



Research Article

How effective are TikTok misinformation debunking videos?

TikTok provides opportunity for citizen-led debunking where users correct other users' misinformation. In the present study (N=1,169), participants either watched and rated the credibility of (1) a misinformation video, (2) a correction video, or (3) a misinformation video followed by a correction video ("debunking"). Afterwards, participants rated both a factual and a misinformation video about the same topic and judged the accuracy of the claim furthered by the misinformation video. We found modest evidence for the effectiveness of debunking on people's ability to subsequently discern between true and false videos, but stronger evidence on subsequent belief in the false claim itself.

Authors: Puneet Bhargava (1), Katie MacDonald (2), Christie Newton (2), Hause Lin (1,3), Gordon Pennycook (1,2)

Affiliations: (1) Hill/Levene Schools of Business, University of Regina, Canada, (2) Department of Psychology, University of Regina, Canada, (3) Massachusetts Institute of Technology, Sloan School of Management, USA

How to cite: Bhargava, P., MacDonald, K. L., Newton, C., Lin, H., & Pennycook, G. (2023). Are TikTok misinformation debunking videos effective? *Harvard Kennedy School (HKS) Misinformation Review*, 4(2).

Received: October 25th, 2022. Accepted: February 27th, 2023. Published: March 29th, 2023.

Research question

- How effective are TikTok misinformation debunking videos?

Essay summary

- We conducted a preregistered survey experiment with 1,169 U.S. American participants who saw TikTok-style videos on one of six misinformation topics (all of which were found on TikTok): aspartame, an artificial sweetener, causes cancer; COVID-19 isn't dangerous given that infected people can be asymptomatic; the accidental shooting by the actor Alec Baldwin on the Rust movie set was on purpose; natural immunity is preferable to vaccinations; Ivermectin, an antiparasitic drug, can effectively treat COVID-19 symptoms; and simple tests can differentiate between left vs. right-brained people.
- Participants were randomly assigned to one of three conditions. They saw a misinformation video (misinformation-only condition), a correction video (correction-only condition), or a misinformation video followed by a correction video (debunking condition).
- Participants rated the credibility of a false and a true video related to their assigned topic and indicated how much they agreed with a false statement about their topic.
- We found a marginally significant interaction between the truth value of the video and the

¹ A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University, John F. Kennedy School of Government.

experimental condition (debunking vs. misinformation-only) for credibility ratings of post-treatment videos, indicating that debunking had a marginal effect on increased people's ability to distinguish between subsequent true and false videos on the same topic.

- Critically, belief in the misinformation claim was significantly lower in the debunking condition compared to the misinformation-only condition.
- Qualitatively, effects were weaker in a correction-only condition relative to the debunking condition, indicating that the correction video was more effective if shown after the initial misinformation video.

Implications

The spread of misinformation on social media is a matter of growing concern (Lazer et al., 2018); suitably, research on the topic has massively grown in recent years (Tucker et al., 2018). In particular, investigating potential interventions against misinformation is a major focus (Pennycook & Rand, 2021), including a great deal of research specifically on the efficacy of fact-checking, corrections, and debunking (Chan et al., 2017; Nieminen & Rapeli, 2018; Porter & Wood, 2021). For example, it has been found that presenting a correction message with arguments that showcase a prior message as misinformation (i.e., “debunking”) can counter misinformation (Chan et al., 2017). Although debunking has shown some promise (Chan et al., 2017; Wood & Porter, 2019), studies on its effectiveness have yielded mixed results (Ecker et al., 2022; Nieminen & Rapeli, 2018). However, these studies tend to use static content, and misinformation comes in many forms. In particular, misinformation videos may pose a uniquely difficult target for debunking attempts because they often appear highly immersive, authentic, and relatable (Wang, 2020), which might cause people to process videos more superficially and believe them more readily (Sundar et al., 2021).

TikTok, a social media platform in which users upload short videos, has recently emerged as the fastest-growing social media platform and reportedly has over a billion monthly users (Bursztynsky, 2021). TikTok's growing popularity and swelling user base raise concerns that TikTok may become a major source of misinformation (Basch et al., 2021). For example, in one analysis taken between January and March in 2020, it was found that 20–32% of the sampled COVID-19-related videos on TikTok contained some misleading or incorrect information (Southwick et al., 2021). Similarly, misinformation was surprisingly common in videos with masking-related hashtags when focusing on either the most-viewed videos and most-liked comments (Baumel et al., 2021), with falsehoods occurring from 6% to 45% of the time depending on the hashtag.

For these reasons, there has been increased focus on how to detect (Shang et al., 2021) and correct (Bautista et al., 2021) misinformation on TikTok. Interestingly, TikTok users can utilize its two unique editing features (“stitch,” which allows incorporation of someone else's video in the beginning of a new video, and “duet,” which facilitates split-screen or picture-in-picture playback of someone else's video and a new video) to incorporate misinformation TikTok videos into their own to create new citizen-led correction videos. Thus, although users may leverage the TikTok format to create compelling misinformation videos, the same can be true of videos that correct misinformation.

The current study aims to test the effectiveness of TikTok correction videos through a survey experiment that presented participants with either a misinformation video, a correction video, or both (in that order). This allowed us to test whether correction videos decreased susceptibility to misinformation by comparing a condition where participants are only exposed to misinformation (i.e., the “misinformation-only” baseline condition) with a condition where they receive a correction either on its own (the “correction-only” condition) or after having seen the misinformation video (the “debunking” condition). To assess the efficacy of corrections, we focused on two outcomes: 1) the ability of users to

distinguish between different (and subsequently presented) true and false videos that are on the same topic (with the goal of assessing subsequent “on platform” behavior); that is, we assessed whether participants rated the subsequent true video as being more credible than the subsequent false video, and 2) whether users actually believed the false claim, measured using agreement with a false statement related to the topic of the misinformation video. To simplify the analysis, we focused on comparing the baseline “misinformation-only” condition against the two correction conditions (i.e., people either received only the correction video [the “correction-only” condition] or the misinformation video and then the correction video [the “debunking” condition]).

Our findings reveal moderate evidence for the effectiveness of debunking on TikTok. Watching a correction video does appear to improve credibility judgments of the same-topic subsequent videos watched on TikTok. However, this effect is weak and only marginally statistically significant in some specifications of the analysis. Furthermore, this weak effect only occurs if the correction video is immediately preceded by the misinformation video (i.e., debunking condition). Watching just the correction video did not improve our participants’ ability to distinguish between subsequent true and false videos on the same topic. Importantly, however, belief in false claims was lower for both the correction-only and debunking conditions relative to the misinformation-only control, indicating that the correction videos were effective in decreasing susceptibility to misinformation and that this occurred regardless of whether the context of the original video was present or not. These findings offer a strong parallel to the existing conclusion in the literature that there is mixed efficacy of using debunking as an intervention against misinformation. Overall, though, the videos were sufficiently effective to be considered a potential avenue for future intervention attempts and certainly a fruitful avenue for future research.

TikTok has become an important medium, not simply for entertainment but also for conveying information, and this is particularly true among younger people. Hence, studying misinformation interventions on TikTok is necessary to work towards creating better information environments into the future. To that end, our findings are useful for content creators on TikTok and for the platform itself. For creators, our results indicate that correction videos that contain misinformation claims in the beginning do have some efficacy in decreasing false beliefs. As such, users are encouraged to continue engaging in citizen-led debunking. In terms of the platform itself, our findings indicate that there is value in up-ranking correction videos to the extent possible. Future research is needed to test for how long TikTok debunking is effective, as well as whether different elements of TikTok videos (e.g., engagement numbers) influence the efficacy of debunking. More broadly, our results indicate that debunking TikTok videos can have some influence, and this is encouraging for those interested in correcting falsehoods on the fast-paced platform.

Findings

Finding 1: Debunking (vs. misinformation-only) marginally improves truth discernment for subsequent TikTok videos.

We first tested whether debunking (i.e., presenting a correction video after a misinformation video) improved people’s ability to distinguish between subsequent (i.e., post-treatment) true and false videos on the same topic. As preregistered (see [OSF](#)), we fitted a 2 (video veracity: true vs. false) x 2 (condition: debunking vs. misinformation-only) mixed analysis of variance (ANOVA). Overall, the post-treatment true videos were found to be more credible than false ones, $F(1, 783) = 757.61, p < .001, d = 0.98$. Crucially, as predicted, we found that this difference between true and false videos was (marginally) greater for the debunking condition relative to the misinformation-only condition, as evidenced by an interaction between veracity and condition, $F(1, 783) = 3.22, p = .073, d = 0.06$, Figure 1. Put differently, debunking

marginally improved overall accuracy for the subsequent video evaluation task. Excluding participants who failed attention checks resulted in a slightly stronger effect, $F(1, 757) = 3.99, p = .046, d = 0.07$ (see Appendix A). In terms of the underlying pattern of data, debunking (vs. misinformation-only) increased subsequent credibility ratings for true videos, $b = 3.68 (1.78), t(1568) = 2.06, p = .039$, but had no effect on credibility ratings for false videos, $b = -1.26 (1.82), t(1568) = -0.70, p = .487$, leading to overall higher truth discernment in the debunking condition relative to the misinformation-only condition. This difference can be seen in Figure 2 (top).

Next, we tested whether a correction video by itself improved people's ability to distinguish between subsequent true and false videos (i.e., unlike in the debunking conditions, participants did not watch the misinformation video prior to watching the correction video—instead, they only watched the correction video). For this, we fitted a 2 (video veracity: true vs. false) x 2 (condition: correction-only vs. misinformation-only) mixed ANOVA (Figure 1). As above, true post-treatment videos were rated as more credible than false ones, $F(1, 777) = 639.48, p < .001, d = 0.91$. However, unlike for the debunking condition, this difference between true and false videos was not greater for the correction-only condition relative to the misinformation-only condition, as evidenced by a non-significant interaction between veracity and condition, $F(1, 777) = 0.08, p = .771, d = 0.01$. Thus, presenting only the correction video without the context of the original falsehood did not produce a significant increase in accuracy for the subsequent videos. Nonetheless, when we compared the debunking versus correction-only conditions with a 2 (video veracity: true vs. false) x 2 (condition: debunking vs. correction-only) mixed ANOVA, there were no differences in credibility or in the effect of veracity on credibility between conditions, $F_s < 2.11, p_s > .147, d_s < 0.05$. Thus, although discernment was higher in the debunking condition relative to the misinformation condition (and this was not true of the correction condition), there was no difference between the debunking and correction conditions. See Appendix A for results from models that excluded participants who failed attention checks.

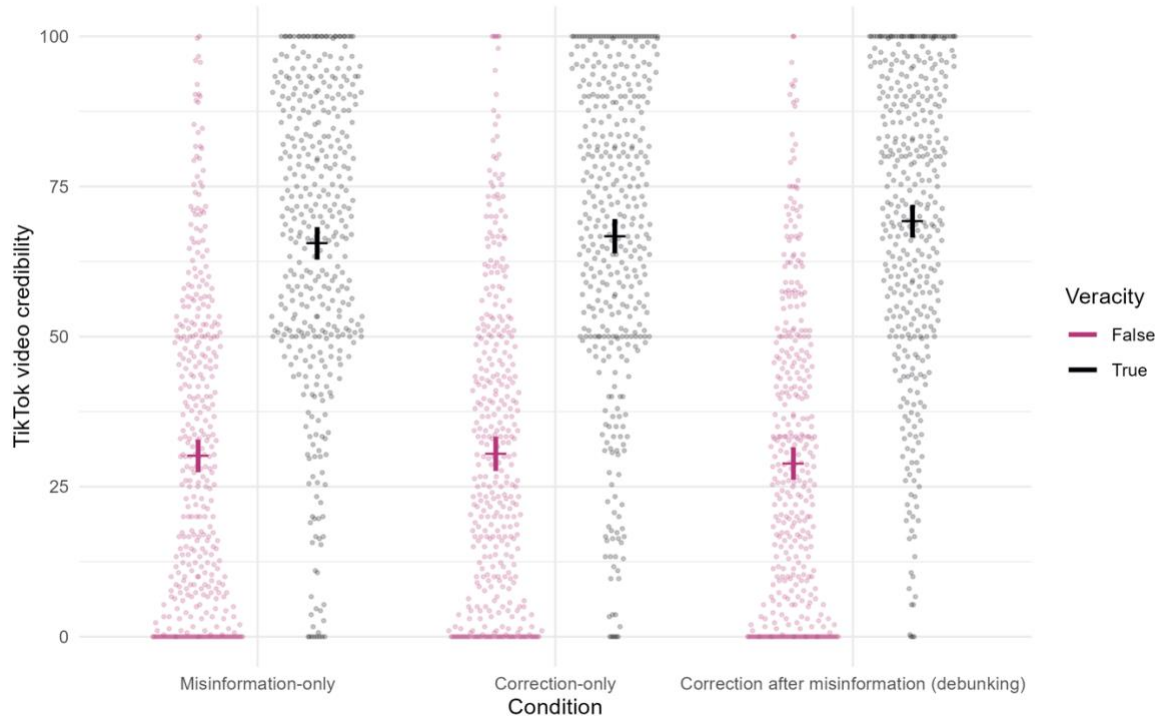


Figure 1. TikTok video credibility ratings for subsequent false and true videos and for the three experimental conditions. Credibility is computed by averaging reliability, unbiasedness, and accuracy ratings. Each dot is one participant's rating. Means (horizontal line forming a cross) and 95% CIs (vertical line forming a cross) are shown.

Finding 2: Debunking (vs. misinformation-only) reduces subsequent false belief.

In addition to rating the credibility of the TikTok videos, participants also indicated to what extent they believed false statements associated with each TikTok video (Figure 2, bottom). Overall, participants in the debunking (vs. misinformation-only) condition believed the false statements less, $b = -0.34$ (0.12), $t(1166) = -2.99$, $p = .003$, $d = -0.17$, and the effects were relatively consistent across topics (Figure 2, bottom left). Similarly, participants in the correction-only (vs. misinformation-only) condition also believed the false statements less, $b = -0.23$ (0.12), $t(1166) = -1.99$, $p = .046$, $d = -0.12$, but the effect was smaller and varied more across topics (Figure 2, bottom right). Together, these results suggest that showing debunking videos on TikTok had relatively weak effects on increasing subsequent video discernment but more robust effects on decreasing belief in false statements.

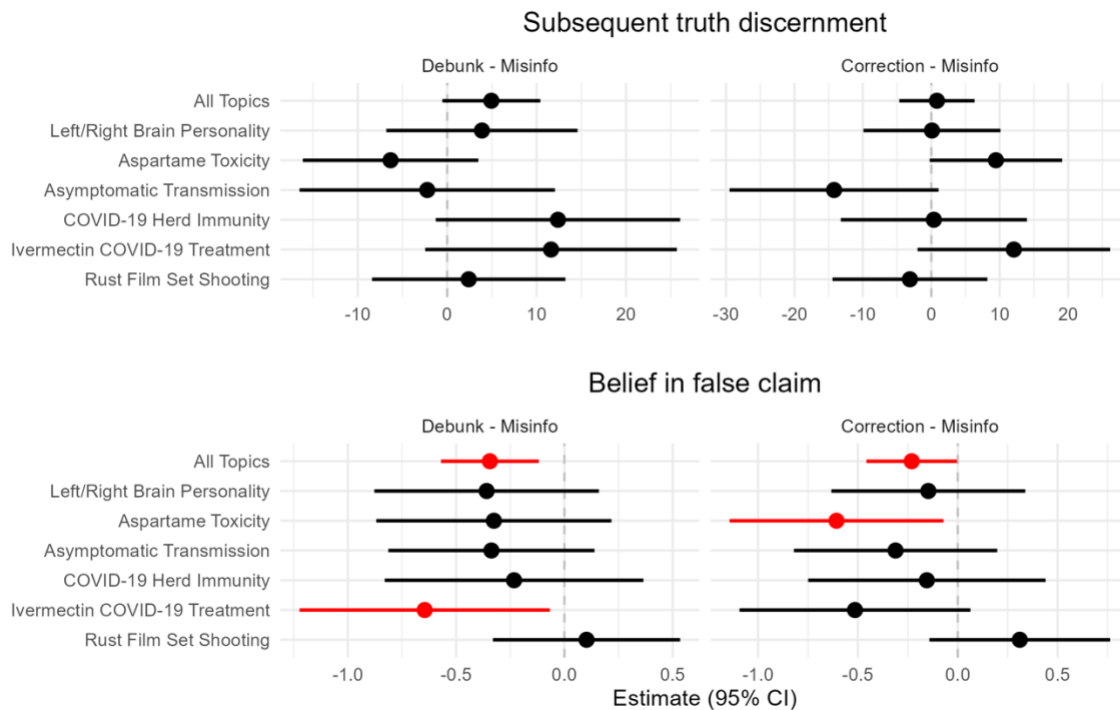


Figure 2. Effect of condition on TikTok video discernment (top) and belief in false statements associated with the TikTok videos (bottom). Subsequent truth discernment refers to the credibility ratings of true videos minus credibility ratings of false videos (i.e., overall accuracy). Left panels contrast the debunking and misinformation-only conditions; right panels contrast the correction-only and misinformation-only conditions. The top row in each panel is the overall estimate for all six topics; the next six rows are the topic-specific estimates. Estimates highlighted in red are statistically significant ($p < .05$).

To summarize, we found that people were marginally better at distinguishing between true and false videos in the debunking relative to the misinformation-only condition, indicating some efficacy of presenting correction videos after an initial false video. Furthermore, belief in a relevant false statement was significantly lower in debunking (vs. misinformation-only) condition. Therefore, debunking showed a stronger effect for false belief reduction, signifying cross-platform applicability of results (for example, when users browse static social media websites after viewing videos on TikTok). Finally, correction videos were not as discernably effective if they were presented outside the context of the original falsehood.

Methods

A total of 1,363 American participants were recruited via Prolific, an online recruitment service that gives a nationally representative sample based on quota-matching age, sex, region, and ethnicity. We removed 161 participants who did not complete the survey. The preregistration can be found on [OSF](#). Following our preregistered exclusion plan, we excluded participants who failed the first attention check ($n = 32$; see Figure 3). Participants were then informed that they would watch videos obtained from TikTok and that we were interested in their thoughts on the video's accuracy. Before watching the TikTok videos, participants were instructed to turn on their audio device's volume and had to verify that the volume was on (one participant was excluded because they could not turn the volume on). In total, 1,169 participants completed the study (mean age = 34.7, $SD = 13.1$; 347 male, 784 female, 38 non-binary/did not respond). There were also 33 participants who failed two additional (post-treatment) attention checks—see Appendix A for analyses with these participants removed. A majority (62%) of the sample had TikTok on their personal device, and the majority (68%) of these users watched videos on TikTok at least daily.

Participants were randomly assigned to watch and rate TikTok videos about one of six misinformation topics: aspartame toxicity (aspartame, an artificial sweetener, causes cancer), COVID-19 asymptomatic transmission (COVID-19 isn't dangerous given that infected people can be asymptomatic), Rust film set shooting (the accidental shooting of Halyna Hutchins on the Rust movie set was not an accident), COVID-19 herd immunity (natural immunity is a better way to end the pandemic than vaccinations), Ivermectin COVID-19 treatment (Ivermectin can effectively treat COVID-19 symptoms), left or right brain personality traits (simple brain tests can differentiate between left vs. right-brained people). These videos were found on TikTok and selected because a) they contained false information, b) were subject to subsequent correction videos (where they were specifically debunked), and c) there were other similar claims that could be found in other TikTok videos (that did not make reference to either the initial misinformation video or its correction). There were no additional selection criteria based on features of the video (beyond that, they contained misinformation or a correction), and hence the videos vary from each other in several aspects. All videos can be found on [OSF](#).

All participants were also randomly assigned to the debunking condition ($N = 390$), the disinformation-only condition ($N = 395$), or the correction-only condition ($N = 384$). Participants in the debunking condition watched a randomly-selected misinformation TikTok video followed by its contrasting debunking TikTok video. Participants in the misinformation-only condition watched a single misinformation TikTok video, whereas participants in the correction-only condition watched a single debunking TikTok video. Afterwards, all participants watched two subsequent videos on the same topic, one misinformation and one factual (the order of which was counterbalanced). The video credibility evaluations measure was given to the participants to answer after watching each TikTok video. For the video credibility evaluations, participants had to rate the accuracy, reliability, and impartiality of the information in the given video on a scale from 0 ("Not at all") to 100 ("Extremely"). The scores on the three items were averaged to create the video credibility ratings. The three measures produced a reliable single credibility score (across items), Cronbach's $\alpha = .92$, although impartiality was not as highly correlated with accuracy ($r = .73$) or reliability ($r = .74$) as accuracy and reliability were correlated with each other ($r = .94$). Means and standard deviations for all videos across conditions can be found in Appendix B.

After watching and rating the TikTok videos, participants answered an attention check question (21 participants failed; see Appendix C for wording). Participants then completed the follow-up questionnaire about the six misinformation topics, wherein they marked their agreement with one false statement per topic on a scale from 1 ("Strongly disagree") to 7 ("Strongly agree"). Participants rated their agreement for all six topics, regardless of which video they watched. For example, for the aspartame topic,

participants were asked to rate statements such as “Aspartame (an ingredient in diet soda) causes cancer.” After the topic questions, participants were asked if TikTok was downloaded on their electronic devices. Participants who answered “yes” were redirected to the two-item questionnaire about their TikTok usage and then directed to complete the Comprehensive Thinking Styles Questionnaire (CTSQ) (Newton et al., 2021). Participants who answered “No” were directly forwarded to the CTSQ. After the CTSQ, participants received another attention check question followed by the demographic questions (12 participants failed). To conclude, we asked participants how long it took them to complete the survey and allowed them to leave any additional comments (see Figure 3).

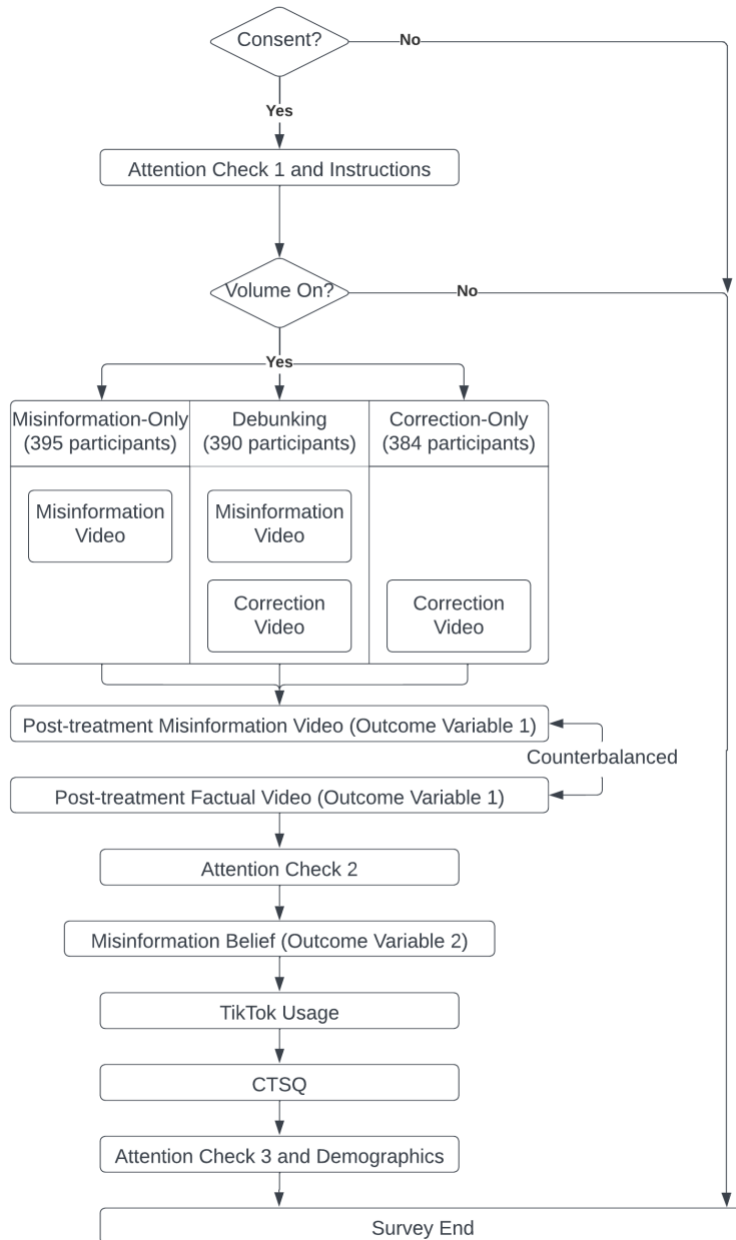


Figure 3. Experiment flowchart.

Bibliography

- Basch, C. H., Meleo-Erwin, Z., Fera, J., Jaime, C., & Basch, C. E. (2021). A global pandemic in the time of viral memes: COVID-19 vaccine misinformation and disinformation on TikTok. *Human Vaccines & Immunotherapeutics*, 17(8), 2373–2377. <https://doi.org/10.1080/21645515.2021.1894896>
- Baumel, N. M., Spatharakis, J. K., Karitsiotis, S. T., & Sellas, E. I. (2021). Dissemination of mask effectiveness misinformation using TikTok as a medium. *Journal of Adolescent Health*, 68(5), 1021–1022. <https://doi.org/10.1016/j.jadohealth.2021.01.029>
- 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. <https://doi.org/10.1016/j.ijmedinf.2021.104375>
- Bursztynsky, J. (2021, September 27). *TikTok says 1 billion people use the app each month*. CNBC. <https://www.cnbc.com/2021/09/27/tiktok-reaches-1-billion-monthly-users.html>
- Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546. <https://doi.org/10.1177/0956797617714579>
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Lazer, D., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 9(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Newton, C., Feeney, J., & Pennycook, G. (2021). *On the disposition to think analytically: Four distinct intuitive-analytic thinking styles*. PsyArXiv. <https://doi.org/10.31234/OSF.IO/R5WEZ>
- Nieminen, S., & Rapeli, L. (2018). Fighting misperceptions and doubting journalists' objectivity: A review of fact-checking literature. *Political Studies Review*, 17(3), 296–309. <https://doi.org/10.1177/1478929918786852>
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
- 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*, 118(37). <https://doi.org/10.1073/pnas.2104235118>
- Shang, L., Kou, Z., Zhang, Y., & Wang, D. (2021, December). A multimodal misinformation detector for COVID-19 short videos on TikTok. In *2021 IEEE international conference on big data (big data)* (pp. 899–908). IEEE. <https://doi.org/10.1109/BigData52589.2021.9671928>
- Southwick, L., Guntuku, S. C., Klinger, E. V., Seltzer, E., McCalpin, H. J., & Merchant, R. M. (2021). Characterizing COVID-19 content posted to TikTok: Public sentiment and response during the first phase of the COVID-19 pandemic. *Journal of Adolescent Health*, 69(2), 234–241. <https://doi.org/10.1016/j.jadohealth.2021.05.010>
- Sundar, S. S., Molina, M. D., & Cho, E. (2021). Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps? *Journal of Computer-Mediated Communication*, 26(6), 301–319. <https://doi.org/10.1093/jcmc/zmab010>
- Tucker, J., Guess, A. M., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature*. SSRN. <https://doi.org/10.2139/ssrn.3144139>

- Wang, Y. (2020). Humor and camera view on mobile short-form video apps influence user experience and technology-adoption intent, an example of TikTok (DouYin). *Computers in Human Behavior*, *110*, 106373. <https://doi.org/10.1016/j.chb.2020.106373>
- Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135–163. <https://doi.org/10.1007/s11109-018-9443-y>

Funding

We gratefully acknowledge research funding from the Social Sciences and Humanities Research Council of Canada, The John Templeton Foundation, the Government of Canada's Digital Citizen Contribution Program, and the U.S. Department of Defense.

Competing interests

G.P. was previously Faculty Research Fellow for Google and has received research funding from them.

Ethics

This research received ethics approval from the University of Regina Research Ethics Board.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Data availability

All materials needed to replicate this study are available via the Harvard Dataverse: <https://doi.org/10.7910/DVN/OBL67B> and OSF: <https://osf.io/xfcsp>

Appendix A: Findings with only attentive participants

Thirty-three participants failed the last two attention checks. Similar to the main findings, below we report the findings after analyzing data of the remaining 1,136 participants.

Finding 1: Debunking (vs. misinformation-only) improves credibility of truthful TikTok videos.

We fitted a 2 (video veracity: true vs. false) x 2 (condition: debunking vs. misinformation-only) mixed ANOVA. Post-treatment true videos were found to be more credible than false ones, $F(1, 757) = 745.85$, $p < .001$, $d = 0.99$, but the credibility of post-treatment videos did not vary between conditions, $F(1, 757) = 1.27$, $p = .261$, $d = 0.04$. Crucially, as predicted, we found a significant interaction effect of veracity and condition on the credibility ratings of the subsequent TikTok videos, $F(1, 757) = 3.99$, $p = .046$, $d = 0.07$ (see Figure 4); that is, debunking improved overall accuracy for the subsequent video evaluation task. More specifically, debunking (vs. misinformation-only) increased subsequent credibility ratings for true videos, $b = 4.13$ (1.80), $t(1516) = 2.29$, $p = .022$, but had no effect on credibility ratings for false videos, $b = -1.48$ (1.86), $t(1516) = -0.79$, $p = .428$, leading to overall higher truth discernment in the debunking condition relative to the misinformation-only condition. This difference can be seen in Figure 5 (top).

Next, we fitted a 2 (video veracity: true vs. false) x 2 (condition: correction-only vs. misinformation-only) mixed ANOVA (Figure 4). True post-treatment videos were rated as more credible than false ones, $F(1, 758) = 634.00$, $p < .001$, $d = 0.91$, but credibility did not vary between conditions, $F(1, 758) = 0.40$, $p = .526$, $d = 0.02$, and the effect of veracity on credibility ratings of post-treatment videos, also, did not vary between conditions, $F(1, 758) = 0.17$, $p = .680$, $d = 0.01$. Thus, presenting only the correction video without the context of the original falsehood did not produce a significant increase in accuracy for the subsequent videos. Nonetheless, when we compared the debunking versus correction-only conditions with a 2 (video veracity: true vs. false) x 2 (condition: debunking vs. correction-only) mixed ANOVA, true (vs. false) post-treatment videos were found to be more credible, $F(1, 751) = 742.10$, $p < .001$, $d = 0.99$; but there were no differences in credibility or in the effect of veracity on credibility between conditions $F_s < 2.39$, $p_s > .122$, $d_s < 0.06$. Thus, although discernment was higher in the debunking condition relative to the misinformation condition (and this was not true of the correction condition), there was no difference between the debunking and correction conditions.

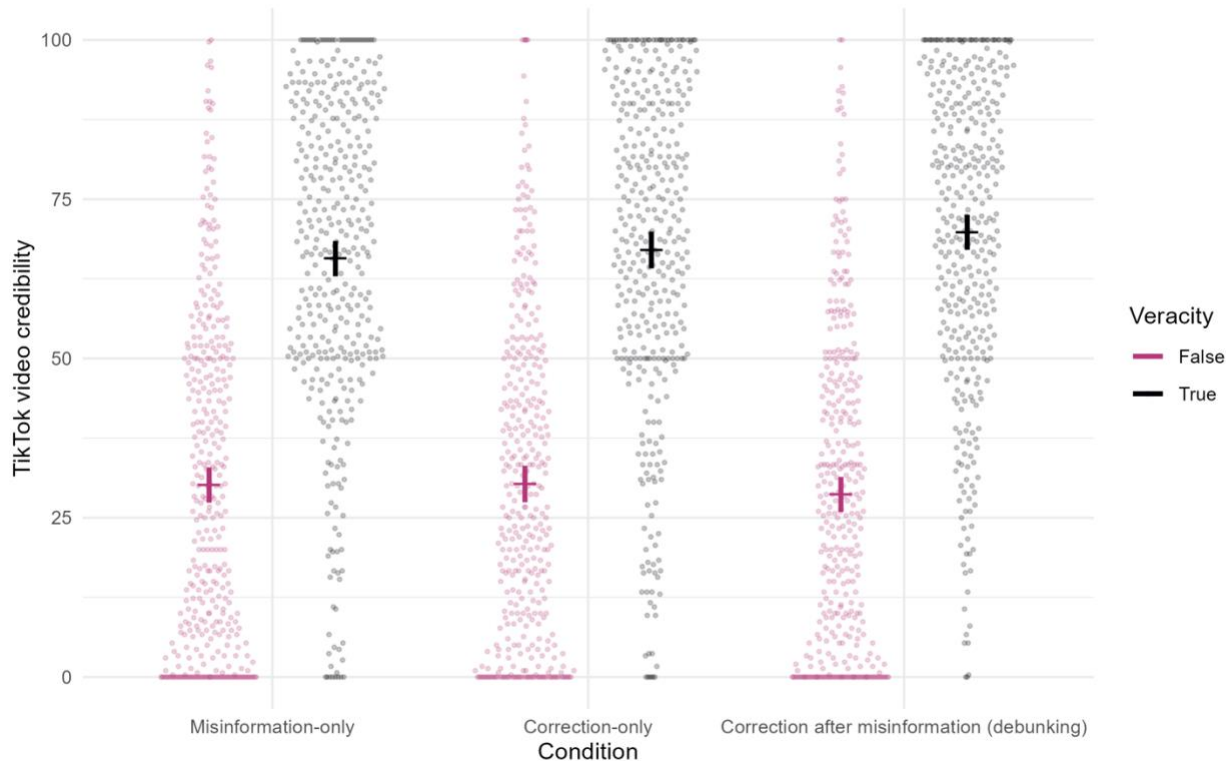


Figure 4. TikTok video credibility ratings for false and true videos and for the three experimental conditions. Credibility is computed by averaging reliability, unbiasedness, and accuracy ratings. Each dot is one participant's rating. Means and 95% CIs are shown.

Finding 2: Debunking (vs. misinformation-only) improves discernment, reduces false belief.

To further probe the effect of veracity on credibility ratings of post-treatment videos between conditions, we computed discernment scores for the video credibility ratings (credibility of true minus false videos (Figure 5 top). Replicating the mixed ANOVA results above, we find that, when collapsing across topics, discernment was higher in the debunking than misinformation-only condition, $b = 5.61 (2.84)$, $t(1133) = 1.97$, $p = .049$, $d = 0.12$; Figure 5 top left), but discernment was not significantly higher in the correction-only than misinformation-only condition, $b = 1.18 (2.84)$, $t(1133) = 0.42$, $p = .677$, $d = 0.02$ (Figure 5 top right). Although debunking increased discernment when collapsing across topics, the effects varied substantially across the six topics (Figure 5 top left).

In addition to rating the credibility of the TikTok videos, participants also indicated to what extent they agreed with or believed false statements associated with each TikTok video (Figure 5 bottom). Overall, participants in the debunking (vs. misinformation-only) condition believed the false statements less, $b = -0.37 (0.12)$, $t(1133) = -3.11$, $p = .002$, $d = -0.18$, and the effects were relatively consistent across topics (Figure 5 bottom left). Similarly, participants in the correction-only (vs. misinformation-only) condition also believed the false statements less, $b = -0.24 (0.12)$, $t(1133) = -2.06$, $p = .039$, $d = -0.12$, but the effect was smaller and varied more across topics (Figure 5 bottom right). Together, these results suggest that showing debunking videos on TikTok increased subsequent video discernment (but effects were small), whereas both showing debunking videos and correction decreased belief in false statements (effects were stronger for debunking than correction).

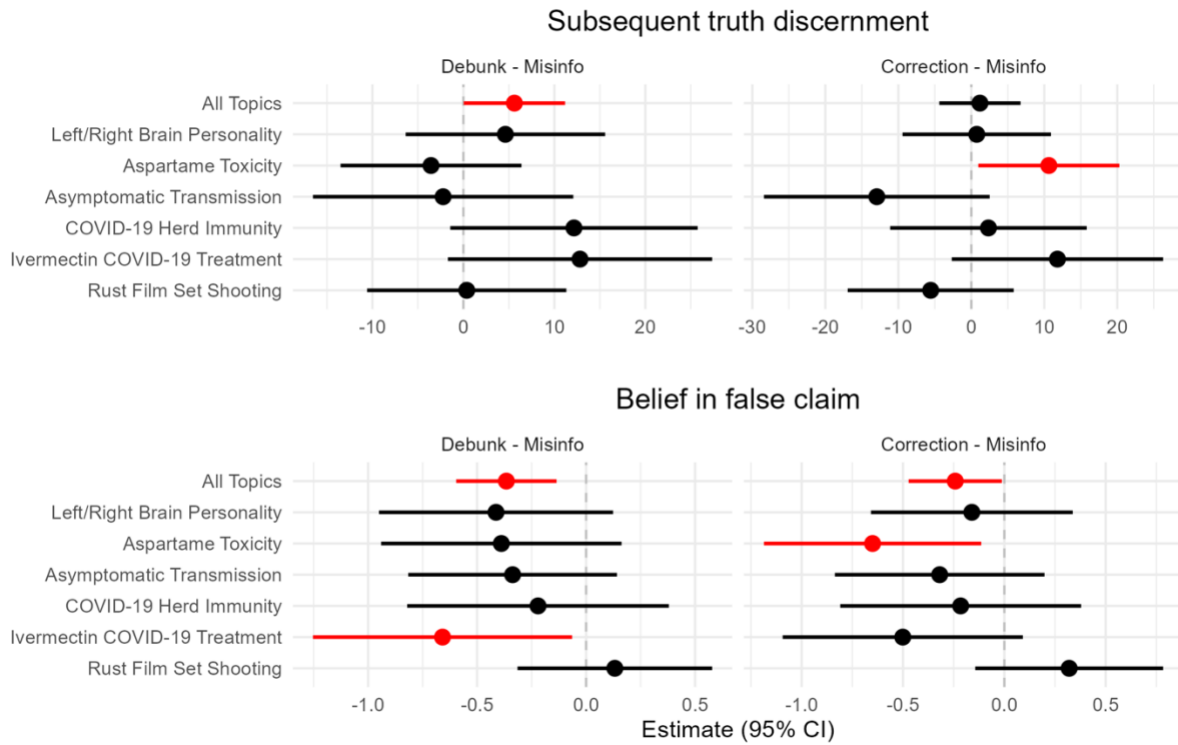


Figure 5. Effect of condition on TikTok video discernment (top) and belief in false statements associated with the TikTok videos (bottom). Discernment is the credibility ratings of true videos minus credibility ratings of false videos. Left panels contrast the debunking and misinformation-only conditions; right panels contrast the correction-only and misinformation-only conditions. The top row in each panel is the overall estimate for all six topics; the next six rows are the topic-specific estimates. Estimates highlighted in red are statistically significant ($p < .05$).

Appendix B: Means and SDs for all videos

Table 1. Means and SDs for all videos by condition.

Topic	Video Type	Condition					
		Misinfo-Only		Debunking		Correction-Only	
		M	SD	M	SD	M	SD
aspartame - an artificial sweetener - causes cancer	Misinformation Video	42.86	25.78	54.36	23.97	-	
	Correction Video	-		66.93	22.68	63.75	29.11
	Post-treatment Misinformation Video	43.23	27.45	45.69	24.19	32.89	24.39
	Post-treatment Factual Video	50.78	21.26	46.91	19.1	49.88	25.01
COVID-19 isn't dangerous given that infected people can be asymptomatic	Misinformation Video	27.57	25.47	29.2	31.34	-	
	Correction Video	-		78.71	24.29	73.17	26.58
	Post-treatment Misinformation Video	15.57	21.38	18.31	23.93	26.86	26.79
	Post-treatment Factual Video	75.64	23.05	76.15	26.6	72.72	28.97
the accidental shooting by the actor Alec Baldwin on the Rust movie set was on purpose	Misinformation Video	29.02	21.92	26.87	21.13	-	
	Correction Video	-		64.31	24.48	61.83	21.4
	Post-treatment Misinformation Video	13.26	16.23	15.04	18.52	18.62	20.58
	Post-treatment Factual Video	68.87	25.74	73.06	22.97	71.11	18.57
natural immunity is preferable to vaccinations	Misinformation Video	46.55	25.21	38.03	25.32	-	
	Correction Video	-		72.56	19.33	62.12	27.52
	Post-treatment Misinformation Video	29.91	25.97	27.97	23.22	31.84	27.43
	Post-treatment Factual Video	63.94	26.2	74.38	20.2	66.24	27.63

Ivermectin - an antiparasitic drug - can effectively treat COVID- 19 symptoms	Misinformation Video	27.45	18.71	33.42	28.62	-	
	Correction Video	-		75.47	19.16	72.73	25
	Post-treatment Misinformation Video	37.62	23.72	38.79	24.13	35.64	27.21
	Post-treatment Factual Video	54.43	24.56	67.22	24.12	64.52	26.58
simple tests can differentiate between left vs. right brained people	Misinformation Video	42.71	23.35	40.44	20.97	-	
	Correction Video	-		69.57	22.09	64.95	23.92
	Post-treatment Misinformation Video	38.93	24.22	30.96	20.87	34.32	23.18
	Post-treatment Factual Video	78.64	19.23	74.56	20.8	74.13	20.83

Appendix C: Attention check questions

We would like to get a sense of your general preferences.

Most modern theories of decision making recognize that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables can greatly impact the decision process. To demonstrate that you've read this much, just go ahead and select both red and green among the alternatives below, no matter what your favorite color is. Yes, ignore the question below and select both of those options.

What is your favorite color?

White	Pink
Black	Green
Red	Blue

Figure 6. Attention check 1.

Below is a list of leisure activities. If you are reading this, please choose the "other" box below and type in "I read the instructions".

Sports
Television
Computer games
Surfing the internet
Board games
Camping
Spending time with family and friends
Other
<input type="text"/>

Figure 7. Attention check 2.

People are very busy these days and many do not have time to follow what goes on in the government. We are testing whether people read questions. To show that you've read this much, answer both "extremely interested" and "very interested":

Not at all interested	Slightly interested	Moderately interested	Very interested	Extremely interested
-----------------------	---------------------	-----------------------	-----------------	----------------------

Figure 8. Attention check 3.