Fight Bias with Bias? Two Interventions for Mitigating the Selective Avoidance of Clicking Uncongenial Facts

Yuko Tanaka (tanaka.yuko@nitech.ac.jp)
Graduate School of Engineering, Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555 Japan

Hiromi Arai (hiromi.arai@riken.jp)
Center for Advanced Intelligence Project, RIKEN
1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027 Japan

Miwa Inuzuka (minuzuka@u-gakugei.ac.jp)
Department of Education, Tokyo Gakugei University
4-1-1 Nukui-kita-machi, Koganei, Tokyo, 184-8501 Japan

Yoichi Takahashi (yoichi.takahashi.c8@tohoku.ac.jp)
Graduate School of Information Sciences, Tohoku University
6-6-05 Aza-Aoba, Aramaki, Aoba-ku, Sendai, 980-8579 Japan

Minao Kukita (minao.kukita@i.nagoya-u.ac.jp)
Graduate School of Informatics, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601 Japan

Kentaro Inui (kentaro.inui.e2@tohoku.ac.jp)
Graduate School of Information Sciences, Tohoku University
6-6-05 Aza-Aoba, Aramaki, Aoba-ku, Sendai, 980-8579 Japan

Abstract

Selective avoidance of facts that are uncongenial to preexisting false beliefs is a biased click behavior that decreases the effect of correcting misinformation. This study examined the strength of this avoidance tendency and whether interventions could reduce it. In a preregistered experiment with 1,203 participants, we compared two different types of interventions: an intervention with instruction that directly calls for reflection via text (instruction intervention); an intervention with a ranking-biased order that induces people to click on what they easily see and vice versa (ranking-biased intervention). Contrary to our expectation, the results showed no significant effect of the instruction intervention. On the other hand, ranking-biased intervention showed preventive outcomes regarding participants' selective avoidance behaviors and promoted clicking on links to uncongenial facts. We discuss the limitation of a direct call for deliberation and the implications of the interaction between the interventions and click behavior based on cognitive characteristics.

Keywords: misinformation; confirmation bias; debunking; selective avoidance; click behavior; uncongenial facts

Introduction

Previous studies have demonstrated that correcting misinformation effectively decreases the perceived accuracy of misinformation (Lewandowsky et al., 2020; Porter & Wood, 2021; Walter & Murphy, 2018). However, people usually need to click on a link to read correcting stores in online environment. For example, fact-checking websites such as Snopes.com and politifact.com display questionable claims as a list on the landing page. Each claim is retracted with a simple "true/false" label. People access a full fact-checking story against misinformation via two phases: clicking on a link that indicates misinformation with a simple "false" label (click phase) and reading the refutation story (reading phase). Although fact-checking websites commonly use this simple retraction format, its effect in reducing false beliefs does not last longer than that of the refutation format (Ecker et al., 2020).

Accordingly, a key question regarding the utilization of online correction to mitigate false beliefs is how to encourage individuals to move from the clicking phase to the reading phase. The primary obstacle to this is a psychological bias, called "selective exposure." Selective exposure is the human tendency to seek like-minded information while neglecting information that is uncongenial to preexisting beliefs (Fischer et al., 2005; Frey, 1986). Information-seeking click behaviors related to misinformation and refutation can be categorized in online environment. For example, if individuals believe a misinformation item to be true, then a corresponding fact would be uncongenial. Further, clicking on the link to read a refutation story based on this fact would lead them to examine their false beliefs (belief-examining click). A

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

corresponding fact will be congenial if individuals do not believe a misinformation item. Clicking on the link to read this fact would lead confirm their belief (belief-confirming click). Individuals can also choose to avoid clicking on the links (avoidance).

A recent study demonstrated that participants selectively avoided clicking on links revealing uncongenial facts while clicking on links revealing congenial facts. Surprisingly, because of this selective avoidance behavior, 93% of the participants' false beliefs were left unexamined (Tanaka et al., 2023). Additionally, the selective avoidance of clicking on uncongenial facts was not unusual and was observed in 43% of the participants. These findings indicate that many participants hesitated during the click phase and did not attend the reading phase. Furthermore, the selective avoidance of belief-examining clicks is related to a cognitive characteristic, *reflexiveness*, measured using the Bullshit Receptivity Scale (BSR) (Pennycook et al., 2015)— in contrast to *reflectiveness*, Pennycook and Rand (2020) assumed that BSR measured the human tendency to accept claims reflexively. That is, the result indicated that participants with higher reflexiveness tended to avoid clicking on links that revealed uncongenial facts.

With the growing need to understand human factors and designing technological environments to support individuals' deliberative thinking (Dingler et al., 2021), it is essential to focus on individual differences to increase the minimal level of deliberative thinking. Due to the lack of research, it is not yet clear how individuals who tend to selectively avoid of clicking on links that reveal uncongenial facts can be supported. Therefore, this study investigated design interventions that mitigate the selective avoidance of belief-examining clicks and encourage individuals to move toward the reading phase. As the strength of this selective avoidance tendency is unclear, this study focused on two different interventions to examine this tendency.

The first intervention—the instruction intervention—resembles the "stop and think" instruction: this includes giving an instruction via text that encourages individuals to reflect on the possibility that some of their preexisting beliefs are false and to click the links revealing uncongenial facts subsequently. Calling for direct deliberation is a typical strategy governments and public institutions adopt to control the spread of misinformation. In ideal conditions, this straightforward approach would be effective; however, there is a possibility that the instruction intervention interacts with individual differences in reflexiveness related to click behavior and the perception of misinformation (Pennycook & Rand, 2020; Tanaka et al., 2023). Thus, instruction intervention may be ineffective for individuals with high reflexiveness. They may reflexively disregard instructions because they requires reflectiveness. This case emphasizes the need for a different intervention, effective even when individuals have reflexive tendencies.

The second intervention— the ranking-biased intervention—is intended to be effective for individuals with high- and low reflexiveness. For the ranking-biased intervention, this study focused on the ranking bias, which is one of the selection biases observed on the web and refers to a bias wherein individuals tend to click on what they easily see (e.g., the top-ranked result) but not to click on what they do not (Baeza-Yates, 2018). Ranking bias is assumed to be a strong human tendency repeatedly observed in previous studies (Draws et al., 2021; Epstein & Robertson, 2015; Epstein et al., 2017); thus, the ranking-biased intervention aims to be effective regardless of the individuals' level of reflexiveness. By placing links related to uncongenial facts on a web page that easily attracts individuals' attention and on which individuals subsequently click easily in the online context, would they be less likely to avoid belief-examining clicks, or would they keep selectively avoiding the clicks? It is essential to determine the controllability of the selective avoidance tendencies through interventions.

We conducted a preregistered experiment to examine whether the instruction and ranking-biased interventions enhance individuals' tendencies to click on links revealing facts uncongenial to their preexisting false beliefs. Specifically, we tested the following hypotheses:

H1: Both the instruction intervention (H1a) and the ranking-biased intervention (H1b) enhance individuals' tendencies to click links revealing uncongenial facts.

H2: Individuals' reflexiveness interacts with the two design interventions: specifically, high reflexiveness decreases the effect of the instruction intervention regarding individuals' clicking on links revealing uncongenial facts (H2a), whereas reflexiveness does not interact with the ranking-biased intervention (H2b).

Method

Participants

The sample size and data exclusion criteria were set before the study and preregistered (see AsPredicted #109330, https://aspredicted.org/zc3rj.pdf). A total of 1,203 participants (600 women, 602 men, and one other; M_{age} = 47.1, SD_{age} = 11.4), who were recruited from Cross Marketing, Inc., completed the entire online experiment.

Experimental Design

Corresponding to the preregistered design, the participants were randomly assigned into one of the following four conditions: instruction intervention (with/without) × ranking-biased intervention (with/without). In the instruction intervention, the participants were explicitly encouraged to reflect with the following instruction: "some links labeled as 'false' may include information that you believed to be 'true' last time. To mitigate the impact of misinformation, why don't you focus on the information that you believed to be 'true'?" Ranking-biased intervention is an intervention design that takes advantage of click behavior affected by the ranking bias and that places links revealing uncongenial facts where individuals tend to click, for example, the beginning and end of the list in this experimental environment. The Institutional Review Board of Nagoya Institute of Technology exempted our research protocol.

Materials

A preliminary survey identified 36 false messages, the corresponding 36 correction messages, and 24 filler messages. False and correction messages were obtained from fat-checking stories related to COVID-19 from third-party fact-checkers (mainich.jp, buzzfeed.com, and infact.press), whereas the filler messages were real news collected from a significant news website (nhk.co.jp). Each message was summarized into approximately 130 Japanese characters. The complete set of stimuli and other supplementary materials is available at the OSF (https://osf.io/82tzn/?view_only=a1ddc963e0734e1391561ad4980ada7f).

To measure the participants' reflexiveness, we employed the BSR (Pennycook et al., 2015), comprising 10 pseudo-profound bullshit sentences including random patches of abstract buzzwords. Bullshit sentences were presented randomly, mixed with 10 prototypically profound and 10 mundane sentences. Participants were asked to rate the profoundness of each sentence on a 5-point Likert scale (1 = *not at all profound*, 5 = *very profound*). The average of the participants' ratings of the bullshit sentences was used to index individual differences in reflexiveness.

Procedure

The participants completed the experiment over 2 days between October 17–21, 2022.

Day 1: Measuring Preexisting Belief and Click Behavior pre-Intervention

After providing informed consent and answering to demographic questions, participants were asked, "*How accurate do you think this information is?* (1= *Not at all*, 6= *Highly accurate*)" for each of the 60 messages. The messages, including two attention check questions, were displayed one at a time in a random order. Participants who believed that more than two of the 36 false messages were accurate were invited to participate in Day 2.

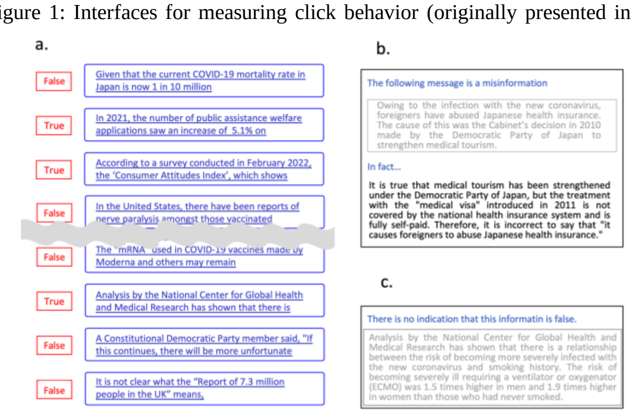
Subsequently, the pre-intervention click behaviors were measured. The participants were instructed: "It was revealed that the messages provided previously could be divided into correct information and misinformation as a result of fact-checking. An explanation of each correct and false item can be read by clicking on the link. Click on at least four links of your choice. Note that a button to proceed to the next page will appear after 3 min." They subsequently proceeded to the next page, which displayed a list of 30 links related to 18 false and 12 filler messages in random order (Figure 1a). The 18 false messages included the half of the false messages that a participant believed to be accurate, and the other half were presented on day 2. On the right side of the web page, a link displays the first 40 characters in each message. On the left, each link was labeled "false" for false messages or "true" for the filler messages.

Based on the click behaviors, the participants were divided into the fact-avoidance group or fact-exposure group using an index described in the following section. The participants of each group were randomly assigned into one of the four conditions: instruction intervention (with/without) × the ranking-biased intervention (with/without).

Day 2: Measuring Click Behavior post-Intervention and Reflexiveness

First, the participants were instructed: "The fact-check results presented on Day 1 were half of the total due to many fact-checks. On Day 2, we presented the other half of the fact-check results." The participants' second click behaviors were measured under four conditions: they were shown a different set of 30 links composed of 18 false and 12 filler messages. Except for the set of links and the four conditions, the procedures were identical to those on Day 1. The combination of the messages displayed on Days 1 and 2 was counterbalanced among the participants. Subsequently, the participants answered the 30 questions on the BSR. Finally, they were debriefed about the study's purpose and provided with a list of the false messages along with a warning.

Figure 1: Interfaces for measuring click behavior (originally presented in



Japanese). a) The main page displaying the list of 30 links. b) A fact-checking story page that was displayed after clicking on a corresponding link labeled "false," in which a false message "The following message is a misinformation" was displayed first, followed by a corresponding fact message "In fact…" (see Table 1 for the translations of false message and fact message). c) A page that was displayed after clicking on a corresponding link labeled "true," with the filler message "There is no indication that this information is false."

Measurement of Click Behavior

To measure click behavior related to uncongenial facts and divide participants into either the fact-avoidance or fact-exposure group, we used the Fact Avoidance/Exposure Index (FAEI) proposed by a previous study (Tanaka, et al., 2023). The FAEI is calculated as follows:

$$FAEI = x - EV \quad (1)$$

where *x* signifies the participant's targeted click behavior; this behavior is measured by the observed number of links clicked relating to uncongenial facts corresponding to preexisting false beliefs. *EV* indicates the expected value of the number of links related to uncongenial facts that can be clicked if any links are randomly clicked under a specific condition.

EV is calculated using the following formula:

$$EV = \sum_{i=0}^{k} \frac{C(a,i) \times C(n-a,b-i)}{C(n,b)} \times i \qquad (2)$$

where *n* is the total number of links presented, *a* is the number of false messages that the participant believes to be accurate, *b* is the total number of links clicked by the participant, and *k* is the smaller number of *a* or *b*. Here, *i* signifies the possible number of links that reveal uncongenial facts clicked when the participant clicks randomly any links *b* times. It extends from 0 to *k*. Here, *k* takes on a smaller value of *a* or *b* because the number of links revealing uncongenial facts clicked cannot be larger than the smaller value of *a* or *b*. FEAI is calculated concerning each participant, as the parameters *x*, *a*, *b*, and *EV* vary among the participants. The FAEI value becomes negative when a participant is less engaged in clicking links revealing uncongenial facts corresponding to preexisting false beliefs than expected (fact avoidance). Contrastingly, it becomes positive when a participant is more engaged in clicking links revealing uncongenial facts than expected (fact exposure). Noteworthy, the participants who received an FAEI score of 0 did not proceed to participate on Day 2.

Results

Each participant's FAEI score was calculated based on the click data collected on Day 1. A total of 617 participants with negative FAEI scores were categorized into the fact-avoidance group, and 586 participants with positive FAEI scores were categorized into the fact-exposure group.

3.1 Effects of Two Interventions

H1 stated that both the instruction- and ranking-biased interventions encourage participants to click links revealing uncongenial facts (*H1*). To test this hypothesis, we calculated the click ratio by dividing the number of clicked links by the number of links displayed. The click ratio was employed because the number of uncongenial and congenial fact-related links differed among the participants. Descriptive statistics regarding the click ratios of uncongenial facts, congenial facts, and fillers for days 1 and 2 are presented in Table 1.

Table 1: Means (standard deviations) for the click ratio of uncongenial facts, congenial facts, and fillers. The upper and lower values correspond to Day 1 (pre-intervention) and Day 2 (post-intervention), respectively.

	Fact-avoidance group (*n* = 617)				Fact-exposure group (*n* = 586)			
	Instruction (1), Ranking-bias (1)	Instruction (0), Ranking-bias (1)	Instruction (1), Ranking-bias (0)	Instruction (0), Ranking-bias (0)	Instruction (1), Ranking-bias (1)	Instruction (0), Ranking-bias (1)	Instruction (1), Ranking-bias (0)	Instruction (0), Ranking-bias (0)
n	145	157	162	153	139	152	155	140
Uncongenial facts	0.10 (0.13) 0.48 (0.30)	0.11 (0.16) 0.46 (0.32)	0.11 (0.15) 0.26 (0.28)	0.12 (0.16) 0.26 (0.28)	0.49 (0.21) 0.51 (0.30)	0.50 (0.24) 0.51 (0.30)	0.53 (0.24) 0.29 (0.26)	0.50 (0.24) 0.28 (0.23)
Congenial facts	0.29 (0.19) 0.21 (0.16)	0.31 (0.21) 0.19 (0.16)	0.27 (0.18) 0.22 (0.16)	0.31 (0.19) 0.24 (0.19)	0.26 (0.16) 0.19 (0.15)	0.29 (0.20) 0.21 (0.16)	0.27 (0.19) 0.28 (0.21)	0.29 (0.20) 0.30 (0.21)
Fillers	0.31 (0.19) 0.23 (0.20)	0.33 (0.21) 0.24 (0.20)	0.31 (0.18) 0.26 (0.19)	0.33 (0.19) 0.30 (0.20)	0.23 (0.17) 0.19 (0.17)	0.23 (0.16) 0.19 (0.16)	0.21 (0.16) 0.27 (0.20)	0.20 (0.19) 0.27 (0.19)

Table 2: Multiple regression coefficients of the BSR, the instruction intervention, and the ranking-biased intervention of the fact-avoidance group.

	Model (Fact-avoidance group, *n* = 616)				
	B	*SE*	*t*	*p*	95% CI
BSR: Instruction (0): Ranking-bias (0)	-0.065	0.059	-3.206	.001**	[-0.104, -0.025]
BSR: Instruction (1): Ranking-bias (0)	-0.059	0.020	-3.022	.003**	[-0.098, -0.021]
BSR: Instruction (0): Ranking-bias (1)	0.005	0.020	0.236	.814**	[-0.035, 0.044]
BSR: Instruction (1): Ranking-bias (1)	0.102	0.019	0.525	.600**	[-0.028, 0.048]
F	20.69				
R^2	0.114				

Note: BSR: bullshit receptivity; CI: confidence interval. ** $p < .01$

Subsequently, to examine the effects of the two intervention designs on the click ratio of uncongenial fact-related links, we conducted a mixed four-way analysis of variance (ANOVA) with the instruction intervention (with/without), ranking-biased intervention (with/without), FAEI group (fact avoidance/ face exposure), and time (Day 1: pre-intervention/Day 2: post-intervention) using the belief-examining click ratio as the dependent variable. The results revealed significant main effects of ranking-biased order, $F(1, 1194)=88.3$, $p<.001$, $\eta^2=0.042$, and FAEI group, $F(1, 1194)=381.5$, $p<.001$, $\eta^2=0.161$, and time, $F(1, 1194)=71.3$, $p<.001$, $\eta^2=0.023$. The main effect of the instruction was not significant. We also found a significant interaction effect between the ranking-biased order and time, $F(1, 1194)=173.5$, $p<.001$, $\eta^2=0.055$, and a significant interaction effect between the FAEI group and time, $F(1, 1194)=409.6$, $p<.001$, $\eta^2=0.12$.

For the interaction between ranking-biased order and time, post-hoc analysis showed significant simple main effects of time with the ranking-biased order, $F(1, 184)=137.0$, $p<.005$, $\eta^2=0.104$, and with the randomized order, $F(1, 1216)=5.38$, $p<.001$, $\eta^2=0.004$. When the link order was ranking-biased, the click ratio of uncongenial facts increased on Day 2 (M=0.49, SD=0.34) than on Day 1 (M=0.30, SD=0.29); Conversely, when the link order was randomized, the click ratio of uncongenial facts decreased on Day 2 (M=0.27, SD=0.26) than on Day 1 (M=0.31, SD=0.29).

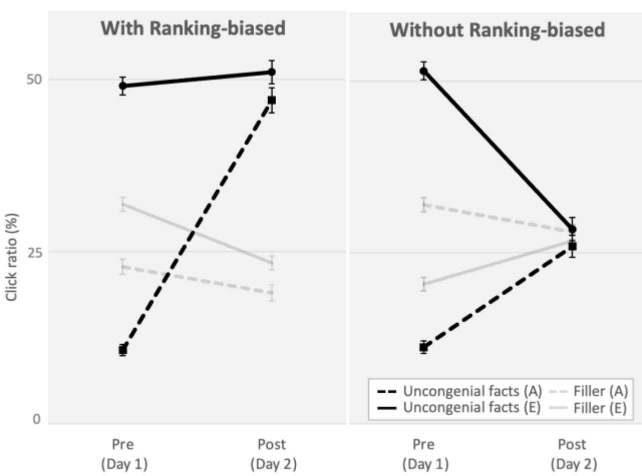


Figure 2: The click ratios of uncongenial facts for the fact-avoidance and fact-exposure groups with ranking-biased intervention (left panel) and without ranking-biased intervention (right panel). The click ratios of filler messages are shown in light grey color. Error bars represent standard error of the means. A and E in legend stand for the fact-avoidance and fact-exposure groups, respectively.

For the interaction between the FAEI group and time, post-hoc analyses showed significant simple main effects of time for the fact-avoidance group, $F(1, 1230)=328.0$, $p<.001$, $\eta^2=0.211$. For the fact-exposure group, $F(1, 1170)=46.8$, $p<.001$, $\eta^2=0.038$: the click ratio of uncongenial facts for the fact-avoidance group increased on Day 2 (M=0.36, SD=0.31) than on Day 1 (M=0.11, SD=0.15); However, the click ratio of uncongenial facts for the fact-exposure group increased in Day 2 (M=0.40, SD=0.30) than on Day 1 (M=0.50, SD=0.23).

In addition to preregistered analyses, we conducted a post-hoc pairwise comparison test with the Bonferroni adjustment as complementary analyses. The results revealed that when the link order was ranking-biased on Day 2, the click ratio of uncongenial facts for the fact-avoidance group increased on Day 2 (M=0.47, SD=0.31) compared to Day 1 (M=0.11, SD=0.15). Contrastingly there were no differences in the click ratio of uncongenial facts for the fact-exposure group on Day 2 (M=0.51, SD=0.30) or on Day 1 (M=0.49, SD=0.22) (Figure 2, left panel). When the link order was randomized on Day 2, the click ratio of uncongenial facts for the fact-avoidance group increased on Day 2 (M=0.26, SD=0.28) compared to Day 1 (M=0.11, SD=0.16). In contrast, the click ratio of uncongenial facts for the fact-exposure group decreased on Day 2 (M=0.28, SD=0.25) compared to Day 1 (M=0.52, SD=0.24) (Figure 2, right panel). Figure 2 also shows the click ratios of the filler messages.

3.2 Relationship between the two interventions and Bullshit Receptivity Scale

H2 focused on the individual differences in the effect of the instruction and ranking-biased interventions. We predicted that participants' reflexiveness would interact with the two design interventions. The BSR was used to measure the participants' reflexiveness. Higher BSR scores indicated higher reflexiveness. Cronbach's alpha for the BSR was reliable (α=.82).

To examine this hypothesis, we conducted a multiple regression analysis to predict the difference in the click ratio of uncongenial facts between days 1 and 2 for the fact-avoidance group, with the positive values indicating that the click ratio of uncongenial facts increased post-intervention. BSR, instruction intervention (with/without), and ranking-biased intervention (with/without) were entered into the model as predictors. Table 2 presents the models' coefficient estimates, standard errors, and *t*-values. The regression analysis results revealed that the BSR interacted with the two interventions, supporting *H2*. However, the interaction patterns differed partially from our hypothesizes. For the instructional intervention, we predicted that high reflexiveness would decreases the effect of this intervention on the participants' clicking on links, revealing uncongenial facts (*H2a*).

Contrary to this hypothesis, the effect of the BSR did not differ with or without the instructional intervention. However, we hypothesized that reflexiveness would not interact with the ranking-biased intervention (*H2b*). Supporting *H2b*, BSR was not significantly related to the difference in the click ratios when the ranking-biased intervention was implemented. The BSR significantly predicted decreased click ratio of uncongenial facts from Day 1 to Day 2 only when the ranking-biased intervention was not provided.

Discussion

In focusing on the selective avoidance tendency of individuals regarding click links revealing uncongenial facts, this study investigated two intervention designs that mitigate this tendency and encourage individuals to click on links revealing uncongenial facts. To understand the strength of this selective avoidance tendency, we compared two different types of interventions: an instruction-based intervention and ranking-biased intervention.

Contrary to our hypothesis (*H1a*), the instruction intervention did not enhance participants' clicking on links that revealed uncongenial facts. Even though it is a commonly used strategy to directly call for reflective thinking, the result implies that the selective avoidance tendency is stronger than the effect of such intervention.

On the other hand, the ranking-biased intervention, which placed the targeted links where people tend to click on a web page, enhanced participants' clicking on uncongenial facts as compared to the randomized order, supporting *H1b*. Complementary analyses revealed that the fact-avoidance group, which selectively avoided clicking on uncongenial facts and retained 89% of their false beliefs intact, clicked on 47% of the links that revealed uncongenial facts post-intervention (Figure 2, left panel). This proportion was at the same level as that for the fact-exposure group (52%). While the ranking-biased intervention neither increased nor decreased participants' clicking on links that revealed uncongenial facts for the fact-exposure group, it would be appropriate to interpret this result as indicating that the ranking-biased intervention allowed the click ratio to remain high at approximately 50%. Compared with the click ratios of filler links of approximately 25%, which were similar to the general click ratios in previous studies (Draws et al., 2021; Rieger et al., 2021), the results suggest that click ratios of approximately 50% are sufficiently high and that a ceiling effect might have occurred at this level.

This interpretation is also consistent with the result that the participants' clicking on links revealing uncongenial facts decreased on Day 2 without ranking-biased intervention. This could be due to the general tendency of click behavior to be active on the first page but to become inactive on the second page (Epstein et al., 2017): the participants actively engaged in click behavior on Day 1; thus, we found noticeable differences in the participants' clicking on links that revealed uncongenial facts between the fact-avoidance and fact-exposure groups. However, the complementary analyses showed that the difference between the two groups became insignificant on day 2, and their click behaviors regarding accessing uncongenial facts became closer to that of fillers (Figure 2). Conversely, this suggests the effect of the ranking-biased intervention in promoting participants' clicking on uncongenial facts for the fact-avoidance group and maintaining the click ratio at a high level for the fact-exposure group. This would revert to the ordinary level on Day 2 without the ranking-bias intervention.

We also predicted that the participants' reflexiveness would interact with the two intervention designs (*H2*). Contrary to *H2a*, high reflexiveness was significantly related to a low click ratio for uncongenial facts, regardless of the instruction intervention (Table 2). This is attributed to the lack of a significant effect of the instruction. The relationship between high reflexiveness and low click behavior for uncongenial facts is consistent with the results of a previous study (Tanaka, et al., 2023). On the other hand, the results supported *H2b*: with the ranking-biased intervention, reflexiveness did not predict clicking on links that revealed uncongenial facts. This suggests that the ranking-biased intervention' effect in fighting participants' selective avoidance regarding clicking on links revealing uncongenial facts was stronger than reflexiveness. Even if participants have the cognitive characteristic of high reflexiveness, which tends to induce the selective avoidance of clicking on links that reveal uncongenial facts, this can be mitigated with an intervention that display the links where people easily see and tend to click.

Limitations and Future Directions

One limitation of this study is that we focused on the click phase alone. Further engagement is needed, including reading stories, reflecting on false beliefs, and updating the beliefs to update false beliefs based on the reading corrections. The ranking-biased intervention was effective in inducing the participants' click behavior toward uncongenial facts. However, it does not mean it was effective in enhancing reflective thinking. Instead, ranking-biased interventions utilize participants' reflexiveness to facilitate their movement from the click phase to the reading phase on fact-checking websites. Therefore, whether this type of intervention, which display targeted links where people tend to click, contributes to the effectiveness of the reading phase remains unclear.

Similarly, while this study did not support the effect of instruction intervention on enhancing participants' click behavior, it cannot be stated that giving instruction that calls for reflection is entirely ineffective because it may enhance fact-checking engagement in the reading phase. Further research is needed to examine the effects of different interventions on different phases of human behavior regarding exposure to correction in online environment.

Concluding Remarks

The selective avoidance of uncongenial facts was so strong that it could not be mitigated by directly calling for deliberation via plain text. However, it was improved by taking advantage of another click tendency regarding ranking bias. The latter intervention is useful for individuals with high reflexiveness as well as individuals with low reflexiveness. These findings will contribute to explaining the cognitive characteristic behind click behavior and considering interventions that are effective for individuals who are less likely to benefit from correcting misinformation.

Acknowledgments

This work was supported by JST CREST Grant Number JPMJCR20D2 and JSPS KAKENHI Grant Number 21K12605, Japan.

References

Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM, 61*(6), 54–61. https://doi.org/10.1145/3209581

Dingler, T., Tag, B., Lorenz-Spreen, P., Vargo, A. W., Knight, S., & Lewandowsky, S. (2021). Workshop on Technologies to Support Critical Thinking in an Age of Misinformation. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI EA'21), Article No. 105, 1–5. https://doi.org/10.1145/3411763.3441350

Draws, T., Tintarev, N., Gadiraju, U., Bozzon, A., & Timmermans, B. (2021). This Is Not What We Ordered: Exploring Why Biased Search Result Rankings Affect User Attitudes on Debated Topics. *SIGIR 2021 - Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 295–305. https://doi.org/10.1145/3404835.3462851

Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology*, *111*(1), 36–54. https://doi.org/10.1111/bjop.12383

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(33), E4512–E4521. https://doi.org/10.1073/pnas.1419828112

Epstein, R., Robertson, R. E., Lazer, D., & Wilson, C. (2017). Suppressing the Search Engine Manipulation Effect (SEME). *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW). https://doi.org/10.1145/3134677

Fischer, P., Jonas, E., Frey, D., & Schulz-Hardt, S. (2005). Selective exposure to information: The impact of information limits. *European Journal of Social Psychology*, *35*(4), 469–492. https://doi.org/10.1002/ejsp.264

Frey, D. (1986). Recent Research on Selective Exposure to Information. *Advances in Experimental Social Psychology*, *19*(C), 41–80. https://doi.org/10.1016/S0065-2601(08)60212-9

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., … Zaragoza, M. S. (2020). *The debunking handbook 2020*. https://sks.to/db2020

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, *10*(6), 549–563. https://doi.org/10.5281/zenodo.1067051

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185–200. https://doi.org/10.1111/jopy.12476

Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(37), 1–7. https://doi.org/10.1073/pnas.2104235118

Rieger, A., Draws, T., Theune, M., & Tintarev, N. (2021). This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias. In *HT 2021 - Proceedings of the 32nd ACM Conference on Hypertext and Social Media* (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3465336.3475101

Tanaka, Y., Inuzuka, M., Arai, H., Takahashi, Y., Kukita, M., & Inui, K. (2023). Who does not benefit from fact-checking websites? A psychological characteristic predicts the selective avoidance of clicking uncongenial facts. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI'23), Article No.664, 1-17. https://doi.org/10.1145/3544548.3580826

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, *85*(3), 423–441. https://doi.org/10.1080/03637751.2018.1467564