z, BUSEMEYER J R. A quantum question order model supported by empirical tests of a priori and precise predictions[J]. Topics in cognitive science, 2013, 5(4): 689-710. [53] DZHAFAROV E N, ZHANG R, KUJALA J. Is there contextuality in behavioral and social systems?[J]. Philosophical transactions of the Royal Society A: mathematical, physical and engineering sciences, 2016, 374(2058): 20150099. [54] MOORE D W. Measuring new types of question-order effects: additive and subtractive[J]. The public opinion quarterly, 2002, 66(1): 80-91. [55] BELKIN N J. Anomalous states of knowledge as a basis for information retrieval[J]. The Canadian journal of information science, 1980, 5(1): 133-143. [56] VAKKARI P. Searching as learning: a systematization based on literature[J]. Journal of information science, 2016, 42(1): 7-18. [57] CAO Shujin, LI Jiena, WANG Zhihong. Research on fine-grained aggregation unit metadata for web information resource aggregation search[J]. Journal of Library Science in China, 2017, 43(4): 74-92. [58] WANG Yicheng, WANG Ping, WANG Meiyue. Research on hierarchical sensitive influencing factors of online information cocoons based on SVM[J]. Information and Documentation Services, 2019, 40(6): 90-97. [59] DZUNG H, LUO S. Search result diversification in resource selection for federated search[C]//International ACM SIGIR conference on research & development in information retrieval. New York: ACM, 2013: 612-622. [60] YU H, JATOWT A, BLANCO R, et al. Decoding multi-click search behavior based on marginal utility[J]. Information retrieval, 2017, 20(1): 25-52. [61] YEARSLEY J M, BUSEMEYER J R. Quantum cognition and decision theories: a tutorial[J]. Journal of mathematical psychology, 2016, 74: 99-116. [62] BUSEMEYER J R, WANG Z. What is quantum cognition, and how is it applied to psychology?[J]. Current directions in psychological science, 2015, 24(3): 163-169.

Author Contributions: Wang Zhihong designed the experimental protocol, conducted the experiment, and drafted the manuscript; Cao Shujin revised the experimental protocol and manuscript.

Note: Figure translations are in progress. See original paper for figures.