The Spanish Journal of Psychology (2021), **24**, e25, 1–9.

doi:10.1017/SJP.2021.23

How Can Psychological Science Help Counter the Spread of Fake News?

Sander van der Linden, Jon Roozenbeek, Rakoen Maertens, Melisa Basol, Ondřej Kácha, Steve Rathje and Cecilie Steenbuch Traberg

University of Cambridge (UK)

Abstract. In recent years, interest in the psychology of fake news has rapidly increased. We outline the various interventions within psychological science aimed at countering the spread of fake news and misinformation online, focusing primarily on corrective (*debunking*) and pre-emptive (*prebunking*) approaches. We also offer a research agenda of open questions within the field of psychological science that relate to how and why fake news spreads and how best to counter it: the longevity of intervention effectiveness; the role of sources and source credibility; whether the sharing of fake news is best explained by the motivated cognition or the inattention accounts; and the complexities of developing psychometrically validated instruments to measure how interventions affect susceptibility to fake news at the individual level.

Received 28 December 2020; Revised 10 January 2021; Accepted 25 February 2021

Keywords: debunking, fake news, inoculation, misinformation, prebunking

Fake news can have serious societal consequences for science, society, and the democratic process (Lewandowsky et al., 2017). For example, belief in fake news has been linked to violent intentions (Jolley & Paterson, 2020), lower willingness to get vaccinated against the coronavirus disease 2019 (COVID-19; Roozenbeek, Schneider, et al., 2020), and decreased adherence to public health guidelines (van der Linden, Roozenbeek, et al., 2020). Fake rumours on the WhatsApp platform have inspired mob lynchings (Arun, 2019) and fake news about climate change is undermining efforts to mitigate the biggest existential threat of our time (van der Linden et al., 2017).

In light of this, interest in the "psychology of fake news" has skyrocketed. In this article, we offer a rapid review and research agenda of how psychological science can help effectively counter the spread of fake news, and what factors to take into account when doing so.[1]

Correspondence concerning this article should be addressed to Sander van der Linden. University of Cambridge. Department of Psychology. Downing Street. CB2 3EB Cambridge (UK).

E-mail: sander.vanderlinden@psychol.cam.ac.uk

Conflicts of Interest: None.

Funding Statement: This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

[1]For more comprehensive reviews of misinformation research, we refer to Lazer et al. (2018), Lewandowsky et al. (2017), Flynn et al. (2017), Greifeneder et al. (2020), Lewandowsky and van der Linden (2021) and van Bavel et al. (2021).

Current Approaches to Countering Misinformation

Scholars have largely offered two different approaches to combat misinformation, one reactive, the other proactive. We review each approach in turn below.

Reactive Approaches: Debunking and Fact-Checking

The first approach concerns the efficacy of debunking and debiasing (Lewandowsky et al., 2012). Debunking misinformation comes with several challenges, as doing so reinforces the (rhetorical frame of the) misinformation itself. A plethora of research on the *illusory truth effect* suggests that the mere repetition of information increases its perceived truthfulness, making even successful corrections susceptible to unintended consequences (Effron & Raj, 2020; Fazio et al., 2015; Pennycook et al., 2018). Despite popular concerns about potential backfire-effects, where a correction inadvertently increases the belief in—or reliance on—misinformation itself, research has not found such effects to be commonplace (e.g., see Ecker et al., 2019; Swire-Thompson et al., 2020; Wood & Porter, 2019). Yet,

How to cite this article:
van der Linden, S., Roozenbeek, J., Maertens, R., Basol, M., Kácha, O., Rathje, S., & Traberg, C. S. (2021). How can psychological science help counter the spread of fake news? *The Spanish Journal of Psychology, 24.* e25. Doi:10.1017/SJP.2021.23

there is reason to believe that debunking misinformation can still be challenging in light of both (politically) motivated cognition (Flynn et al., 2017), and the continued influence effect (CIE) where people continue to retrieve false information from memory despite acknowledging a correction (Chan et al., 2017; Lewandowsky et al., 2012; Walter & Tukachinsky, 2020). In general, effective debunking requires an alternative explanation to help resolve inconsistencies in people's mental model (Lewandowsky, Cook, et al., 2020). But even when a correction is effective (Ecker et al., 2017; MacFarlane et al., 2020), fact-checks are often outpaced by misinformation, which is known to spread faster and further than other types of information online (Petersen et al., 2019; Vosoughi et al., 2018).

Proactive Approaches: Inoculation Theory and Prebunking

In light of the shortcomings of debunking, scholars have called for more proactive interventions that reduce whether people believe and share misinformation in the first place. *Prebunking* describes the process of *inoculation*, where a forewarning combined with a pre-emptive refutation can confer psychological resistance against misinformation. Inoculation theory (McGuire, 1970; McGuire & Papageorgis, 1961) is the most well-known psychological framework for conferring resistance to persuasion. It posits that *pre-emptive* exposure to a weakened dose of a persuasive argument can confer resistance against future attacks, much like a medical vaccine builds resistance against future illness (Compton, 2013; McGuire, 1964). A large body of inoculation research across domains has demonstrated its effectiveness in conferring resistance against (unwanted) persuasion (for reviews, see Banas & Rains, 2010; Lewandowsky & van der Linden, 2021), including misinformation about climate change (Cook et al., 2017; van der Linden et al., 2017), conspiracy theories (Banas & Miller, 2013; Jolley & Douglas, 2017), and astroturfing by Russian bots (Zerback et al., 2020).

In particular, the distinction between *active* vs. *passive* defences has seen renewed interest (Banas & Rains, 2010). As opposed to traditional passive inoculation where participants receive the pre-emptive refutation, during active inoculation participants are tasked with generating their own "antibodies" (e.g., counter-arguments), which is thought to engender greater resistance (McGuire & Papageorgis, 1961). Furthermore, rather than inoculating people against specific issues, research has shown that making people aware of both their own vulnerability and the manipulative intent of others can act as a more general strategy for inducing resistance to deceptive persuasion (Sagarin et al., 2002).

Perhaps the most well-known example of active inoculation is *Bad News* (Roozenbeek & van der Linden, 2019b), an interactive fake news game where players are forewarned and exposed to weakened doses of the common techniques that are used in the production of fake news (e.g., conspiracy theories, fuelling intergroup polarization). The game simulates a social media feed and over the course of 15 to 20 minutes lets players actively generate their own "antibodies" in an interactive environment. Similar games have been developed for COVID-19 misinformation (*Go Viral!*, see Basol et al., in press), climate misinformation (*Cranky Uncle*, see Cook, 2019) and political misinformation during elections (*Harmony Square*, see Roozenbeek & van der Linden, 2020). A growing body of research has shown that after playing "fake news" inoculation games, people are; (a) better at spotting fake news, (b) more confident in their ability to identify fake news, and (c) less likely to report sharing fake news with others in their network (Basol et al., 2020; Roozenbeek, van der Linden, et al., 2020; Roozenbeek & van der Linden, 2019a, 2019b, 2020). Figure 1 shows screenshots from each game.

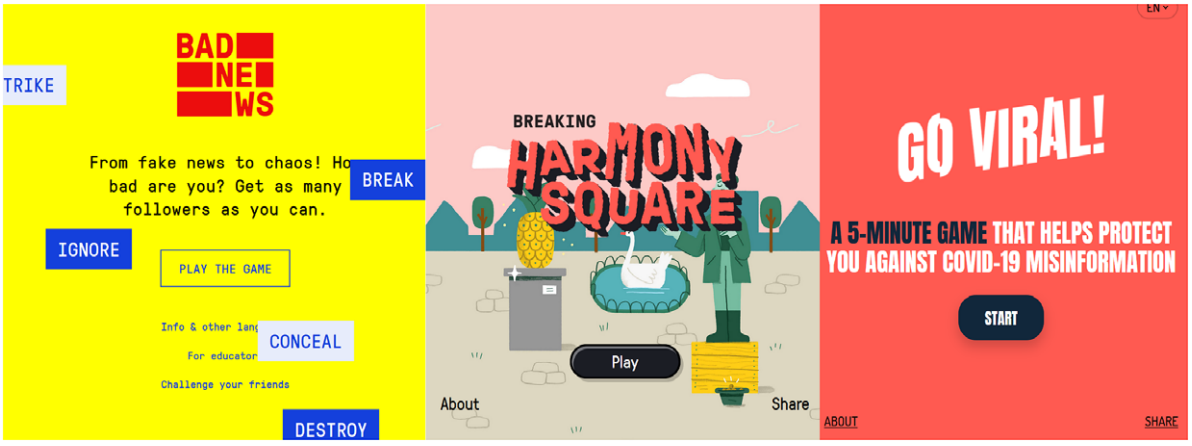

Figure 1. Screenshots of the Bad News (www.getbadnews.com), Go Viral! (www.goviralgame.com), and Harmony Square (www.harmonysquare.game) landing pages.

https://doi.org/10.1017/SJP.2021.23 Published online by Cambridge University Press

An Agenda for Future Research on Fake News Interventions

Although these advancements are promising, in this section, we outline several open questions to bear in mind when designing and testing interventions aimed at countering misinformation: How long their effectiveness remains detectable, the relevance of source effects, the role of inattention and motivated cognition, and the complexities of developing psychometrically validated instruments to measure how interventions affect susceptibility to misinformation.

The Longevity of Intervention Effects

Reflecting a broader lack of longitudinal studies in behavioural science (Hill et al., 2013; Marteau et al., 2011; Nisa et al., 2019), most research on countering misinformation does not look at effects beyond two weeks (Banas & Rains, 2010). While Swire, Berinsky, et al. (2017) found that most effects had expired one week after a debunking intervention, Guess et al. (2020) report that three weeks after a media-literacy intervention effects can either dissipate *or* endure.

Evidence from studies comparing interventions indicates that expiration rates may vary depending on the method, with *inoculation-based* effects generally staying intact for longer than *narrative*, *supportive*, or *consensus-messaging* effects (e.g., see Banas & Rains, 2010; Compton & Pfau, 2005; Maertens, Anseel, et al., 2020; Niederdeppe et al., 2015; Pfau et al., 1992). Although some studies have found inoculation effects to decay after two weeks (Compton, 2013; Pfau et al., 2009; Zerback et al., 2020), the literature is converging on an average inoculation effect that lasts for at least two weeks but largely dissipates within six weeks (Ivanov et al., 2018; Maertens, Roozenbeek, et al., 2020). Research on *booster sessions* indicates that the longevity of effects can be prolonged by repeating interventions or regular assessment (Ivanov et al., 2018; Maertens, Roozenbeek, et al., 2020; Pfau et al., 2006).

Gaining deeper insights into the longevity of different interventions, looking beyond immediate effects, and unveiling the mechanisms behind decay (e.g., *interference* and *forgetting*), will shape future research towards more enduring interventions.

Source Effects

When individuals are exposed to persuasive messages (Petty & Cacioppo, 1986; Wilson & Sherrell, 1993) and evaluate whether claims are true or false (Eagly & Chaiken, 1993), a large body of research has shown that source credibility matters (Briñol & Petty, 2009; Chaiken & Maheswaran, 1994; Maier et al., 2017; Pornpitakpan, 2004; Sternthal et al., 1978). A significant factor contributing to source credibility is similarity between the source and message receiver (Chaiken & Maheswaran, 1994; Metzger et al., 2003), particularly attitudinal (Simons et al., 1970) and ideological similarity (Marks et al., 2019).

Indeed, when readers attend to source cues, source credibility affects evaluations of online news stories (Go et al., 2014; Greer, 2003; Sterrett et al., 2019; Sundar et al., 2007) and in some cases, sources impact the believability of misinformation (Amazeen & Krishna, 2020; Walter & Tukachinsky, 2020). In general, individuals are more likely to trust claims made by ideologically congruent news sources (Gallup, 2018) and discount news from politically incongruent ones (van der Linden, Panagopoulos, et al., 2020). Furthermore, polarizing sources can boost or retract from the persuasiveness of misinformation, depending on whether or not people support the attributed source (Swire, Berinsky, et al., 2017; Swire, Ecker, et al., 2017).

For debunking, organizational sources seem more effective than individuals (van der Meer & Jin, 2020; Vraga & Bode, 2017) but only when information recipients actively assess source credibility (van Boekel et al., 2017). Indeed, source credibility may matter little when individuals do not pay attention to the source (Albarracín et al., 2017; Sparks & Rapp, 2011), and despite highly credible sources the continued influence of misinformation may persist (Ecker & Antonio, 2020). For prebunking, evidence suggests that inoculation interventions are more effective when they involve high-credibility sources (An, 2003). Yet, sources may not impact accuracy perceptions of obvious fake news (Hameleers, 2020), political misinformation (Dias et al., 2020; Jakesch et al., 2019), or fake images (Shen et al., 2019), potentially because these circumstances reduce news receivers' attention to the purported sources. Overall, relatively little remains known about how people evaluate sources of political and non-political fake news.

Inattention versus Motivated Cognition

At present, there are two dominant explanations for what drives susceptibility to and sharing of fake news. The *motivated reflection* account proposes that reasoning can increase bias. *Identity-protective* cognition occurs when people with better reasoning skills use this ability to come up with reasons to defend their ideological commitments (Kahan et al., 2007). This account is based on findings that those who have the highest levels of education (Drummond & Fischhoff, 2017), cognitive reflection (Kahan, 2013), numerical ability (Kahan et al., 2017), or political knowledge (Taber et al., 2009) tend to show more partisan bias on controversial issues.

https://doi.org/10.1017/SJP.2021.23 Published online by Cambridge University Press

The *inattention* account, on the other hand, suggests that people want to be accurate but are often not thinking about accuracy (Pennycook & Rand, 2019, 2020). This account is supported by research finding that deliberative reasoning styles (or *cognitive reflection*) are associated with better discernment between true and false news (Pennycook & Rand, 2019). Additionally, encouraging people to pause, deliberate, or think about accuracy before rating headlines (Bago et al., 2020; Fazio, 2020; Pennycook et al., 2020) can lead to more accurate identification of false news for both politically congruent and politically-incongruent headlines (Pennycook & Rand, 2019).

However, both theoretical accounts suffer from several shortcomings. First, it is difficult to disentangle whether partisan bias results from motivated reasoning or selective exposure to different (factual) beliefs (Druckman & McGrath, 2019; Tappin et al., 2020). For instance, although ideology and education might interact in a way that enhances motivated reasoning in correlational data, exposure to facts can neutralize this tendency (van der Linden et al., 2018). Paying people to produce more accurate responses to politically contentious facts also leads to less polarized responses (Berinsky, 2018; Bullock et al., 2013; Bullock & Lenz, 2019; Jakesch et al., 2019; Prior et al., 2015; see also Tucker, 2020). On the other hand, priming partisan identity-based motivations leads to increased motivated reasoning (Bayes et al., 2020; Prior et al., 2015).

Similarly, a recent re-analysis of Pennycook and Rand (2019) found that while cognitive reflection was indeed associated with better truth discernment, it was not associated with less partisan bias (Batailler et al., in press). Other work has found large effects of partisan bias on judgements of truth (see also Tucker, 2020; van Bavel & Pereira, 2018). One study found that animosity toward the opposing party was the strongest psychological predictor of sharing fake news (Osmundsen et al., 2020). Additionally, when Americans were asked for top-of-mind associations with the word "fake news," they most commonly answered with news media organizations from the opposing party (e.g., Republicans will say "CNN," and Democrats will say "Fox News"; van der Linden, Panagopoulos, et al., 2020). It is therefore clear that future research would benefit from explicating how interventions target both motivational and cognitive accounts of misinformation susceptibility.

Psychometrically Validated Measurement Instruments

To date, no psychometrically validated scale exists that measures *misinformation susceptibility* or people's *ability to discern fake from real news.* Although related scales exist, such as the *Bullshit Receptivity scale* (BSR; Pennycook et al., 2015) or the *conspiracy mentality scales* (Brotherton et al., 2013; Bruder et al., 2013; Swami et al., 2010), these are only proxies. To measure the efficacy of fake news interventions, researchers often collect (e.g., Cook et al., 2017; Guess et al., 2020; Pennycook et al., 2020; Swire, Berinsky, et al., 2017; van der Linden et al., 2017) or create (e.g., Roozenbeek, Maertens, et al., 2020; Roozenbeek & van der Linden, 2019b) news headlines and let participants rate the *reliability* or *accuracy* of these headlines on binary (e.g., *true* vs. *false*) or Likert (e.g., reliability 1-7) scales, resulting in an index assumed to depict how skilled people are at detecting misinformation. These indices are often of limited psychometric quality, and can suffer from varying reliability and specific item-set effects (Roozenbeek, Maertens, et al., 2020).

Recently, more attention has been given to the correct detection of both factual *and* false news, with some studies indicating people improving on one dimension, while not changing on the other (Guess et al., 2020; Pennycook et al., 2020; Roozenbeek, Maertens, et al., 2020). This raises questions about the role of *general scepticism*, and what constitutes a "good" outcome of misinformation interventions in a post-truth era (Lewandowsky et al., 2017). Relatedly, most methods fail to distinguish between veracity discernment and response bias (Batailler et al., in press). In addition, stimuli selection is often based on a small pool of news items, which limits the representativeness of the stimuli and thus their external validity.

A validated psychometric test that provides a general score, as well as reliable subscores for false and factual news detection (Roozenbeek, Maertens, et al., 2020), is therefore required. Future research will need to harness modern psychometrics to develop a new generation of scales based on a large and representative pool of news headlines. An example is the new Misinformation Susceptibility Test (MIST, see Maertens, Götz, et al., 2020).

Better measurement instruments combined with an informed debate on desired outcomes, should occupy a central role in the fake news intervention debate.

Implications for Policy

A number of governments and organizations have begun implementing prebunking and debunking strategies as part of their efforts to limit the spread of false information. For example, the *Foreign, Commonwealth and Development Office* and the *Cabinet Office* in the United Kingdom and the *Department of Homeland Security* in the United States have collaborated with researchers and practitioners to develop evidence-based tools to counter misinformation using inoculation theory and prebunking games that have been scaled across millions of people (Lewsey, 2020; Roozenbeek

https://doi.org/10.1017/SJP.2021.23 Published online by Cambridge University Press

& van der Linden, 2019b, 2020). Twitter has also placed inoculation messages on users' news feeds during the 2020 United States presidential election to counter the spread of political misinformation (Ingram, 2020).

With respect to debunking, Facebook collaborates with third-party fact checking agencies that flag misleading posts and issue corrections under these posts (Bode & Vraga, 2015). Similarly, Twitter uses algorithms to label dubious Tweets as misleading, disputed, or unverified (Roth & Pickles, 2020). The United Nations has launched "Verified"—a platform that builds a global base of volunteers who help debunk misinformation and spread fact-checked content (United Nations Department of Global Communications, 2020).

Despite these examples, the full potential of applying insights from psychology to tackle the spread of misinformation remains largely untapped (Lewandowsky, Smillie, et al., 2020; Lorenz-Spreen et al., 2020). Moreover, although individual-level approaches hold promise for policy, they also face limitations, including the uncertain long-term effectiveness of many interventions and limited ability to reach sub-populations most susceptible to misinformation (Nyhan, 2020; Swire, Berinsky, et al., 2017). Hence, interventions targeting consumers could be complemented with top-down approaches, such as targeting the sources of misinformation themselves, discouraging political elites from spreading misinformation through reputational sanctions (Nyhan & Reifler, 2015), or limiting the reach of posts published by sources that were flagged as dubious (Allcott et al., 2019).

Conclusion

We have illustrated the progress that psychological science has made in understanding how to counter fake news, and have laid out some of the complexities to take into account when designing and testing interventions aimed at countering misinformation. We offer some promising evidence as to how policy-makers and social media companies can help counter the spread of misinformation online, and what factors to pay attention to when doing so.

References

Albarracín, D., Kumkale, G. T., & Vento, P. P.-D. (2017). How people can become persuaded by weak messages presented by credible communicators: Not all sleeper effects are created equal. *Journal of Experimental Social Psychology*, *68*, 171–180. https://doi.org/10.1016/j.jesp.2016.06.009

Allcott, H., Gentzkow, M., & Yu, C. (2019). *Trends in the diffusion of misinformation on social media* (Working Paper 25500). *National Bureau of Economic Research*. https://doi.org/10.3386/w25500

Amazeen, M., & Krishna, A. (2020, August 7). Correcting vaccine misinformation: Recognition and effects of source type on misinformation via perceived motivations and credibility. *SSRN*. https://doi.org/10.2139/ssrn.3698102

An, C. (2003). *Efficacy of inoculation strategies in promoting resistance to political attack messages: Source credibility perspective* [Unpublished doctoral dissertation]. University of Oklahoma.

Arun, C. (2019). On WhatsApp, rumours, lynchings, and the Indian Government. *Economic & Political Weekly*, *54*(6), 30–35.

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

Banas, J. A., & Miller, G. (2013). Inducing resistance to conspiracy theory propaganda: Testing inoculation and metainoculation strategies. *Human Communication Research*, *39*(2), 184–207. https://doi.org/10.1111/hcre.12000

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, *77*(3), 281–311. https://doi.org/10.1080/03637751003758193

Basol, M., Roozenbeek, J., McClanahan, P., Berriche, M., Uenal, F., & van der Linden, S. (in press). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID–19 misinformation. *Big Data & Society*.

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, *3*(1), Article 2. https://doi.org/10.5334/joc.91

Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (in press). *A signal detection approach to understanding the identification of fake news*. *Perspectives on Psychological Science*.

Bayes, R., Druckman, J. N., Goods, A., & Molden, D. C. (2020). When and how different motives can drive motivated political reasoning. *Political Psychology*, *41*(5), 1031–1052. https://doi.org/10.1111/pops.12663

Berinsky, A. J. (2018). Telling the truth about believing the lies? Evidence for the limited prevalence of expressive survey responding. *The Journal of Politics*, *80*(1), 211–224. https://doi.org/10.1086/694258

Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *The Journal of Communication*, *65*(4), 619–638. https://doi.org/10.1111/jcom.12166

Briñol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology*, *20*(1), 49–96. https://doi.org/10.1080/10463280802643640

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in Psychology*, *4*, Article 279. https://doi.org/10.3389/fpsyg.2013.00279

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy Mentality Questionnaire. *Frontiers in Psychology*, *4*(279), Article 225. https://doi.org/10.3389/fpsyg.2013.00225

https://doi.org/10.1017/SJP.2021.23 Published online by Cambridge University Press

Bullock, J. G., Gerber, A. S., Hill, S. J., & Huber, G. A. (2013). *Partisan bias in factual beliefs about politics* (Working Paper 19080). *National Bureau of Economic Research*. https://doi.org/10.3386/w19080

Bullock, J. G., & Lenz, G. (2019). Partisan bias in surveys. *Annual Review of Political Science*, *22*, 325–342. https://doi.org/10.1146/annurev-polisci-051117-050904

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, *66*(3), 460–473. https://doi.org/10.1037//0022-3514.66.3.460

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

Compton, J. (2013). Inoculation theory. In J. P. Dillard & L. Shen (Eds.), *The Sage handbook of persuasion: Developments in theory and practice* (2nd Ed., pp. 220–237). Sage. https://doi.org/10.4135/9781452218410.n14

Compton, J. A., & Pfau, M. (2005). Inoculation theory of resistance to influence at maturity: Recent progress in theory development and application and suggestions for future research. *Annals of the International Communication Association*, *29*(1), 97–146. https://doi.org/10.1080/23808985.2005.11679045

Cook, J. (2019). *Using mobile gaming to improve resilience against climate misinformation* (American Geophysical Union: Fall Meeting Abstracts, Abstract PA13A–10). SAO/NASA Astrophysics Data System. https://ui.adsabs.harvard.edu/abs/2019AGUFMPA13A..10C

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, *12*(5), Article e0175799. https://doi.org/10.1371/journal.pone.0175799

Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review*, *1*(1). https://doi.org/10.37016/mr-2020-001

Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, *9*(2), 111–119. https://doi.org/10.1038/s41558-018-0360-1

Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences of the United States of America*, *114* (36), 9587–9592. https://doi.org/10.1073/pnas.1704882114

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.

Ecker, U. K. H., & Antonio, L. M. (2020). Can you believe it? *An investigation into the impact of retraction source credibility on the continued influence effect*. PsyArXiv. https://doi.org/10.31234/osf.io/qt4w8

Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, *6*(2), 185–192. https://doi.org/10.1016/j.jarmac.2017.01.014

Ecker, U. K. H., Lewandowsky, S., Jayawardana, K., & Mladenovic, A. (2019). Refutations of equivocal claims: No evidence for an ironic effect of counterargument number. *Journal of Applied Research in Memory and Cognition*, *8*(1), 98–107. https://doi.org/10.1016/j.jarmac.2018.07.005

Effron, D. A., & Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychological Science*, *31*(1), 75–87. https://doi.org/10.1177/0956797619887896

Fazio, L. K. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*, *1*(2). https://doi.org/10.37016/mr-2020-009

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology. General*, *144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Advances in Political Psychology*, *38*, 127–150. https://doi.org/10.1111/pops.12394

Gallup. (2018). *An online experimental platform to assess trust in the media*. https://knightfoundation.org/reports/an-online-experimental-platform-to-assess-trust-in-the-media

Go, E., Jung, E. H., & Wu, M. (2014). The effects of source cues on online news perception. *Computers in Human Behavior*, *38*, 358–367. https://doi.org/10.1016/j.chb.2014.05.044

Greer, J. D. (2003). Evaluating the credibility of online information: A test of source and advertising influence. *Mass Communication and Society*, *6*(1), 11–28. https://doi.org/10.1207/S15327825MCS0601_3

Greifeneder, R., Jaffe, M. E., Newman, E. J., & Schwarz, N. (2020). The psychology of fake news: Accepting, sharing, and correcting misinformation. *Routledge*. https://doi.org/10.4324/9780429295379

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

Hameleers, M. (2020). Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication and Society*. https://doi.org/10.1080/1369118X.2020.1764603

Hill, S. J., Lo, J., Vavreck, L., & Zaller, J. (2013). How quickly we forget: The duration of persuasion effects from mass communication. *Political Communication*, *30*(4), 521–547. https://doi.org/10.1080/10584609.2013.828143

Ingram, D. (2020, October 26). *Twitter launches "pre-bunks" to get ahead of voting misinformation*. NBC News. https://www.nbcnews.com/tech/tech-news/twitter-launches-pre-bunks-get-ahead-voting-misinformation-n1244777

https://doi.org/10.1017/SJP.2021.23 Published online by Cambridge University Press

Ivanov, B., Parker, K. A., & Dillingham, L. L. (2018). Testing the limits of inoculation-generated resistance. *Western Journal of Speech Communication*, *82*(5), 648–665. https://doi.org/10.1080/10570314.2018.1454600

Jakesch, M., Koren, M., Evtushenko, A., & Naaman, M. (2019). The role of source, headline and expressive responding in political news evaluation. *SSRN*. https://doi.org/10.2139/ssrn.3306403

Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology*, *47*(8), 459–469. https://doi.org/10.1111/jasp.12453

Jolley, D., & Paterson, J. L. (2020). Pylons ablaze: Examining the role of 5G COVID–19 conspiracy beliefs and support for violence. *The British Journal of Social Psychology*, *59*(3), 628–640. https://doi.org/10.1111/bjso.12394

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 407–424.

Kahan, D. M., Braman, D., Gastil, J., Slovic, P., & Mertz, C. K. (2007). Culture and identity-protective cognition: Explaining the white-male effect in risk perception. *Journal of Empirical Legal Studies*, *4*(3), 465–505. https://doi.org/10.1111/j.1740-1461.2007.00097.x

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54–86. https://doi.org/10.1017/bpp.2016.2

Lazer, D. M. J., 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*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

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. *Databrary*. https://doi.org/10.17910/b7.1182

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lewandowsky, S., Smillie, L., Garcia, D., Hertwig, R., Weatherall, J., Egidy, S., Robertson, R. E., O'Connor, C., Kozyreva, A., Lorenz-Spreen, P., Blaschke, Y., & Leiser, M. R. (2020). Technology and democracy: Understanding the influence of online technologies on political behaviour and decision-making. *Publications Office of the European Union*. https://doi.org/10.2760/709177

Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. European Review of Social Psychology. Advance online publication. https://doi.org/10.1080/10463283.2021.1876983

Lewsey, F. (2020). *Cambridge game "pre-bunks" COVID–19 conspiracies as part of the UK government's fight against fake news*. University of Cambridge. https://www.cam.ac.uk/stories/goviral

Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C. R., & Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature Human Behaviour*, *4*(11), 1102–1109. https://doi.org/10.1038/s41562-020-0889-7

MacFarlane, D., Tay, L. Q., Hurlstone, M. J., & Ecker, U. K. H. (2020). Refuting spurious COVID–19 treatment claims reduces demand and misinformation sharing. *PsyArXiv*. http://doi.org/10.31234/osf.io/q3mkd

Maertens, R., Anseel, F., & van der Linden, S. (2020). Combatting climate change misinformation: Evidence for longevity of inoculation and consensus messaging effects. *Journal of Environmental Psychology*, *70*, Article 101455. https://doi.org/10.1016/j.jenvp.2020.101455

Maertens, R., Götz, F. M., Schneider, C., Roozenbeek, J., Kerr, J., Stieger, S., McClanahan, W. P., Drabot, K., & van der Linden, S. (2021). *The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment [Paper presentation]*. Society for Personality and Social Psychology (SPSP) Annual Convention 2021. Virtual Convention.

Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2020). *Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments*. Journal of Experimental Psychology: Applied. Advance online publication. https://doi.org/10.1037/xap0000315

Maier, M., Adam, S., & Maier, J. (2017). Does the messenger matter? A comparison of the effects of Eurosceptic messages communicated by mainstream and radical right-wing parties on citizens' EU attitudes. *Journal of Elections, Public Opinion and Parties*, *27*(3), 330–349. https://doi.org/10.1080/17457289.2016.1273227

Marks, J., Copland, E., Loh, E., Sunstein, C. R., & Sharot, T. (2019). Epistemic spillovers: Learning others' political views reduces the ability to assess and use their expertise in nonpolitical domains. *Cognition*, *188*, 74–84. https://doi.org/10.1016/j.cognition.2018.10.003

Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: Can nudging improve population health? *BMJ*, *342*, Article d228. https://doi.org/10.1136/bmj.d228

McGuire, W. J. (1964). Inducing resistance to persuasion: Some contemporary approaches. *Advances in Experimental Social Psychology*, *1*, 191–229. https://doi.org/10.1016/S0065-2601(08)60052-0

McGuire, W. J. (1970). Vaccine for brainwash. *Psychology Today*, *3*(9), 36–39.

McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *Journal of Abnormal and Social Psychology*, *62*(2), 327–337. https://doi.org/10.1037/h0042026

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & Mccann, R. M. (2003). Credibility for the 21st Century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Annals of the International Communication Association*, *27*(1), 293–335. https://doi.org/10.1080/23808985.2003.11679029

Niederdeppe, J., Heley, K., & Barry, C. L. (2015). Inoculation and narrative strategies in competitive framing of three health policy issues. *The Journal of Communication*, *65*(5), 838–862. https://doi.org/10.1111/jcom.12162

Nisa, C. F., Bélanger, J. J., Schumpe, B. M., & Faller, D. G. (2019). Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. *Nature Communications*, *10*(1), Article 4545. https://doi.org/10.1038/s41467-019-12457-2

Nyhan, B. (2020). Facts and myths about misperceptions. *The Journal of Economic Perspectives*, *34*(3), 220–236. https://doi.org/10.1257/jep.34.3.220

Nyhan, B., & Reifler, J. (2015). The effect of fact-checking on elites: A field experiment on U.S. state legislators. *American Journal of Political Science*, *59*(3), 628–640. https://doi.org/10.1111/ajps.12162

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2020). Partisan polarization is the primary psychological motivation behind "fake news" sharing on Twitter. *PsyArXiv*. https://doi.org/10.31234/osf.io/v45bk

Pennycook, G., Cannon, T., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

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.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID–19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780. https://doi.org/10.1177/0956797620939054

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. https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2020). The cognitive science of fake news. *PsyArXiv*. https://doi.org/10.31234/osf.io/ar96c

Petersen, A. M., Vincent, E. M., & Westerling, A. L. (2019). Discrepancy in scientific authority and media visibility of climate change scientists and contrarians. *Nature Communications*, *10*(1), Article 3502. https://doi.org/10.1038/s41467-019-09959-4

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In R. E. Petty & J. T. Cacioppo (Eds.), *Communication and persuasion: Central and peripheral routes to attitude change. Springer Series in Social Psychology* (pp. 1–24). Springer. https://doi.org/10.1007/978-1-4612-4964-1_1

Pfau, M., Compton, J., Parker, K. A., An, C., Wittenberg, E. M., Ferguson, M., Horton, H., & Malyshev, Y. (2006). The conundrum of the timing of counterarguing effects in resistance: Strategies to boost the persistence of counterarguing output. *Communication Quarterly*, *54*(2), 143–156. https://doi.org/10.1080/01463370600650845

Pfau, M., Semmler, S. M., Deatrick, L., Mason, A., Nisbett, G., Lane, L., Craig, E., Underhill, J., & Banas, J. (2009). Nuances about the role and impact of affect in inoculation. *Communication Monographs*, *76*(1), 73–98. https://doi.org/10.1080/03637750802378807

Pfau, M., van Bockern, S., & Kang, J. G. (1992). Use of inoculation to promote resistance to smoking initiation among adolescents. *Communication Monographs*, *59*(3), 213–230. https://doi.org/10.1080/03637759209376266

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Prior, M., Sood, G., & Khanna, K. (2015). You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science*, *10*(4), 489–518. https://doi.org/10.1561/100.00014127

Roozenbeek, J., Maertens, R., McClanahan, W., & van der Linden, S. (2020). Disentangling item and testing effects in inoculation research on online misinformation: Solomon revisited. *Educational and Psychological Measurement*. Advance online publication. https://doi.org/10.1177/0013164420940378

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID–19 around the world. *Royal Society Open Science*, *7*(10), Article 201199. https://doi.org/10.1098/rsos.201199

Roozenbeek, J., & van der Linden, S. (2019a). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*, *22*(5), 570–580. https://doi.org/10.1080/13669877.2018.1443491

Roozenbeek, J., & van der Linden, S. (2019b). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), Article 65. https://doi.org/10.1057/s41599-019-0279-9

Roozenbeek, J., & van der Linden, S. (2020). Breaking Harmony Square: A game that "inoculates" against political misinformation. *The Harvard Kennedy School Misinformation Review*, *1*(8). https://doi.org/10.37016/mr-2020-47

Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on the psychological theory of "inoculation" can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*, *1*(2). http://doi.org/10.37016//mr-2020-008

Roth, Y., & Pickles, N. (2020, May *11*). Updating our approach to misleading information. Twitter Blog. https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html

Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology*, *83*(3), 526–541. https://doi.org/10.1037//0022-3514.83.3.526

https://doi.org/10.1017/SJP.2021.23 Published online by Cambridge University Press

Shen, C., Kasra, M., Pan, W., Bassett, G. A., Malloch, Y., & O'Brien, J. F. (2019). Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. *New Media & Society*, *21*(2), 438–463. https://doi.org/10.1177/1461444818799526

Simons, H. W., Berkowitz, N. N., & Moyer, R. J. (1970). Similarity, credibility, and attitude change: A review and a theory. *Psychological Bulletin*, *73*(1), 1–16. https://doi.org/10.1037/h0028429

Sparks, J. R., & Rapp, D. N. (2011). Readers' reliance on source credibility in the service of comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(1), 230–247. https://doi.org/10.1037/a0021331

Sternthal, B., Dholakia, R., & Leavitt, C. (1978). The persuasive effect of source credibility: Tests of cognitive response. *Journal of Consumer Research*, *4*(4), 252–260. https://doi.org/10.1086/208704

Sterrett, D., Malato, D., Benz, J., Kantor, L., Tompson, T., Rosenstiel, T., Sonderman, J., & Loker, K. (2019). Who shared it?: Deciding what news to trust on social media. *Digital Journalism*, *7*(6), 783–801. https://doi.org/10.1080/21670811.2019.1623702

Sundar, S. S., Knobloch-Westerwick, S., & Hastall, M. R. (2007). News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology*, *58*(3), 366–378. https://doi.org/10.1002/asi.20511

Swami, V., Chamorro-Premuzic, T., & Furnham, A. (2010). Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs. *Applied Cognitive Psychology*, *24*(6), 749–761. https://doi.org/10.1002/acp.1583

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), Article 160802. https://doi.org/10.1098/rsos.160802

Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(12), 1948–1961. https://doi.org/10.1037/xlm0000422

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition*, *9*(3), 286–299. https://doi.org/10.1016/j.jarmac.2020.06.006

Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, *31*(2), 137–155. https://doi.org/10.1007/s11109-008-9075-8

Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Rethinking the link between cognitive sophistication and politically motivated reasoning. *Journal of Experimental Psychology: General*. Advance online publication. https://doi.org/10.1037/xge0000974

Tucker, J. A. (2020, October 21). The truth about fake news: Measuring vulnerability to fake news online. *Digital Life Seminar @Cornell Tech*. https://www.dli.tech.cornell.edu/seminars/The-Truth-About-Fake-News%3A-Measuring-Vulnerability-to-Fake-News-Online

United Nations Department of Global Communications. (2020, May 28). *"Verified" initiative aims to flood digital space with facts amid COVID–19 crisis*. United Nations. https://www.un.org/en/coronavirus/%E2%80%98verified%E2%80%99-initiative-aims-flood-digital-space-facts-amid-covid-19-crisis

van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political psychology in the digital (mis) information age: A model of news belief and sharing. *Social Issues and Policy Review*, *15*(1), 84–113. https://doi.org/10.1111/sipr.1207

van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

van Boekel, M., Lassonde, K. A., O'Brien, E. J., & Kendeou, P. (2017). Source credibility and the processing of refutation texts. *Memory & Cognition*, *45*(1), 168–181. https://doi.org/10.3758/s13421-016-0649-0

van der Linden, S., Leiserowitz, A., & Maibach, E. (2018). Scientific agreement can neutralize politicization of facts. *Nature Human Behaviour*, *2*(1), 2–3. https://doi.org/10.1038/s41562-017-0259-2

van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), Article 1600008. https://doi.org/10.1002/gch2.201600008

van der Linden, S., Panagopoulos, C., & Roozenbeek, J. (2020). You are fake news: Political bias in perceptions of fake news. *Media Culture & Society*, *42*(3), 460–470. https://doi.org/10.1177/0163443720906992

van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID–19. *Frontiers in Psychology*, *11*, Article 566790. https://doi.org/10.3389/fpsyg.2020.566790

van der Meer, T. G. L. A., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication*, *35*(5), 560–575. https://doi.org/10.1080/10410236.2019.1573295

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, *39*(5), 621–645. https://doi.org/10.1177/1075547017731776

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research*, *47*(2), 155–177. https://doi.org/10.1177/0093650219854600

Wilson, E. J., & Sherrell, D. L. (1993). Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal of the Academy of Marketing Science*, *21*(2), Article 101. https://doi.org/10.1007/bf02894421

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

Zerback, T., Töpfl, F., & Knöpfle, M. (2020). The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them. *New Media & Society*. Advance online publication. https://doi.org/10.1177/1461444820908530

https://doi.org/10.1017/SJP.2021.23 Published online by Cambridge University Press