

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The Effects of Refutative Elements in Others' Comments on Accepting Health-Related Fallacious Claims

Permalink

<https://escholarship.org/uc/item/4jc93021>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors

Inuzuka, Miwa

Tanaka, Yuko

Fujimoto, Kazunori

Publication Date

2023

Peer reviewed

The Effects of Refutative Elements in Others' Comments on Accepting Health-Related Fallacious Claims

Miwa Inuzuka (minuzuka@u-gakugei.ac.jp)

Department of Education, Tokyo Gakugei University, 4-1-1 Nukuikita-machi
Koganei, Tokyo 184-8501, Japan

Yuko Tanaka (tanaka.yuko@nitec.ac.jp)

Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku
Nagoya, Aichi 466-8555, Japan

Kazunori Fujimoto (kfujimoto@kindai.ac.jp)

Faculty of Business Administration, Kindai University, 3-4-1 Kowakae
Higashi-Osaka, Osaka 577-8502, Japan

Abstract

This study investigated the impact of refutative elements in others' comments on accepting fallacious claims about food nutrition. Four types of comments were used, two of which included refutative elements (challenging evidence and deductive process). A multiple regression analysis was conducted with 506 participants' agreement with the fallacious claim as the dependent variable and the type of comment and their agreement before exposure to the comments as independent variables. The study also considered individual differences such as media literacy, information literacy, cognitive reflection, and interest in and familiarity with the topic. The results showed that presenting comments that challenged the deductive process significantly decreased agreement when participants had higher agreement before exposure. Further analysis of participants with high initial agreement revealed significant effects of all three comment types. A supplemental survey ($n=182$) suggested that the perceived negativity and usefulness of the comments influenced the participants' agreement.

Keywords: health-related fallacious claim; refutation; others' comments; knowledge update.

Introduction

The present study focuses on fallacious claims about food nutrition, which can be defined as a type of health-related misinformation. Online health-related information may serve to form and reinforce fallacious beliefs (Albarracín et al., 2018; Chua & Banerjee, 2017, 2018). Fallacious food information can be defined as the claim of benefit or harm of certain foods not supported by scientific evidence and expert opinion (e.g., eating grapefruit with every meal helps burn fat leading to weight loss). Typical fallacious claims are based on faulty evidence or inappropriate interpretation of evidence to explain erroneous knowledge. Such claims are often overlooked because they generally seem acceptable and are less harmful than medical misinformation. However, there is a continuum between accepting such seemingly harmless fallacious beliefs and fatal misinformation.

Non-professional readers could struggle with denying fallacious claims, because validating such information requires expertise (Bautista, Zhang, & Gwizdka, 2021).

Misinformation is constantly emerging, making it challenging to provide reliable guidance. However, comments posted by other readers online can serve as a resource to refute false claims. User comments can help update the interpretation and evaluation of online text. If these comments highlight inaccuracies or encourage critical thinking, readers may question the text. This study examines how other people's comments influence readers' agreement with false claims and how knowledge is updated. Through this study, we aim to identify the characteristics of online text that is essential to debunking fallacious beliefs.

Users' Comments and Their Impact

Comments provided to web articles are usually short sentences expressing impressions or evaluations; they influence evaluations and attitudes toward the topic. Cumulative studies have indicated that the negative valence of comments leads to a decrease in the evaluation of information (Go, Jung, & Wu, 2014; Kümpel & Unkel, 2020; Waddell & Bailey, 2019; Waddell, 2020; Weber, Prochazka, & Schweiger, 2019). The stronger effects of negativity compared to the positivity (Baumeister et al., 2001) were observed in user comments (Cameron & Geidner, 2014; Inuzuka, Tanaka, & Tsubakimoto, 2019; Waddell, 2020; Waddell & Bailey, 2019; Waddell & Sundar, 2017).

The explanation regarding negative comments' effects suggests that the comments could evoke the bandwagon effect, which is a tendency of people to conform to attitudes or actions based on the perception of widespread peer adoption (Sundar, 2008) or facilitate elaboration processes. Regarding the bandwagon effect explanation, people tend to regard a small number of comments as representing the majority opinion (Zerback & Fawzi, 2017) and follow them heuristically (Sundar, 2008; Sundar, Jia, Waddell, & Huang, 2015; Waddell & Sundar, 2020). The other rationale suggests negativity, implies that the message is diagnostic, and attracts readers' focus (Skowronski & Carlston, 1989).

While the effects of comments' negativity are widely acknowledged, few studies have investigated the quality of arguments. Substantive information included in the

comments is suggested to be a facilitative element for elaboration that ends in attitude change (Inuzuka, Tanaka, & Fujimoto, 2023; Kim & Gonzalez, 2022). Thus, the present study aims to further clarify the quality of substantive information that prevents the acceptance of fallacious claims.

Refutation of Fallacious Claims

Appropriately rejecting inaccurate information can be challenging; Rapp and Salovich (2018) pointed out that exposure to inaccurate information, even if clearly incorrect, can confuse readers and make it difficult to integrate it with pre-existing knowledge. People may even begin to doubt correct knowledge and use incorrect information (Hinze et al., 2014; Marsh & Fazio, 2006; Marsh et al., 2003).

Providing refutation text is a promising solution to update knowledge and remove erroneous parts. Refutation texts include three components: identifying misinformation, rejecting misinformation, and explaining correct information (Hynd, 2001). The presentation of refutation texts has been shown to be effective for knowledge revision in various areas, including healthcare (Trevors & Kendeou, 2020). According to Kendeou and O'Brien (2014a), correct and incorrect information must be integrated to explain why misinformation is erroneous. Thus, refutation is effective for knowledge revision when it provides causal explanations of the negation (Kendeou, Smith, & O'Brien, 2013; Kendeou, Walsh, Smith, & O'Brien, 2014b; Kendeou, Butterfuss, Kim, & Van Boekel, 2019).

We concentrated on evidence and warrant critiques as possible types of negation explanations. Considering that the link between claims and premises is important in argumentation (Freeman, 2008; Nussbaum, 2011), the element that denies the evidence or breaks the link would work effectively as a refutation.

To investigate the effects of refutative elements, we included in the others comments, we investigate the following four categories of comments. First, two categories of comments contain elements that could function as explanations: challenges to the evidence (C-Evidence) and challenges to the reasoning process (C-Process). C-Evidence comments point out that there are problems with the facts on which the fallacious claim is based. In a health-related article, for example, comments indicating that a medical theory or experiment introduced is contradictory or inappropriate fall into this category. C-Process is a category of comments that question the warrant. The C-Process comments provide remarks on the excessive generalizations or leaps in deriving claims from the facts presented as evidence. The C-Process comments raise concerns regarding the link between evidence and claim, implying that the claim can be rejected even if an evidential fact is accepted. As these comments explain why the final claim is incorrect and provide information about the premise and its relationship to the claim, presenting a comment in one of these categories is more likely to facilitate the updating of one's knowledge. A comment that is as critical as these two types of comments but offers no explanation is included in the category, the

challenge to the conclusion (C-Conclusion). Comments that do not explain why and reject the conclusion as non-essential or sense-making fall into this category. Since C-Conclusion comments convey negative comments, they would have the effect of decreasing agreement with the false claim. As they lack an explanation for incorrect knowledge in an integrated way, the effects of C-Conclusion comments can be weaker than those of C-Evidence and C-Process. The fourth category is associative comments that neither challenge nor explain fallacious claims (No-Challenge).

Thus, the present study examines the following hypotheses:

H1-a: Challenging comments that provide the refutative element, the explanations on the fallaciousness (C-Evidence and C-process), decrease readers' agreement with fallacious claims.

H1-b: The comments that include refutative elements are more effective in decreasing readers' agreement than other comments (C-Conclusion and No-Challenge).

H2: Critical comments without a refutative element (C-Conclusion) decrease the readers' agreement with the fallacious claim compared to comments with no challenge (No-Challenge).

Individual Differences

Individual differences exist in the propensity to believe misinformation. Although the relevance of these individual differences is not the central issue of the present study, individual differences that increase the tendency to become critical about the claim may mask the effects of others' comments. Thus, we include their index as control variables. Firstly, we include the index of media literacy and information literacy (Jones-Jang et al., 2019). The literature on media literacy emphasizes people's perceived beliefs about their ability to access, analyze, evaluate, and communicate information (Aufderheide, 1993). Media literacy is generally expected to be related to one's ability to critically evaluate information. In the present study, we measure media literacy in accordance with this definition. Information literacy is defined as the ability to understand, find, evaluate, and use information (ACRL, 2000). Studies suggested that the information literacy scale predicted the accurate identification of fake news (Jones-Jang et al., 2019; Jones-Jang, Mortensen, & Liu, 2021)

Another individual difference relates to cognitive style. Each individual differs in their tendency toward reflectiveness. Frederick (2005) developed a measurement for reflectiveness called the cognitive reflection test (CRT), which presents the correlation between the CRT score and other tests of analytic thinking (Toplak, West, & Stanovich, 2011). Recent studies have revealed that CRT score explains the ability to distinguish false news from accurate news (Pennycook & Rand, 2019).

Method

Open Practices and Ethical Statements

The materials used in the study and the data included in the analysis are available at OSF (<https://osf.io/jzgdxf/>). The research protocol was deemed exempt by the Institutional Review Board (IRB) of Tokyo Gakugei University (approval number 156-2). The participants consented freely to participate in the experiment, and no personal information was obtained.

Participants

Participants were recruited from Cross Marketing Inc. Panels. The participants received a monetary honorarium determined by the panel's criteria for participating. The terms and conditions for participation were explained at the beginning of the survey page, and the participants' consent was confirmed by clicking the "I Agree" button. A total of 916 participants completed all survey items.

Materials

Fallacious Article on Food Nutrition

An article regarding the health benefits of *Natto-kinase* was created and presented in the appearance of a website providing health information (Figure 1). *Natto-kinase* is an enzyme extracted from Natto, a traditional Japanese colored product. *Natto-kinase* exhibits strong fibrinolytic activity when in contact with human blood and blood clots. As it is derived from a traditional Japanese food that generally has a healthy reputation, there are high expectations for its various health effects other than its fibrinolytic activity. In this study, we prepared an article claiming that *Natto-kinase* improves the human immune system (immunity) and contributes to overall health. The claim was based on an experiment suggesting that mice with influenza recovered after being fed *Natto-kinase*. The article should be treated as fallacious since it contained following drawbacks: the details of the experiments presented are unclear and it is impossible to judge whether the data are valid, and the data on the mice experiment is directly extended to human health in general.

Attention-Checking Quiz

Participants answered an attention-checking quiz to ensure that they read the text properly. The quiz was in the form of selecting all correct statements regarding the content of the text from the seven alternatives.

Agreement with the Claims

Seven items (e.g., "*Natto* makes the body less susceptible to illness") related to the text's claims were presented, and participants were asked to respond on a 5-point Likert scale (1: strongly disagree–5: strongly agree). The same items were used to measure the strength of the reader's agreement before (Agreement-Before) and after (Agreement-After) reading the comments. The items consisted of the following three areas: the mouse experiment on which the text was based, the

efficacy of *Natto-kinase* in humans, and the importance of immunity.

Comments on the Article

Four types of comments were created (Table 1) and displayed at the top of the fallacious article on food nutrition (Figure 1). Two types of comments were crafted to explain the fallacy: C-Evidence and C-Process. C-Evidence presented questions or critical remarks on the validity of the experiment in which the claim of the text was used as evidence. The C-Process left the validity of the experiment untouched, but questioned the appropriateness of the conclusion deduced from it. Another critical comment was C-Conclusion, which questioned the conclusion but presented no remark on the data or deduction process. Control conditions were established, with comments mentioning associations concerning Natto (No-Challenge).



Figure 1: Screenshot of article and comments Article without the comments (left panel) and comments displayed above the article in the callout (right panel).

Table 1: Examples of comments

Condition	Examples
C-Evidence	If you don't know the details of the experiment, you can't trust it. Is it really that effective?
C-Process	Do mouse experiments work for humans? A mouse body is not the same as a human body.
C-Conclusion	If you eat Natto and become healthy, society will not need doctors.
No-Challenge	I have never had the flu before—is this rare?

Indexes for Individual Differences

Cognitive Reflection Test

Three items originally from Frederick (2005) were translated and used (e.g., “A bat and a ball cost \$1.10. The bat costs \$1 more than the ball. How much does the ball cost?”). We changed the terms and monetary units (bats and balls were changed to candy and gum, and units were changed from dollars to the currency of the country where the experiment was conducted) so that the numerical values fit the participants’ daily experiences. The participants were asked to write their answers in answer boxes.

Media Literacy Scale

For the media literacy scale (MLS), we selected questionnaire items from the media literacy literature (Joe-Jang et al., 2019; Ye, Toshimori, & Horita, 2016; Simons, Meeus, & T’Sas, 2017), and similar items were merged to create 18 items (e.g., “I know how media production and distribution works”). The questionnaire items consisted of descriptions of the handling of information equipment, understanding of media, and usual behavior, to which the participants responded on a 5-point scale (1: Not at all applied to me at all–5: Very much applied to me).

Information Literacy Questionnaire

We employed Seven items of information literacy questionnaire (ILQ). Five of the items were used by Jones-Jang et al. (2019) and two items related to food safety were included from Boh Podgornik et al. (2016). ILQ required the participants to respond to each question (e.g., “Which of the data listed below are ‘raw’ unprocessed data?”) by selecting one of four options (e.g., “share prices published at the end of a trading day,” “weather maps,” “population growth data presented in tables,” and “population growth data presented diagrammatically”).

Prior Knowledge and Attitude

The participants were required to self-assess their interest in and knowledge of health information using one item each. They were also asked to rate their knowledge of the topic (*Natto-kinase*) and whether they thought it would be effective with one item for each. All four items were rated on a 5-point scale (1: strongly disagree–5: strongly agree).

Procedures

The terms and conditions of the experiment were presented, and participants’ consent was confirmed. The participants were first asked to provide basic information, such as their age, and then to answer four questions on prior knowledge and attitudes. They were then presented with the fallacious article on food nutrition and asked to answer an attention-checking quiz after reading it. They then answered seven items regarding agreement with the claims of the article to measure their agreement before reading others’ comments (Agreement-Before). The participants were divided into four groups and presented with different types of comments in each group (C-Evidence, C-Process, C-Conclusion, or No-

Challenge). Comments were presented by randomly selecting two of the five comments in the comment type. The comments were presented above the article so that the subject could compare them with the article. After reading the comments, the subjects answered the scales again for agreement with the claim (Agreement-After). The participants then completed the MLS, ILQ, and CRT. Participants who had completed all questions disclosed that the text presented was inaccurate.

Results

Data Selection

We excluded the data of participants who failed to answer more than three items of the attention check quizzes and those who failed to follow the instructions on the two items in which they were asked to respond as instructed. Finally, data from 506 participants were included in the analysis.

Scoring

The MLS was categorized into three subscales based on exploratory factor analysis and previous studies: communication (ML-C: e.g., “I would caution people around me about the negative sides and negative effects of media”), knowledge (ML-K; e.g., “I know how media production and distribution works”), and skills (ML-S: e.g., “I can consciously choose between different media devices, based on their function”). We calculated the mean of the questionnaire items for each subscale and used them as scale scores for subsequent analysis (Table 2). Cronbach’s alpha coefficient was calculated for each questionnaire item scale. For the ILQ and CRT, correct answers were scored as 1 point and incorrect answers as 0 points, and the total score was used for the analysis.

Table 2: Mean scores (*SD*) for scales

Scales	Mean	<i>SD</i>
Agreement Before ($\alpha = .87$)	3.61	0.63
Agreement After ($\alpha = .89$)	3.61	0.65
ILQ	3.60	1.51
MLS-C ($\alpha = .72$)	2.50	0.79
MLS-K ($\alpha = .87$)	3.22	0.61
MLS-S ($\alpha = .87$)	3.00	0.81
CRT	1.24	1.05

Effects of Comments on the Change in Agreement

Multiple regression analysis (stepwise) was conducted using R (ver. 4.2.2). Agreement-Post was included as the dependent variable, and Agreement-Before, comment condition, three MLS subscale scores (MLS-C, MLS-K, MLS-S), ILQ, and CRT scores as independent variables. The comment condition was coded for the presence of a challenge in each of C-Evidence, C-Process, and C-Conclusion using three

dummy variables. The multiple regression analysis model included interactions between the Agreement-Before score and comment conditions.

The results of the final model (Table 3) showed that the estimate for partial coefficient of C-Conclusion was significant, indicating that the presence of comments decreased the agreement with the fallacious claim. The interaction between C-Process and Agreement-Before was significant. The simple slope estimate for C-Process was significant when Agreement-Before was high ($b = -0.16, SE = 0.05, p = .00$). The simple slope estimate was also significant in the opposite direction when the Agreement-Before scores were low ($b = 0.15, SE = 0.06, p = .01$). Thus, the results indicate that agreement decreases when a comment challenging the process is presented when the agreement is high before reading the comment.

The estimates for the other comment conditions did not reach a significant level. Among the individual difference indexes, the estimate for MLS-K was significant, indicating that people with higher self-reported scores in media literacy knowledge were more likely to agree with the fallacious claim. On the contrary, the ILQ score showed a small but significant negative estimate, meaning that appropriate knowledge of information lessened agreement with the text.

Table 3: Results of the final model of multiple-regression analysis: Estimates of partial coefficients,

Predictor Variables	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	3.23	0.11	28.22	<.001
Agreement-Before	0.90	0.04	24.14	<.001
C-Evidence ^{a)}	-0.07	0.04	-1.58	.114
C-Process ^{a)}	-0.06	0.05	-1.26	.210
C-Conclusion ^{a)}	-0.09	0.04	-2.00	.046
MLS-K	0.14	0.03	5.09	<.001
ILQ	-0.03	0.01	-2.51	.013
Familiarity	-0.03	0.02	-1.54	.123
Attitude	0.05	0.03	1.92	.055
Agreement-Before × C-Process	-0.28	0.06	-4.51	<.001
Agreement-Before × C-Conclusion	-0.12	0.07	-1.77	.078

$R^2 = .70, R^2_{adj.} = .70$

a) Dummy code for comment conditions.

Post-Hoc Analysis

As shown in the results of the multiple regression, the impact of comments may vary depending on the degree of agreement prior to being exposed to the comments. We examined this possibility using the following post-hoc analysis. We conducted multiple regression analyses using the same variables and model on two groups of participants: those who

scored four or higher on Agreement-Before ($n=152$) and those who scored three or lower ($n=105$).

The results of the analysis for the group with higher Agreement-Before scores showed that the effects of all challenge comment dummy variables were significant. The effect of individual differences remained in the final model, although not statistically significant. Familiarity with *Natto-kinase* remained significant, showing a negative impact on the agreement (Table 4).

Meanwhile, the results of the multiple regression analysis on the group with lower Agreement-Before scores revealed no significant effects of challenge comments. The two subscales' scores of individual differences in media literacy were significant, indicating that higher scores for communication and knowledge aspects of media literacy increased agreement (Table 5).

Table 4: The final model of multiple-regression analysis on high Agreement-Before group

Predictor Variables	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	0.43	0.34	1.29	.200
Agreement-Before	0.92	0.08	11.45	<.001
C-Evidence ^{a)}	-0.15	0.07	-2.02	.045
C-Process ^{a)}	-0.20	0.07	-2.82	.005
C-Conclusion ^{a)}	-0.19	0.07	-2.59	.010
MLS-C	-0.05	0.03	-1.51	.134
MLS-K	0.09	0.05	1.82	.071
Familiarity	-0.07	0.02	-2.84	.005

$R^2 = .53, R^2_{adj.} = .50$

a) Dummy code for comment conditions.

Table 5: The final model of multiple-regression analysis on low Agreement-Before group

Predictor Variables	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	0.58	0.37	1.58	.118
Agreement-before	0.43	0.10	4.16	<.001
MLS-C	0.22	0.07	3.32	.001
MLS-K	0.16	0.07	2.19	.031

$R^2 = .32, R^2_{adj.} = .30$

Supplemental Survey

Contrary to H1-a, we found no difference in the impact among the challenging comments. To further investigate the reason for this result, we conducted a supplemental survey to investigate how readers perceived differences in comments. We focused on the negativity and usefulness of evaluating fallacious claims.

Participants ($n=182$) who were not part of this study were asked to rate the negativity and usefulness of each comment on a 10-point scale (Table 6).

The results of a one-factor within-subjects analysis of variance were significant ($F(3,543) = 3.14, p = .00$ for negativity and $F(3,543) = 3.25, p = .00$ for usefulness), and multiple comparisons (Holm) showed that C-Conclusion rated significantly more negatively than the other three types of conditions. C-Evidence and C-Process differed significantly from No-Challenge. In terms of usefulness ratings, C-Evidence and C-Process were rated highest. The difference between the C-Conclusion and No-Challenge groups was significant. The differences between the No-Challenge and other three comment types were also significant. The results suggest that comments are differentiated by the presence of challenges and refutative elements. The effect sizes of the difference between C-Evidence and C-Conclusion were greater in usefulness ($d = 0.81$) than in negativity ($d = 0.46$). The tendency between C-Process and C-Conclusion is the same ($d = 0.84$ and $d = 0.60$, respectively).

Table 6: Mean scores for the negativity and usefulness rating

Comment condition	Negativity (<i>SD</i>)	Usefulness (<i>SD</i>)
C-Evidence	6.37 (1.49) ^{a)}	5.32 (1.76) ^{a)}
C-Process	6.22 (1.26) ^{a)}	5.32 (1.67) ^{a)}
C-Conclusion	7.06 (1.53) ^{b)}	3.95 (1.60) ^{b)}
No-Challenge	2.99 (1.64) ^{c)}	2.41 (1.52) ^{c)}

^{a) b) c)} Significant differences were found between the scores of each comment condition with different alphabets.

Discussion

Effects of Refutative Elements in Comments

The results for the prediction of hypotheses were mixed and partially supported: H1 was not supported, and H2 was supported in an unexpected way. Contrary to H1-b, only C-Conclusion showed a significant impact, and the other challenge comments (C-Evidence and C-Process) did not. These results can be explained by the reaction to negativity. The supplemental survey revealed that the differences among the comment types were perceived as expected. All three types of challenge comments were perceived as negative, as expected, but C-Conclusion was perceived as more negative than the other two types of comments. The stronger impact of C-Conclusion was probably due to readers perceiving negativity more saliently, which led to stronger heuristic processing, or the bandwagon effect (Sundar, 2008; Sundar et al., 2015).

Meanwhile, the interaction between C-Process and Agreement-Before partially supported H1-a, suggesting that the refutative elements in comments would impact those with

a greater need to update their knowledge. For participants with a high level of agreement, comments served as an opportunity to revise the certainty of their representations. However, for those who had already disagreed, the comments served only to confirm their understanding. Thus, the effects of comments were different between the two groups.

The results of the usefulness rating in the supplemental survey reinforce this discussion. The impact of the comments in this study result from perceptions of negativity and usefulness. According to the results of the supplemental survey, negativity stems from comments that are directly critical of the claim itself, whereas usefulness stems from the refutation element. The effect of comments with refutation elements was not observed in the analysis of all participants, since participants with less need for knowledge updates were included. The useful information provided by the refutative elements impacts readers when they need to revise their understanding.

Individual Differences and Accepting Fallacious Claims

The results showed unexpected effects of media literacy score. Higher scores on the knowledge dimension of the media literacy scale predicted higher agreement: the more people regard themselves as knowledgeable about media, the more they tend to accept fallacious claims. We speculate that this is because MLS is based on self-assessment. Jones-Jang et al. (2019) also demonstrated that media literacy questionnaire scores failed to predict the identification of fake news. Although self-reporting is a commonly used method, knowledge about media may be an inaccurate indicator due to ambiguous assessment criteria and insufficient metacognition. The present study suggests that overestimating one's knowledge is related to the tendency to agree with fallacious claims.

Significance of the Study and Future Research Directions

Limited research has focused on the quality of online comments (Kim et al., 2022); hence, this study broadened the understanding of online comments by focusing on refutative features or explanations in comments. From the viewpoint of knowledge updating (Kendeou et al., 2014a; 2014b), the current study implies that the refutative elements in the short comments assist readers update their representation.

A limitation of this study is that it examines only one category of health information. Misinformation is also diffused in areas such as science and politics, but it is unclear whether the effect of the refutative elements of the comments can be replicated. Additional research is needed to determine whether the observed effect applies to diverse domains.

Additionally, we could not exclude the possibility of a floor effect in the cognitive sense. The persuasiveness of the claim shown in the experiment may have resulted in the participants' resistance to lower the agreement below a certain level. Investigating the above possibilities would deepen our understanding of this phenomenon.

Acknowledgments

This work was supported by JSPS KAKENHI Grant number 19K03253.

References

- Albarracín, D., Romer, D., Jones, C., Hall Jamieson, K., & Jamieson, P. (2018). Misleading claims about tobacco products in YouTube videos: Experimental effects of misinformation on unhealthy attitudes. *Journal of Medical Internet Research*, 20(6), e229.
- Association of College and Research Libraries (ACRL). (2000). Information literacy competency standards for higher education. Chicago, IL: ALA—American Library Association. Retrieved from <http://www.ala.org/acrl/standards/informationliteracycompetency>
- Aufderheide, P. (Ed.). (1993). *Media literacy: A report of the National Leadership Conference on Media Literacy*. Aspen, CO: Aspen Institute.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370.
- Bautista, J. R., Zhang, Y., & Gwizdka, J. (2021). Healthcare professionals' acts of correcting health misinformation on social media. *International Journal of Medical Informatics*, 148, 104375.
- Boh Podgornik, B., Dolničar, D., Šorgo, A., & Bartol, T. (2016). Development, testing, and validation of an information literacy test (ILT) for higher education. *Journal of the Association for Information Science and Technology*, 67(10), 2420–2436.
- Cameron, J., & Geidner, N. (2014). Something old, something new, something borrowed from something blue: Experiments on dual viewing TV and Twitter. *Journal of Broadcasting & Electronic Media*, 58(3), 400–419.
- Chua, A. Y., & Banerjee, S. (2017). To share or not to share: The role of epistemic belief in online health rumors. *International Journal of Medical Informatics*, 108, 36–41.
- Chua, A. Y., & Banerjee, S. (2018). Intentions to trust and share online health rumors: An experiment with medical professionals. *Computers in Human Behavior*, 87, 1–9.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
- Freeman, J. B. (2008). Argument strength, the Toulmin model, and ampliative probability. *Informal Logic*, 26(1), 25–40.
- Go, E., Jung, E. H., & Wu, M. (2014). The effects of source cues on online news perception. *Computers in Human Behavior*, 38, 358–367.
- Hinze, S. R., Slaten, D. G., Horton, W. S., Jenkins, R., & Rapp, D. N. (2014). Pilgrims sailing the Titanic: Plausibility effects on memory for misinformation. *Memory & Cognition*, 42(2), 305–324.
- Hynd, C. R. (2001). Refutational texts and the change process. *International Journal of Educational Research*, 35, 699–714.
- Inuzuka, M., Tanaka, Y., & Fujimoto, K. (2023). Effects of the valence and argument substantiveness of others' comments on viewers' validation of and attitudes toward pseudo-scientific claims. *Cogent Social Sciences*, 9(1), 2185287.
- Inuzuka, M., Tanaka, Y., & Tsubakimoto, M. (2019). Does incorporating social media messages into television programs affect the validation of incorrect argument? *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 1942–1948). Montreal.
- Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, 65(2), 371–388.
- Kendeou, P., Butterfuss, R., Kim, J., & Van Boekel, M. (2019). Knowledge revision through the lenses of the three-pronged approach. *Memory & Cognition*, 47(1), 33–46.
- Kendeou, P., & O'Brien, E. J. (2014a). The knowledge revision components (KReC) framework: Processes and mechanisms. In D. Rapp & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. Cambridge, MA: MIT Press.
- Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: why causality matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 854.
- Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014b). Knowledge revision processes in refutation texts. *Discourse Processes*, 51(5-6), 374–397.
- Kim, Y., & Gonzalez, A. L. (2022). When we tolerate online incivility: Dual-Process effects of argument strength and heuristic cues in uncivil user comments. *Computers in Human Behavior*, 131, 107235.
- Kümpel, A. S., & Unkel, J. (2020). Negativity wins at last: How presentation order and valence of user comments affect perceptions of journalistic quality. *Journal of Media Psychology*, 32(2), 89–99.
- Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, 34, 1141–1149.
- Marsh, E. J., Meade, M. L., & Roediger, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language*, 49, 519–536.
- Nussbaum, M. E. (2011). Argumentation, dialogue theory, and probability modeling: Alternative frameworks for argumentation research in education. *Educational Psychologist*, 46(2), 84–106.
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50.
- Rapp, D. N., & Salovich, N. A. (2018). Can't we just disregard fake news? The consequences of exposure to inaccurate information. *Policy Insights from the Behavioral and Brain Sciences*, 5(2), 232–239.

- Simons, M., Meeus, W., & T'Sas, J. (2017). Measuring media literacy for media education: Development of a questionnaire for teachers' competencies. *Journal of Media Literacy Education*, 9(1), 99–115.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131–142.
- Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 72–100). MIT Press.
- Sundar, S. S., Jia, H., Waddell, T. F., & Huang, Y. (2015). Toward a theory of interactive media effects (TIME): Four models for explaining how interface features affect user psychology. In S.S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 47–86). Oxford, UK: Wiley Blackwell.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7), 1275–1289.
- Trevors, G., & Kendeou, P. (2020). The effects of positive and negative emotional text content on knowledge revision. *Quarterly Journal of Experimental Psychology*, 73(9), 1326–1339.
- Waddell, T. F. (2020). The authentic (and angry) audience: How comment authenticity and sentiment impact news evaluation. *Digital Journalism*, 8(2), 249–266.
- Waddell, T. F., & Bailey, E. (2019). Is social television the “anti-laugh track”? Testing the effect of negative comments and canned laughter on comedy reception. *Psychology of Popular Media Culture*, 8(1), 99–107.
- Waddell, T. F., & Sundar, S. S. (2017). #thisshowsucks! The overpowering influence of negative social media comments on television viewers. *Journal of Broadcasting & Electronic Media*, 61(2), 393–409.
- Waddell, T. F. & Sundar, S. S. (2020). Bandwagon effects in social television: How audience metrics related to size and opinion affect the enjoyment of digital media. *Computers in Human Behavior*, 107, 106270.
- Weber, P., Prochazka, F., & Schweiger, W. (2019). Why user comments affect the perceived quality of journalistic content: The role of judgment processes. *Journal of Media Psychology: Theories, Methods, and Applications*, 31(1), 24–34.
- Ye, S., Toshimori, A., & Horita, T. (2016). The causal relationships between college students' media/social media usage and Internet literacy: Controlling for the effects of social skills and gender differences on them. *Japanese Journal of Educational Technology*, 40(3), 165–174.
- Zerback, T., & Fawzi, N. (2017). Can online exemplars trigger a spiral of silence? Examining the effects of exemplar opinions on perception of public opinion and speaking out. *New Media & Society*, 19(7), 1034–1051.