March 2024

EVIDENCE-BASED TECHNIQUES FOR COUNTERING MIS-/DIS-/MAL-INFORMATION

A PRIMER

Megan K. McBride, Pamela G. Faber, Kaia Haney, Patricia J. Kannapel, and Samuel Plapinger

With contributions by Heather M. K. Wolters

DIM-2023-U-035081-Final

Abstract

The Office of Naval Research (ONR) is keenly aware of the uptick of mis-/dis-/mal-information (MDM) activity since 2018 and recognizes a critical need for the US government to design and implement a program to protect US servicemembers from the potentially malign influence of MDM campaigns. ONR is also aware of the increasingly robust body of research on this topic and recognizes the benefit of aggregating such evidence-based research to inform the development of a program to protect US servicemembers. To support ONR's objectives, we offer a plain-language explanation of the evidence-based research on four types of counter-MDM interventions: inoculation, debunking, fact-checking, and media literacy. More specifically, we discuss the origins and logic of each intervention, summarize overall research findings, identify issues of ongoing analysis, and discuss how long the effects of each intervention typically last. We completed this review in support of a broader project whose goal is to recommend a single intervention (or suite of interventions) that the US government might adopt to protect servicemembers from malign foreign influence.

This document contains the best opinion of CNA at the time of issue. The views, opinions, and findings contained in this report should not be construed as representing the official position of the Department of the Navy.

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited. 3/4/2024

This work was created in the performance of Federal Government Contract Number N00014-22-D-7001.

Cover image: CNA

Approved by: **May 2023**

Maria Kingsley, Acting Research Program Director
Countering Threats and Challenges Program
Strategy, Policy, Plans, and Programs Division

Executive Summary

US allies and adversaries have used tools of persuasion and influence throughout the 20th century, and adversary attempts to use persuasion and influence to harm the US and its allies have been considered a national security priority for decades. In this context, the US has long worried about foreign efforts to use persuasion and influence against US servicemembers. Though concern about this issue waned with the end of the Cold War, the need to harness research to actively protect US servicemembers from malign persuasion and influence—from mis-/dis-/mal-information (MDM)—has once again become pressing. According to a May 2020 survey directed by senior Army leadership, almost 90 percent of US Army soldiers and civilian employees had not received any information from their units regarding adversarial propaganda about COVID-19 despite the fact that both Russia and China had been circulating virus-related MDM since March 2020.[1] This lack of awareness—and lack of counter-MDM training—left servicemembers vulnerable to external influence. It also effectively ceded the battlespace to US adversaries, allowing Russia and China to act with uncontested impunity to potentially influence servicemembers in the information sphere.

In promising news, an increasingly robust body of research addresses how to effectively counter MDM. Much of this research builds on 20th-century work by social scientists, with adjustments to compensate for the reality of social media, which has seriously exacerbated the scope of this threat. In this paper, we offer a plain-language explanation of the evidence-based research on four types of counter-MDM interventions: inoculation, debunking, fact-checking, and media literacy. More specifically, we discuss the origins and logic of each intervention, summarize overall research findings, identify issues of ongoing analysis, and discuss how long the effects of each intervention typically last.

Interventions

It is important to acknowledge, at the outset, that a comprehensive human-centric approach to this challenge (versus a technology-centric approach such as using AI to identify MDM) must take into consideration the fact that this is both a psychological and social issue. As such, it is critical that we identify, design, and implement counter-MDM solutions that address both the

[1] Amy Mackinnon, “US Army Failed to Warn Troops About COVID-19 Disinformation,” *Foreign Policy*, Oct. 21, 2021, https://foreignpolicy.com/2021/10/21/us-army-covid-19-disinformation-russia-china/.

psychological vulnerabilities that make us receptive to MDM (e.g., our tendency to accept at face value content that looks official), and the social structures that make organizations vulnerable to the spread of MDM (e.g., our tendency to believe content from authoritative figures in hierarchical organizations). This work focuses on the former, but work on the latter is equally important.

Note that the interventions we describe are not designed to change people's strongly held beliefs—or even people's lightly held opinions. This work simply aims to protect people from being manipulated by the systems and actors trying to circumvent their ability to engage in reasoned and critical thinking.

We also note that protecting oneself from MDM is more complicated than it might seem. Being savvy about the media landscape is not enough, nor is knowing that you are being exposed to MDM. This content works by exploiting normal psychological mechanisms that people use in their day-to-day lives.[2] As an analogy, keeping your front door locked at night is a great first step in protecting your home, but it won't stop a burglar who breaks in through your dryer vent (i.e., something you didn't think of as a vulnerability). In the same way, being intelligent, thoughtful, and critical—and even recognizing MDM in your newsfeed—is not adequate protection against MDM because this type of content circumvents normal defenses. The training interventions outlined in this literature are designed to bolster defenses, including those at the metaphorical front and back doors (which may be strong but not strong enough) and those at the dryer vent and heat exhaust (which may not yet exist).

Inoculation

Inoculation *is the practice of exposing individuals to persuasive messages containing weakened arguments that threaten an attitude or belief in order to "inoculate" them against stronger persuasive messages and attacks on this attitude or belief in the future. Inoculation builds resilience to manipulation (Table 1).*

Table 1. Inoculation key findings

Inoculation is an effective way to increase resistance to persuasion and manipulation.
• Inoculation works if people: ○ have imperfect knowledge of a topic ○ have imperfect knowledge of the techniques of manipulation ○ care that they are being manipulated

[2] Heather Wolters, Kasey Stricklin, Neil Carey, and Megan K. McBride, *The Psychology of (Dis) information: A Primer on Key Psychological Mechanisms*, CNA, 2021.

Inoculation is an effective way to increase resistance to persuasion and manipulation.

- Inoculations can be designed to:
  - target MDM on a specific topic
  - target the techniques used by the creators of MDM
- Inoculations may be more effective when they actively engage the user
- Inoculations can be given before or after exposure to MDM (i.e., prophylactic vs. therapeutic inoculation)
- Inoculations that cite consensus information may be more effective
- Inoculation is a potentially useful as a component of a training program designed to teach US servicemembers how to protect themselves from MDM

Source: CNA.

At its core, the theory of inoculation holds that a "vaccine" consists of two parts: a threat or forewarning of MDM and a refutational preemption (sometimes referred to as a "prebunk"). Through inoculation, the individual is forewarned of or threatened by a counter-attitudinal "attack" (a term used to describe a counterargument) that motivates resistance, and then they receive the skills or information to refute the counterargument. Studies and experiments have found that inoculation theory effectively neutralizes and builds resistance to MDM. According to one article, "Over the last 50 years, a large body of evidence across domains—from health to political campaigning—has revealed that inoculation messages can be effective at conferring resistance to persuasion."[3] This general consensus is borne out by the majority of recently published literature.

Debunking

***Debunking** is the use of a concise correction to MDM that demonstrates that the prior message or messaging campaign was inaccurate (Table 2).*

Table 2. Debunking key findings

Debunking is an effective way to reduce belief in MDM accuracy.

- Debunking can correct specific instances of inaccurate information, but it cannot protect people from influence in general
- Debunking messages appear to be more effective when they:
  - cite high-credibility sources (i.e., sources that have expertise and that are trustworthy)
  - contain detailed corrective information, which is more effective than simple corrections
  - express stronger corrections (e.g., those containing more information)

[3] Stephan Lewandowsky and Sander Van Der Linden, "Countering Misinformation and Fake News Through Inoculation and Prebunking," *European Review of Social Psychology* 32, no. 2 (2021).

Debunking is an effective way to reduce belief in MDM accuracy.
• The tone of the correction (e.g., uncivil, neutral, affirmational) does not appear to change the effect of the correction
• The format of the correction (e.g., truth first, myth first) does not appear to change the effect of the correction

Source: CNA.

The logic of debunking is relatively straightforward: it involves the targeted provision of correct information in response to incorrect information. In this respect, debunking is primarily a therapeutic intervention that responds directly to MDM after it has been circulated, but it can also be a quasi-prophylactic intervention when the correction alerts recipients to specific bad actors or sources who are likely to spread MDM. A common finding of debunking research is that corrections succeed in reducing belief in the accuracy of MDM, or as one debunking expert we consulted put it: "Corrections are wildly effective."[4] Overall, research indicates that MDM debunking, when well executed, is an effective tool for countering MDM. In fact, the limited cross-comparative studies comparing inoculation (a.k.a., "prebunking") and debunking found debunking to be even more effective, although both reduced MDM reliance.[5]

Fact-checking

Fact-checking *is a journalistic practice designed to reject clearly false claims with empirical evidence from neutral or unimpeachable sources (Table 3).*[6]

Table 3. Fact-checking key findings

Fact-checking is an effective way to reduce belief in MDM accuracy.
• Fact-checking can correct specific instances of inaccurate information, but it cannot protect people from influence in general
• Fact-checking is best when integrated into the consumption of news
• Fact-checking is a potentially powerful tool for DOD personnel with communications responsibilities

Source: CNA.

Debunking and fact-checking are quite similar, but fact-checking is primarily employed by journalists and newsrooms, is typically impartial (to the extent that it adheres to journalistic

[4] Interview with Dr. Briony Swire-Thompson, Dec. 5, 2022.

[5] Li Qian Tay et al., "A Comparison of Prebunking and Debunking Interventions for Implied Versus Explicit Misinformation," *British Journal of Psychology* 113, no. 3 (2022), doi: 10.1111/bjop.12551, NLM.

[6] Interview with fact-checking subject matter expert, Dec. 1, 2022.

standards of accuracy), and aims to correct all falsehoods within a given context (e.g., a political debate). The general logic of fact-checking is relatively straightforward: (1) an individual is presented with false information, (2) the individual is presented with subsequent information that corrects the initial information, and (3) the individual updates their belief about the accuracy of the original information to more closely align with the factual information. Fact-checking is, as a result, a therapeutic intervention that directly responds to a specific piece of false information. At present, the findings on fact-checking are slightly mixed, and a recent meta-analysis determined that the overall effectiveness of fact-checking seems to be contingent upon a range of issues, including the political sophistication of the individual reading the fact-check, the nature of the message, and the preexisting beliefs of the individual.

Media literacy

Media literacy *is an individual's ability to critically assess a piece of content. It includes the skills required to evaluate a piece of content, as well as an understanding of the structures that produced that content (Table 4).*

Table 4. Media literacy key findings

Media literacy is an effective way to increase resistance to persuasion and manipulation.

- In-person media literacy training has been found to be effective across a range of topics, behaviors, and outcomes
- Online media literacy training has been shown to positively affect media use in multiple ways:
  - increase trust in media
  - increase the ability to differentiate real from fake headlines
  - lower people's belief that MDM is accurate
- Online *news* media literacy training may be limited in its ability to counter MDM, but it has been shown to:
  - improve self-perceptions of media literacy
  - effectively reinforce lessons learned from in-person trainings
  - improve the quality of the news that people share online

Source: CNA.

Media literacy can be thought of as a process or set of skills based on critical thinking.[7] Media literate individuals have the tools and abilities necessary to critically evaluate a piece of "media," whether that be a tweet, an article, a TV show, or other content. Media literacy training teaches a range of skills including, but not limited to, asking questions, analyzing sources,

[7] Monica Bulger and Patricia Davison, "The Promises, Challenges, and Futures of Media Literacy," *Journal of Media Literacy Education* 10 (2018).

assessing bias, and valuing the role of an independent media. In this sense, media literacy is content neutral: it does not advance or counter specific ideas but rather teaches wholly nonpartisan skills. Media literacy, like inoculation theory, is a preventative or prophylactic intervention, and its original goal was to help protect young people from negative media exposure effects. As described by some scholars, media literacy trainings and interventions are examples of "logic-based inoculations."[8] In these framings, media literacy education and critical-thinking interventions become types of inoculation. Not all scholars call their media literacy programs inoculations, but many cite well-known scholars and specific tenets from the inoculation theory literature. For example, Vraga and Tully (2015) note that a media literacy public service announcement may serve a similar role as a booster in an inoculation campaign by reminding people of things they learned in earlier media literacy education.[9] Because media literacy interventions were found to be increasingly effective outside of classrooms, against both inaccurate headlines and biased media, they emerged as a promising preemptive intervention that might (similar to, but distinct from, inoculation efforts) combat MDM.

Potential concerns

In reviewing the literature on MDM interventions, we identified three concerns that scholars repeatedly raised: the backfire effect, the continued influence effect, and news cynicism. In broad terms, the *backfire effect* is a worry that counter-MDM interventions and trainings will result in unforeseen consequences. However, research shows that worries about the backfire effect are likely overwrought. Although earlier studies detected backfire effects,[10] recent research has found little evidence of these effects, and studies have been unable to show that

[8] Emily K. Vraga, Sojung Claire Kim, and John Cook, "Testing Logic-Based and Humor-Based Corrections for Science, Health, and Political Misinformation on Social Media*," Journal of Broadcasting & Electronic Media* 63, no. 3 (2019): 393-414, doi: 10.1080/08838151.2019.1653102.

[9] Emily K. Vraga and Melissa Tully, "Media Literacy Messages and Hostile Media Perceptions: Processing of Nonpartisan Versus Partisan Political Information," *Mass Communication and Society* 18, no. 4 (2015): 422-448, doi: 10.1080/15205436.2014.1001910.

[10] Brendan Nyhan and Jason Reifler, "When Corrections Fail: The Persistence of Political Misperceptions," *Political Behavior* 32, no. 2 (2010): 303-330; B. Nyhan, J. Reifler, and P. A. Ubel, "The Hazards of Correcting Myths About Health Care Reform," *Medical Care* 51, no. 2 (2013): 127-132, doi: 10.1097/MLR.0b013e318279486b, NLM; Dino P. Christenson, Sarah E. Kreps, and Douglas L. Kriner, "Contemporary Presidency: Going Public in an Era of Social Media: Tweets, Corrections, and Public Opinion," *Presidential Studies Quarterly* 51, no. 1 (2021): 151-165, doi: https://doi.org/10.1111/psq.12687.

they occur under only certain conditions.[11] The *continued influence effect* is a worry that MDM cannot be truly eliminated but will continue to exert an influence even after an intervention or training. Research suggests that this concern is legitimate; however, given the nature of the continued influence effect (i.e., a failure to *fully* eliminate the influence of MDM), this issue is not a reason to avoid counter-MDM trainings or interventions. The third concern, news cynicism, is slightly more complicated. As the research highlights, counter-MDM interventions and training might increase cynicism related to real news. And yet, numerous experts point out that this outcome may not necessarily be bad. Certainly, people doubting the veracity of all information would be a negative outcome, but people approaching all headlines (from both partisan and nonpartisan sites) with a critical eye may be socially healthy.

Conclusion

We completed this review in support of a broader project whose goal is to recommend a single intervention (or suite of interventions) that the US government might adopt to protect servicemembers from malign foreign influence. Although it stands alone as a useful primer for those hoping to understand the state of research on these issues, more specific guidance—in the form of an assessment of applicability to military populations, a list of best practices, and recommendations for near-term implementation—can be found in the companion report: *Protecting Servicemembers from Foreign Influence: A Counter-MDM Toolkit*.

[11] R. Kelly Garrett, Erik C. Nisbet, and Emily K. Lynch, "Undermining the Corrective Effects of Media-Based Political Fact Checking? The Role of Contextual Cues and Naïve Theory," *Journal of Communication* 63, no. 4 (2013): 617–637, doi: https://doi.org/10.1111/jcom.12038; Thomas Wood and Ethan Porter, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence," *Political Behavior* 41 (2019): 135–163, doi: 10.1007/s11109-018-9443-y; Ethan Porter and Thomas J. Wood, "Political Misinformation and Factual Corrections on the Facebook News Feed: Experimental Evidence," *Journal of Politics* 84, no. 3 (2022): 1812-1817, doi: 10.1086/719271.

This page intentionally left blank.

Contents

Introduction

US allies and adversaries have used tools of persuasion and influence throughout the 20th century, and adversary attempts to use persuasion and influence to harm the US have been considered a national security priority for decades. In this context, the US has long been worried about foreign efforts to use persuasion and influence against US servicemembers. Concern about this issue was heightened during the Cold War, when US officials worried about the "brainwashing" of US prisoners of war. During this period, the US Department of Defense (DOD) and US intelligence agencies funded research into both mind control and mind resistance techniques—the most famous of which were the Central Intelligence Agency's (CIA's) MK-ULTRA experiments carried out from the 1950s to the 1970s.[12]

A subset of 21st-century research on how to counter the influence of mis-/dis-/mal-information (MDM) traces directly to these early US government–funded efforts. Inoculation theory was initially developed in the early 1960s by social scientist William McGuire.[13] And though no direct evidence indicates that McGuire was funded by DOD or connected to the MK-ULTRA experiments, he may have been one of the many unwitting researchers unaware of who was paying their bills.

McGuire's research remains compelling because it was relatively mainstream (in contrast to the unethical work of MK-ULTRA), and his theory has experienced consistent expansion since the 1960s. Today, inoculation theory stands alongside others—debunking, fact-checking, and media literacy—as a promising technique for countering MDM.

The need to operationalize this type of work—that is, to harness research to actively protect US servicemembers from malign persuasion and influence—has once again become pressing. According to a May 2020 survey directed by senior Army leadership, almost 90 percent of US Army soldiers and civilian employees had not received any information from their units about adversarial COVID-19 propaganda, despite the fact that both Russia and China had been

[12] US Senate, The Select Committee on Intelligence and the Subcommittee on Health and Scientific Research of the Committee on Human Resources, *Joint Hearing on Project MKULTRA, the CIA'S Program of Research in Behavioral Modification*, 95 Cong., 1st sess., Aug. 3, 1977, https://www.intelligence.senate.gov/sites/default/files/hearings/95mkultra.pdf.

[13] William J. McGuire, "Inducing Resistance to Persuasion. Some Contemporary Approaches," in *Self and Society. An Anthology of Readings*, ed. C. C. Haaland and W. O. Kaelber (Lexington, MA: Ginn Custom Publishing, 1964).

circulating virus-related MDM since March 2020.[14] This lack of awareness—and lack of counter-MDM training—left servicemembers vulnerable to external influence. It also effectively ceded the battlespace to US adversaries, allowing Russia and China to act with uncontested impunity in the information sphere and potentially influence servicemembers.

In promising news, an increasingly robust body of research has explored how to effectively counter MDM. Much of this research builds on 20th-century work, with adjustments and changes to compensate for the reality of social media, which has seriously exacerbated the scope of this threat.

Note that this research—and the interventions described in this review—is not designed to change people's strongly held positions, or even people's lightly held opinions. In fact, research suggests that these interventions don't change general political views, attitudes, and voting preferences, though they may change beliefs about the accuracy of MDM. The goal is narrow and specific: help people sift the true from the false, and protect people from being manipulated by systems and actors aspiring to circumvent their ability to engage in reasoned and critical thinking.

We also note that protecting oneself from MDM is more complicated than it might seem. Being savvy about the media landscape is not sufficient, nor is knowing that you are being exposed to MDM. This content works by exploiting normal psychological mechanisms that people use in their day-to-day lives.[15] As an analogy, keeping your front door locked at night is a great first step in protecting your home, but it won't stop a burglar who breaks in through your dryer vent (i.e., something that you didn't think of as a vulnerability). In the same way, just recognizing MDM in your newsfeed is not sufficient.

There are, in fact, many ways to exploit human thinking. Table 5 outlines some common methods.

[14] Amy Mackinnon, "US Army Failed to Warn Troops About COVID-19 Disinformation," *Foreign Policy*, Oct. 21, 2021, https://foreignpolicy.com/2021/10/21/us-army-covid-19-disinformation-russia-china/.

[15] Wolters et al., *The Psychology of (Dis) information: A Primer on Key Psychological Mechanisms.*

Table 5. Methods to exploit thinking

Method	Impact
Illusory truth effect	"It's easy to be misled. Our feelings of familiarity and truth are often linked. We are more likely to believe things that we have heard many times than new information."[16]
Emotional appeal	"Misinformation is…often steeped in emotional language and designed to be attention-grabbing and have persuasive appeal. This facilitates its spread and can boost its impact, especially in the current online economy in which user attention has become a commodity."[17]
Information processing	When faced with deluge of info and insufficient time to process, people turn to heuristics, i.e., "How does this fit in with what I already believe? How does it fit in with what I know or what I think I know? How does it fit in with what other people believe—in particular trusted other people?"[18]
Cognitive dissonance	"People build mental models of the world…and they want these mental models to be *complete*. They want to understand what's going on. They don't like incomplete models, and they are willing to accept information that is maybe not very reliable or valid if that allows them to build complete models of the world so they have what feels like a complete understanding."[19]
Group, belief, and novelty	"We are more likely to share information with people we consider members of our group, when we believe the information is true, and when it is novel or urgent. If disinformation is coming from a group member with whom we identify, is consistent with our beliefs, or is new information for us, we are more likely to share it."[20]

Source: Multiple sources, see footnotes.

In short, being intelligent, thoughtful, and critical is not adequate protection against MDM because this content circumvents normal defenses. The training interventions in this literature are designed to bolster defenses, including those at the metaphorical front and back doors (which may be strong but not strong enough) and those at the dryer vent and heat exhaust (which may not yet exist).

[16] Stephan Lewandowsky et al., *The Debunking Handbook 2020*, 2020.

[17] Ibid.

[18] Ullrich K. H. Ecker, "Why Rebuttals May Not Work: The Psychology of Misinformation," *Media Asia* 44, no. 2 (2017): 79-87.

[19] Ibid.

[20] Wolters et al., *The Psychology of (Dis) information: A Primer on Key Psychological Mechanisms.*

Goals and features of the literature review

The Office of Naval Research (ONR) is keenly aware of the uptick of MDM activity in the last five years and recognizes a critical need for the US government to design and implement a program to protect US servicemembers from the potentially malign influence of MDM campaigns. ONR is also aware of the increasingly robust body of research on this topic and recognizes the benefit of aggregating such evidence-based research to inform the development of a program to protect US servicemembers. To support ONR's objectives, we offer a plain-language explanation of the evidence-based research on the four best-researched types of counter-MDM interventions: inoculation, debunking, fact-checking, and media literacy. We completed this review in support of a broader project whose goal is to recommend a single intervention (or suite of interventions) that the US government might adopt to protect servicemembers from malign foreign influence.

We intend this literature review to bridge the gap between the academic work being done now and the policy work that will (hopefully) follow. As such, two key features of the literature are worth noting.

First, because the four interventions we describe originate in a range of academic disciplines (social psychology, education, journalism, etc.), the literatures describing these interventions and testing their effectiveness are quite distinct. In writing this paper, we chose to remove as much academic jargon and discipline-specific language as possible.

Second, because this paper seeks to inform policy-makers, we deviated from the conventions of a systematic literature review. Specifically, we chose to summarize the core or overall findings of the field rather than summarizing all of the work in the field. As such, we included only research findings that were replicated and embraced by the field. We determined that a finding articulated by a single article didn't offer adequate evidence to be incorporated into a training program for US servicemembers. Such individual findings may be replicated in the future and become accepted features of the literature, at which point they should be incorporated into the literature that informs training program decisions. But given that such findings may just as likely remain unsubstantiated or be refuted, we excluded them at this stage.

Organization

This research paper has four sections devoted to the following counter-MDM techniques: inoculation, debunking, fact-checking, and media literacy. In each section, we begin with a brief history of the technique, define the technique and describe how it works, and summarize the state of research on the technique. Note that the lengths of these sections are unequal because the literatures on the four intervention types vary considerably. The nature of the literatures can be traced to a range of variables (e.g., norms of the academic fields in which the articles are published, differences of opinion within the fields, our own decisions about how to group intervention types), and the quantity of the research should not be interpreted as a proxy for quality. Following discussions of each counter-MDM technique, we discuss some concerns raised about these interventions before we offer a brief set of concluding thoughts.

Of note to the reader, you can chose to read this paper straight through or focus on individual sections. We hope that it functions as a reference manual for those who need a quick primer—or refresher—on one or more of the counter-MDM techniques it addresses.

Inoculation Theory

***Inoculation** is the practice of exposing individuals to persuasive messages containing weakened arguments that threaten an attitude or belief in order to "inoculate" them against stronger persuasive messages and attacks on this attitude or belief in the future. Inoculation builds resilience to manipulation (Table 6).*

Table 6. Inoculation key findings

Inoculation is an effective way to increase resistance to persuasion and manipulation.

- Inoculation works if people:
  - have imperfect knowledge of a topic
  - have imperfect knowledge of the techniques of manipulation
  - care that they are being manipulated
- Inoculations can be designed to:
  - target MDM on a specific topic
  - target the techniques used by the creators of MDM
- Inoculations may be more effective when they actively engage the user
- Inoculations can be given before or after exposure to MDM (i.e., prophylactic vs. therapeutic inoculation)
- Inoculations that cite consensus information may be more effective
- Inoculation is a potentially useful as a component of a training program designed to teach US servicemembers how to protect themselves from MDM

Source: CNA.

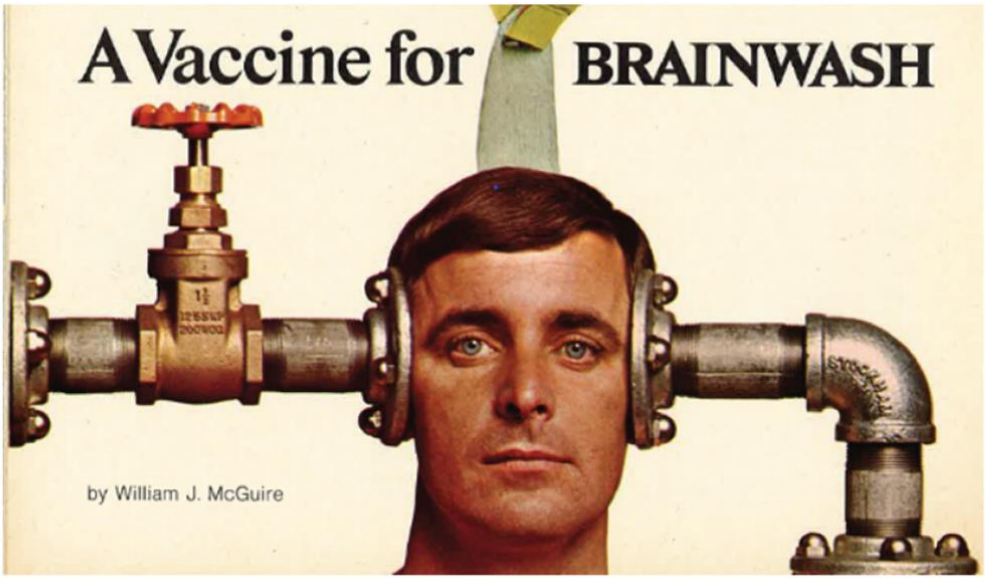
A brief history of inoculation theory

Inoculation theory can be traced to mid-20th-century concerns around the psychological manipulation and "brainwashing" of US prisoners of war in Korea. This issue was particularly worrisome and a national security priority because some US prisoners of war chose to remain with their captors after being given the opportunity to return to the US.[21] One hypothesis for why they made this decision was that US soldiers were unable to psychologically defend their

[21] "The True Story of Brainwashing and How It Shaped America," *Smithsonian Magazine*, May 22, 2017, https://www.smithsonianmag.com/history/true-story-brainwashing-and-how-it-shaped-america-180963400/.

beliefs against the strong persuasion that occurred during captivity.[22] To understand—and defend against—this problematic trend, social scientists worked to improve their understanding of persuasion. Inoculation theory, initially developed in the early 1960s by social scientist William McGuire, was part of this research effort (see Figure 1).[23]

Figure 1. Early image on inoculation theory



Source: Photo from McGuire (1970) in *Psychology Today*. Copyright held by an unknown person. Taken from Lewandowsky and van der Linden (2021).

Over the decades, inoculation theory shifted from focusing on protecting against attacks on preexisting beliefs (the original idea articulated by McGuire) to focusing on protecting against the manipulative techniques used by those who spread inaccurate information, such as MDM. As a result, interest in the theory and its application has experienced an unsurprising uptick since 2016.

In what follows, we first provide more background on inoculation, explore the state of the overall research, and end with a brief discussion of how long an individual retains the benefits of inoculation after the intervention.

[22] Elspeth Cameron Ritchie, “Psychiatry in the Korean War: Perils, PIES, and Prisoners of War,” *Military Medicine* 167, no. 11 (2002).

[23] McGuire, “Inducing Resistance to Persuasion. Some Contemporary Approaches.”

Definition and logic of inoculation theory

Inoculation theory is a psychosocial hypothesis that uses a biological metaphor as its central premise. The earliest work on inoculation theory suggested that people can be inoculated against "persuasive attacks on their attitudes" in a similar way to viral immunization.[24] Critically, inoculation is not meant to be a tool of persuasion in and of itself, but rather a tool for reducing the effectiveness of undue persuasion on populations. In this model, exposing individuals to messages containing a weakened argument against an attitude or belief can "inoculate" them against stronger attacks on this attitude or belief.[25] Moreover, much as protecting people from viruses also includes warning them that viruses are circulating (and thus encouraging people to wash their hands, etc.), the theory of inoculation argues that attitudinal resistance can be induced by forewarning an individual of an impending attack. As one example from the era, a soldier might be told to expect arguments against democracy should he be taken prisoner of war, and then exposed to weak arguments against American democracy to stimulate a defensive attitude. According to McGuire:

> We can develop belief resistance in people as we develop disease resistance in a biologically overprotected [person][26] or animal: by exposing the person to a weak dose of the attacking material, strong enough to stimulate [the person's] defenses, but not strong enough to overwhelm them.[27]

Originally, inoculation theory was applied to noncontroversial or "germ-free" beliefs (i.e., beliefs that had not been infected by any seeds of doubt), such as the value of toothbrushing. However, inoculation theory has been applied to a greatly expanding number of topics over time.[28]

Inoculation theory is dependent on two assumptions. First, people have imperfect knowledge of specific topics (e.g., climate change, vaccine safety); second, people have imperfect knowledge about how they can be manipulated. If both assumptions hold, then a population can be inoculated. If, however, a population has exacting knowledge on a topic (or is fully

[24] Ibid.

[25] John A. Banas and Stephen A. Rains, "A Meta-Analysis of Research on Inoculation Theory," *Communication Monographs* 77, no. 3 (2010).

[26] We altered this quote to remove the generic use of "man" and "his."

[27] McGuire cited in Michael Pfau, "Designing Messages for Behavioral Inoculation," in *Designing Health Messages: Approaches from Communication Theory and Public Health Practice*, (Thousand Oaks, CA: Sage Publications, 1995), doi: 10.4135/9781452233451.n6.

[28] Interview with Dr. Jon Roozenbeek, Nov. 22, 2022.

convinced that they know the truth) or does not care about being psychologically manipulated, then they are unlikely to be affected by an inoculation.[29]

At its core, inoculation theory holds that a "vaccine" consists of two parts: a threat or forewarning of misinformation, and a refutational preemption (sometimes referred to as a "prebunk").[30] The individual is forewarned or threatened[31] by a counter-attitudinal attack (a term for a counterargument) that is designed to motivate resistance, and then provided the skills or information to refute the attack. Where inoculation differs from biological vaccines, though, is in the nature of the defense. When a biological vaccine is given, the goal is to provoke the body to develop antibodies to a specific virus. In an attitudinal inoculation, the vaccine can be designed to produce different types of cognitive antibodies depending on the desired target.[32]

Issue (narrow) versus technique (broad)

One approach is issue-based inoculation, which focuses on conferring psychological resistance against misinformation on specific subject areas, such as vaccine hesitancy or climate change. In short, issue-based inoculation gives people the skills to refute future persuasive challenges about a specific issue.[33] The second approach is technique-based inoculation, which focuses on "building resistance against the rhetorical techniques and strategies that are commonly used to mislead people, such as the use of emotionally manipulative language, evoking outgroup animosity and polarisation, logical fallacies, fake experts, or conspiratorial reasoning."[34] In

[29] Ibid.

[30] In more recent literature, the "forewarning" is also described as a "threat," and the "refutational preemption" is also described as "prebunking." See Josh Compton et al., "Inoculation Theory in the Post-Truth Era: Extant Findings and New Frontiers for Contested Science, Misinformation, and Conspiracy Theories," *Social and Personality Psychology Compass* 15, no. 6 (2021).

[31] The threat is important because according to the theory, this process motivates an individual to defend or protect a desirable position they see as being at risk (see Banas and Rains, "A Meta-Analysis of Research on Inoculation Theory"). Additional scholarship has confirmed the need for a perceived threat but is unresolved on the optimal amount of perceived threat.

[32] Note that the literature inconsistently defines the stages and details of inoculation (e.g., what is included in the "vaccine" stage, where threat and forewarning fit into the process, the definitions of *refutational preemption* and *prebunk*). Although the details differ, the general flow of the inoculation process is similar. This paper combines a number of inoculation theory descriptors; as a result, it may not precisely match any individual study. However, we have taken care to combine all the relevant elements of inoculation theory across the literature to maximize understanding.

[33] Compton et al., "Inoculation Theory in the Post-Truth Era: Extant Findings and New Frontiers for Contested Science, Misinformation, and Conspiracy Theories."

[34] Jon Roozenbeek and Sander van der Linden, "How to Combat Health Misinformation: A Psychological Approach," *American Journal of Health Promotion* 36, no. 3 (2022).

technique-based inoculation, the vaccine provides information on an underlying rhetorical strategy used to mislead. According to the theory, this information gives the individual skills to refute future persuasive challenges that use this technique, no matter the topic. In both cases, the forewarning and the preemptive refutation prompt an individual to develop "cognitive antibodies" that, in theory, lead to psychological inoculation, though in different ways.[35]

Elsewhere in the literature, these approaches are described as "narrow-spectrum inoculation," which protects against a specific argument or issue, and "broad-spectrum inoculation," which protects against the techniques that underlie manipulation and persuasion.[36] Narrow-spectrum inoculations cultivate immunity against a specific topic or idea, whereas broad-spectrum inoculations train people to detect flawed argumentation styles that could be used in MDM on a range of topics. Recent research increasingly supports the latter alternative. These findings suggest that inoculation can focus less on specific issues and more broadly on training people to identify and counter underlying persuasion techniques.[37] This research is connected to the growing literature on "cross-protection," which holds that inoculation can confer resistance to yet untreated attitudes.[38] For example, if an individual is trained to recognize the underlying techniques for persuasion on the issue of climate change, they can then apply this knowledge to an unrelated topic, such as vaccine hesitancy.

Both issue-based/narrow-spectrum and technique-based/broad-spectrum inoculations have been found to be effective. Moreover, because they aspire to do different things (inoculating against specific ideas versus against specific persuasive techniques), it is not possible to say that one has more value than the other. That said, technique-based/broad-spectrum inoculations have been theoretically linked to the idea of "herd immunity." In this logic, the biological metaphor of inoculation is expanded to include the question of whether herd immunity is possible via the social diffusion of widely applicable counter-MDM skills (e.g., improved ability to recognize emotional manipulation) via widespread inoculation. The ideal outcome, herd immunity, would occur if the inoculation occurred widely enough to effectively slow (and even stop) the spread of MDM through populations. Dai et al. posited that the effect

[35] Sander van der Linden, "Misinformation: Susceptibility, Spread, and Interventions to Immunize the Public," *Nature Medicine* 28, no. 3 (2022).

[36] Lewandowsky and Linden, "Countering Misinformation and Fake News Through Inoculation and Prebunking."

[37] Ibid.

[38] Kimberly A. Parker, Stephen A. Rains, and Bobi Ivanov, "Examining the 'Blanket of Protection' Conferred by Inoculation: The Effects of Inoculation Messages on the Cross-Protection of Related Attitudes," *Communication Monographs* 83, no. 1 (2016), doi: 10.1080/03637751.2015.1030681.

of word-of-mouth can be integrated into spreading the inoculation process, though evidence is currently lacking to support this theory.[39]

Active versus passive

The scholarship also addresses the distinction between active and passive inoculations.[40] In active inoculation, individuals are responsible for developing counterarguments to weakened, controversial arguments (i.e., the "virus"). In passive inoculation (the more traditional approach), individuals are passively provided counterarguments.[41] Passive inoculation includes reading short texts or viewing an infographic or video. In recent years, the number of active inoculation activities has risen, including games such as Bad News and Go Viral!. The value of active inoculation is that participants are actively stimulated to build resistance through direct engagement.[42] Researchers have found that active inoculation can train people to be "more attuned to specific deception strategies" than they had been prior to the training.[43] Although both passive and active inoculations effectively inoculate, there is evidence that active inoculation has longer lasting benefits than passive inoculation.[44]

Refutational same versus refutational different

There are two types of refutational arguments (sometimes called "prebunks"): those present in subsequent attack messages ("refutational same"), and those not present in subsequent attack messages ("refutational different").[45] A refutational same argument might use a specific argument to inoculate people against misinformation to help them develop antibodies against this specific argument in future misinformation. A refutational same inoculation might take the form depicted in Figure 2.

[39] Yue Nancy Dai et al., "The Effects of Self-Generated and Other-Generated eWOM in Inoculating Against Misinformation," *Telematics and Informatics* 71 (2022), doi: 101835.

[40] Lewandowsky and Linden, "Countering Misinformation and Fake News Through Inoculation and Prebunking."

[41] Jon Roozenbeek and Sander Van der Linden, "Fake News Game Confers Psychological Resistance Against Online Misinformation," *Palgrave Communications* 5, no. 1 (2019).

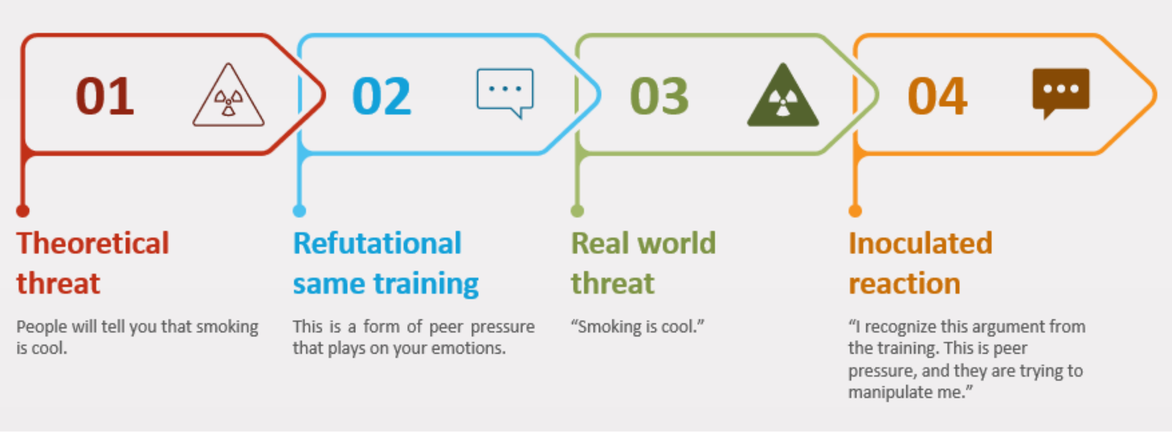
[42] Melisa Basol et al., "Towards Psychological Herd Immunity: Cross-Cultural Evidence for Two Prebunking Interventions Against COVID-19 Misinformation," *Big Data & Society* 8, no. 1 (2021).

[43] Roozenbeek and Linden, "Fake News Game Confers Psychological Resistance Against Online Misinformation."

[44] Basol et al., "Towards Psychological Herd Immunity: Cross-Cultural Evidence for Two Prebunking Interventions Against COVID-19 Misinformation."

[45] Banas and Rains, "A Meta-Analysis of Research on Inoculation Theory."

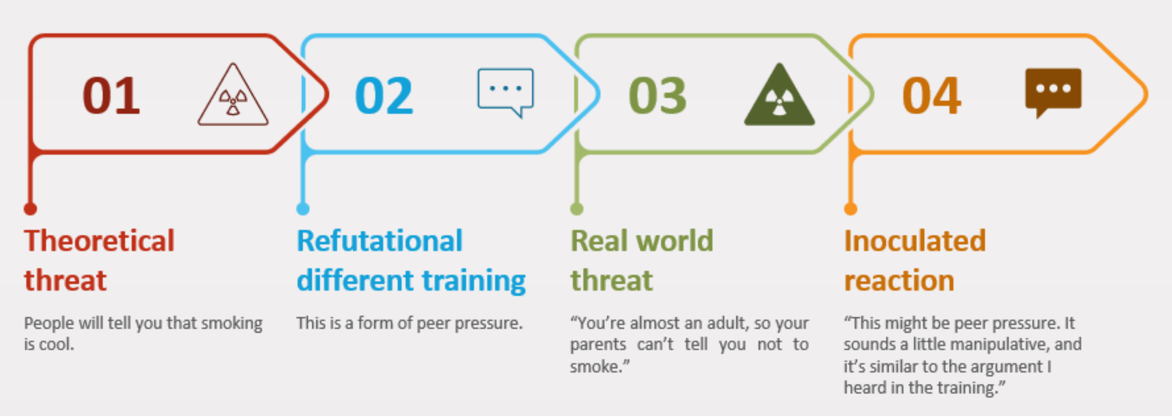
Figure 2. Refutational same messaging


01
Theoretical threat
People will tell you that smoking is cool.
02
Refutational same training
This is a form of peer pressure that plays on your emotions.
03
Real world threat
"Smoking is cool."
04
Inoculated reaction
"I recognize this argument from the training. This is peer pressure, and they are trying to manipulate me."

Source: CNA.

In contrast, a refutational different argument would use messages that are distinct from those expected in future attacks but that would (in theory) still elicit an inoculated response (Figure 3).

Figure 3. Refutational different messaging


01
Theoretical threat
People will tell you that smoking is cool.
02
Refutational different training
This is a form of peer pressure.
03
Real world threat
"You're almost an adult, so your parents can't tell you not to smoke."
04
Inoculated reaction
"This might be peer pressure. It sounds a little manipulative, and it's similar to the argument I heard in the training."

Source: CNA.

Researchers hold different views on the effectiveness of refutational same and different messages, and on how they are expressed, because it is not yet clear whether inoculation works better when people face the same or novel arguments.

Therapeutic versus prophylactic

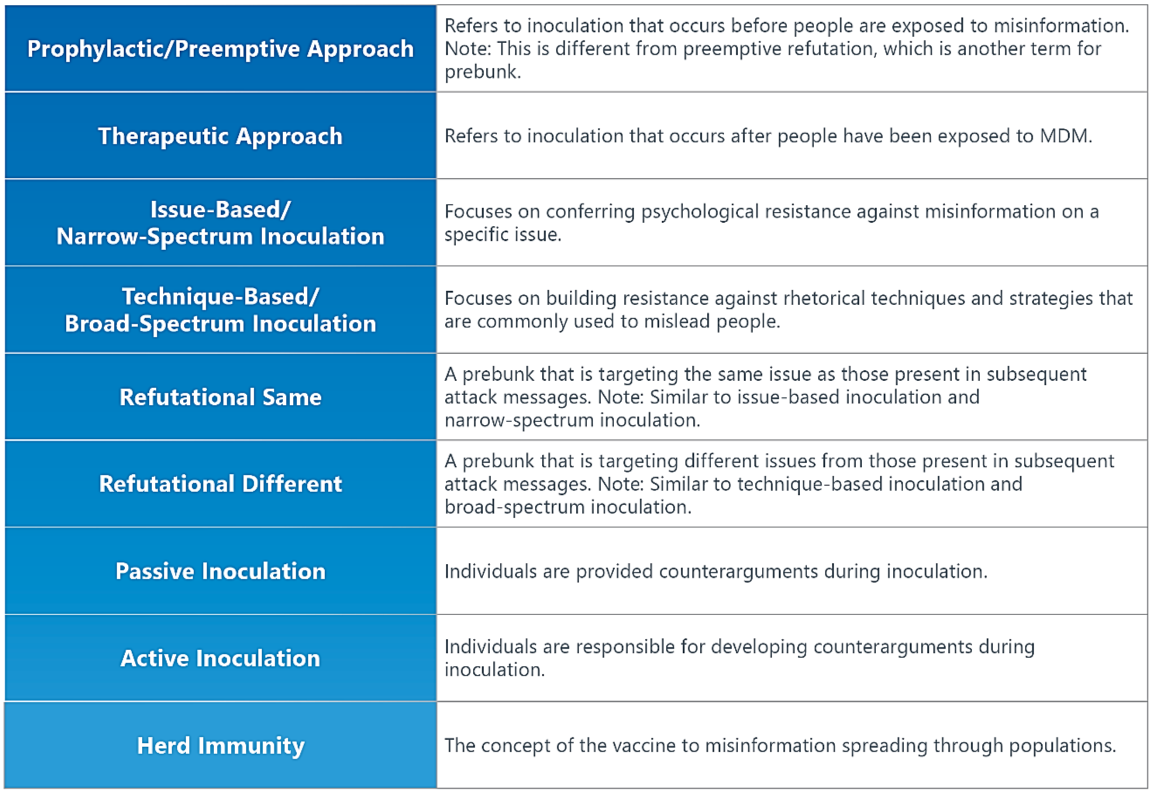
The literature distinguishes between therapeutic inoculations—those that occur *after* exposure to MDM—and prophylactic (or preemptive) inoculations—those that occur *before* exposure to MDM.[46] Fully prophylactic approaches occur before people have been exposed to any misinformation, a perhaps unrealistic standard in the Internet Age.[47] However, someone may be exposed to an issue but not know much about it; as a result, an intervention can still be quasi-prophylactic, even though the person is no longer eligible for a purely prophylactic intervention. This concept is also linked to whether interventions are effective in the face of people's preexisting beliefs, an issue discussed in later sections.

As the material above makes clear, the broad category of inoculation theory contains tremendous variety. Just juxtaposing two of the categories above—issue versus technique and passive versus active—generates four possible intervention types: issue-based/passive, issue-based/active, technique-based/passive, and technique-based/active. As a result, literally dozens of combinations can be tested and explored. Moreover, for most of the issues outlined above, the research has not yet reached clear consensus regarding the most effective path forward. Figure 4 summarizes the issues outlined above.

[46] Compton et al., "Inoculation Theory in the Post-Truth Era: Extant Findings and New Frontiers for Contested Science, Misinformation, and Conspiracy Theories."

[47] Basol et al., "Towards Psychological Herd Immunity: Cross-Cultural Evidence for Two Prebunking Interventions Against COVID-19 Misinformation."

Figure 4. Inoculation concepts



Prophylactic/Preemptive Approach	Refers to inoculation that occurs before people are exposed to misinformation. Note: This is different from preemptive refutation, which is another term for prebunk.
Therapeutic Approach	Refers to inoculation that occurs after people have been exposed to MDM.
Issue-Based/ Narrow-Spectrum Inoculation	Focuses on conferring psychological resistance against misinformation on a specific issue.
Technique-Based/ Broad-Spectrum Inoculation	Focuses on building resistance against rhetorical techniques and strategies that are commonly used to mislead people.
Refutational Same	A prebunk that is targeting the same issue as those present in subsequent attack messages. Note: Similar to issue-based inoculation and narrow-spectrum inoculation.
Refutational Different	A prebunk that is targeting different issues from those present in subsequent attack messages. Note: Similar to technique-based inoculation and broad-spectrum inoculation.
Passive Inoculation	Individuals are provided counterarguments during inoculation.
Active Inoculation	Individuals are responsible for developing counterarguments during inoculation.
Herd Immunity	The concept of the vaccine to misinformation spreading through populations.

Source: CNA.

Overall findings

Research, studies, and experiments have found that inoculation theory effectively neutralizes and builds resistance to MDM. According to Lewandowsky and van der Linden, “Over the last 50 years, a large body of evidence across domains—from health to political campaigning—has revealed that inoculation messages can be effective at conferring resistance to persuasion.”[48] This general consensus is borne out by the majority of recently published literature.

A 2010 meta-analysis of research on inoculation theory offered support for a wide range of hypotheses being explored by researchers.[49] Although this meta-analysis did not include the

[48] Lewandowsky and Linden, “Countering Misinformation and Fake News Through Inoculation and Prebunking.”

[49] Banas and Rains, “A Meta-Analysis of Research on Inoculation Theory.”

spike in literature since 2016, it provides a useful outline of the major conclusions reached before the 2016 spike. The meta-analysis found support for the following hypotheses:

1. Inoculated participants were more resistant to attack than those who were not inoculated.
2. Inoculated participants would be less susceptible to an attack than those who received some other type of supportive treatment.
3. Resistance to persuasion conferred by inoculation generalizes beyond the arguments refuted in those treatments.

The study also found **no** support for the following hypotheses:

1. A moderate delay between inoculation and attack would be more effective than no delay or a long delay.
2. Inoculation would be more effective among individuals who are moderately involved with the issue (as compared to individuals who have low or high involvement with an issue).
3. Inoculation would be more effective among participants who felt greater levels of threat.

More recent work has affirmed some of these earlier findings. For example, in a 2019 real-world experiment, Roozenbeek and van der Linden had 15,000 participants play an online inoculation game called "Bad News" as an active inoculation, which significantly reduced the participants' perception of the reliability of tweets containing misinformation strategies.[50] Because the game was public, this experiment unfortunately had no control group. An additional series of experiments found that inoculation videos improved participants' ability to recognize manipulation techniques, boosted their confidence in spotting these techniques, increased their ability to discern trustworthy from untrustworthy content, and improved the quality of their sharing decisions.[51]

A second online inoculation game called Go Viral!, which focuses on COVID-19 misinformation, has also been the subject of significant experimentation.[52] Go Viral! (a game created in collaboration with the United Kingdom (UK) Cabinet Office, the Disinformation Intervention Model (DROG), the World Health Organization (WHO), and the United Nations (UN)) identifies three techniques used to spread COVID-19 misinformation: fear mongering, fake experts, and

[50] Roozenbeek and Linden, "Fake News Game Confers Psychological Resistance Against Online Misinformation."

[51] See: Jon Roozenbeek, Cecilie S. Traberg, and Sander van der Linden, "Technique-Based Inoculation Against Real-World Misinformation," *Royal Society Open Science* 9, no. 5 (2022), doi: doi:10.1098/rsos.211719, https://royalsocietypublishing.org/doi/abs/10.1098/rsos.211719.

[52] Basol et al., "Towards Psychological Herd Immunity: Cross-Cultural Evidence for Two Prebunking Interventions Against COVID-19 Misinformation."

conspiracy theories. One large-sample experiment found that people who played Go Viral! assessed misinformation to be more manipulative after playing than before. Similarly, a preregistered randomized controlled trial (the gold standard of scientific experimentation) found that interventions such as Go Viral! significantly increased the perceived manipulativeness of misinformation about COVID-19.[53] It also found that people who had played the game were more confident in their ability to identify manipulation, were less willing to share misinformation with others, and were better able to distinguish real news and misinformation about COVID-19 after playing.

Although most of the research found inoculation effective, not all work on the topic has come to this conclusion. In a 2020 experiment by Williams and Bond, some participants were exposed to (1) information on scientific consensus related to climate change, (2) an inoculation message warning that they might see MDM on climate change, and (3) MDM related to climate change. These participants then demonstrated an increased perception of scientific consensus related to climate change.[54] Unfortunately, Williams and Bond found a similar pattern when the inoculation was left out. In other words, participants who were exposed to (1) information on scientific consensus related to climate change and (3) MDM related to climate change *also* demonstrated an increased perception of scientific consensus related to climate change.[55] This finding doesn't suggest that inoculation is ineffective, but it led the researchers to write that they were "unable to conclude that the inoculation intervention provides an additional benefit beyond the simple consensus-treatment intervention (i.e., telling participants about the percentage of scientists who agree that climate change is occurring due to human activities)."[56] Similarly, a 2022 article by Dai et al. found that exposure to inoculation messages did not significantly increase resistance to misinformation.[57] These findings, though significant in their dissent, remain rare in the literature.

Consensus and expertise

Inoculation literature uses the term *consensus* in several ways. In some cases, the literature touches on a phenomenon called the "consensus effect," which is when people are influenced by a *perceived* consensus on an issue; for example, social media algorithms respond to user activity by providing more confirmatory reporting and limiting contradictory reporting,

[53] Ibid.

[54] Matt N. Williams and Christina M. C. Bond, "A Preregistered Replication of 'Inoculating the Public Against Misinformation About Climate Change,'" *Journal of Environmental Psychology* 70 (2020), doi: 101456.

[55] Ibid.

[56] Ibid.

[57] Dai et al., "The Effects of Self-Generated and Other-Generated eWOM in Inoculating Against Misinformation."

creating the impression that consensus exists. In other cases, the literature explores instances in which *accurate* consensus information is included in an inoculation intervention (e.g., 99 percent of scientists agree that anthropogenic climate change is real). This section discusses the latter type of consensus.

Multiple researchers have explored the extent to which consensus information (alone or in conjunction with inoculation techniques) can protect against the effect of misinformation. Although van der Linden et al. and Cook et al. used different methodologies, their research yielded similar results: both teams found that the use of consensus information in an inoculation (specifically, providing information on consensus in the scientific community regarding climate change) protected participants from misinformation.[58]

That said, van der Linden et al. found that the *consensus information* negated the effects of misinformation, while Cook et al. found that the *inoculation* negated the effects of misinformation. Cook et al. also found that even though the inoculation protected against misinformation, the consensus material (not the inoculation) increased participants' sense of whether or not there was consensus on the topic.

Taken together, these studies suggest that consensus information has an overwhelmingly positive impact, either by protecting against the effects of misinformation or by changing people's perceptions of consensus; however, the relationship between inoculation and consensus information is complex and not yet fully understood.

Preexisting beliefs

Additional research has focused on the relationship between an individual's preexisting views on an issue and their subsequent suitability for successful inoculation.[59]

To begin, Cook et al. looked for differences in how preexisting views affected engagement information.[60] In one experiment, they found that false-balance media coverage (giving non-scientific contrarian views equal airtime as scientific consensus views) lowered perceived

[58] See John Cook, Stephan Lewandowsky, and Ullrich K. H. Ecker, "Neutralizing Misinformation Through Inoculation: Exposing Misleading Argumentation Techniques Reduces Their Influence," *PLOS ONE* 12, no. 5 (2017), doi: 10.1371/journal.pone.0175799, https://doi.org/10.1371/journal.pone.0175799; and Sander van der Linden et al., "Inoculating the Public Against Misinformation About Climate Change," *Global Challenges* 1, no. 2 (2017), doi: https://doi.org/10.1002/gch2.201600008, https://onlinelibrary.wiley.com/doi/abs/10.1002/gch2.201600008.

[59] Tatyana Deryugina and Olga Shurchkov, "The Effect of Information Provision on Public Consensus About Climate Change," *PLOS ONE* 11, no. 4 (2016), doi: 10.1371/journal.pone.0151469, https://doi.org/10.1371/journal.pone.0151469.

[60] Cook, Lewandowsky, and Ecker, "Neutralizing Misinformation Through Inoculation: Exposing Misleading Argumentation Techniques Reduces Their Influence."

consensus overall, although the effect was greater among free-market supporters (a proxy for conservatives). More plainly, when media gave equal coverage to contrarian beliefs and to widely accepted scientific findings, conservatives expressed more doubt (than liberals) about whether consensus existed on the issue.

They also found that misinformation that confuses people about the level of scientific agreement regarding anthropogenic global warming (AGW) had a polarizing effect, with high free-market supporters (again, a proxy for conservatives) reducing their acceptance of AGW and with low free-market supporters (a proxy for liberals) increasing their acceptance of AGW. In other words, content designed to confuse people had stronger effects on conservatives (who expressed less confidence in the belief that humans had contributed to global warming) than on liberals (who expressed more confidence in the belief that humans had contributed to global warming).

A second Cook et al. experiment found that the source of misinformation affects people differently depending on their preexisting beliefs. Misinformation from fake experts had a polarizing impact, such that it caused people with high free-market support (presumed conservatives) to demonstrate decreased climate acceptance, while people with low free-market support (presumed liberals) demonstrated increased climate acceptance.[61] In other words, fake experts drove conservatives to be more skeptical about climate change information and drove liberals to be more accepting of climate change information.

Finally, in a third experiment, Cook et al. found that high free-market supporters responded to a misinformation message by *reducing* acceptance of AGW, while low free-market supporters responded by *increasing* acceptance of AGW.[62] In other words, they concluded that people who are highly invested in their worldviews may respond to misinformation messaging by reaffirming their preexisting beliefs.

Despite these differences in how preexisting beliefs might influence people's perceptions of misinformation, Cook et al. also found that inoculation messages designed to (1) explain the flawed argumentation technique used in the misinformation or (2) highlight the scientific consensus on climate change were successful in neutralizing the adverse effects of misinformation (including its polarizing effect) for both groups. This finding suggests that inoculation *can* be effective with individuals holding a wide range of political positions.

[61] Ibid.

[62] Cook et al. selected "free market supporters" based off an assumption that this is linked to political views. They state, "Accepting the evidence that human activities drive climate change suggests embracing behavioral change, including support of increased regulation of free markets. This sits uncomfortably with conservative values of liberty and freedom. Accordingly, climate change perceptions and attitudes have repeatedly been found to be strongly associated with political worldview....Free-market support was used as a proxy for political ideology, given the strong relationship between free-market support and climate attitudes." See: ibid.

Moreover, in a series of two experiments in 2017, van der Linden et al. found that it was possible to effectively inoculate people against climate change misinformation even in a politically charged environment. The studies found strong support for the efficacy of communicating the scientific consensus on human-caused climate change (in this case, countering MDM that cited the Oregon Petition Project,[63] which claims that scientific consensus on climate change is lacking). Importantly, the experiments also found that the consensus built through inoculation techniques proved equally effective across the political spectrum.[64] Lewandowsky and van der Linden similarly found—in an experiment testing susceptibility to misinformation with and without inoculation and across political beliefs—that the inoculation treatments equally protected against misinformation (and boosted belief in the scientific consensus) for those with positive, neutral, and negative prior attitudes toward the issue.[65]

More research is needed to determine the conditions under which preexisting beliefs may lead to additional polarization. If holding moderate views is eventually determined to be essential to successful inoculation, this finding could be quite limiting, especially because researchers in recent years have focused on controversial or polarizing topics. That said, in aggregate, this research suggests that people with varying political views are vulnerable to misinformation in different ways, and are thus differently affected by inoculation efforts; even so, it is in fact possible to successfully inoculate a diverse population against the influence of MDM.

Longevity of effect

Because inoculation research incorporates a wide range of scenarios (e.g., laboratory and real world, passive and active, broad and specific), there is no single answer to the question of how long an inoculation will last. That said, researchers have tested the longevity of an inoculation intervention in multiple studies and found that the effect lasts from one week to three months. Basol et al. found that Go Viral! positively affects people's ability to identify misinformation about the virus for at least one week after playing, and it significantly reduces intentions to share misinformation with others.[66] A study by Maertens et al. found that the benefits of playing "Bad News" wore off after two months without further interventions, but the benefits

[63] The Oregon Petition Project urged the US to reject the 2007 Kyoto Climate Protocol and challenged the consensus around climate change. The petition claimed to have over 30,000 signatories. Since the creation of the petition, climate deniers have used it to spread misinformation around the consensus over anthropogenic climate change.

[64] van der Linden et al., "Inoculating the Public Against Misinformation About Climate Change."

[65] Sander van der Linden and Jon Roozenbeek, "Psychological Inoculation Against Fake News," in *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation*, (New York, NY: Routledge/Taylor & Francis Group, 2021), doi: 10.4324/9780429295379-11.

[66] Basol et al., "Towards Psychological Herd Immunity: Cross-Cultural Evidence for Two Prebunking Interventions Against COVID-19 Misinformation."

remained intact for three months if the retention interval included a "booster shot" in the form of additional exposure to training content.[67] Inoculation treatments typically decay over a number of weeks,[68] which is a similar timeline to the decay of conventional rebuttal efforts.[69] Research by Zerback et al. and Niederdeppe et al. showed that inoculation wore off after a two-week delay.[70]

Studies have also found longevity differences in the effectiveness of passive and active inoculation techniques. As one example, Basol et al. found that those who played the Go Viral! game displayed minimal decay of the inoculation effect over a week, whereas those who just read infographics (a less active form of inoculation) were less able to identify manipulation. Moreover, one week after the intervention, people who played Go Viral! remained significantly more confident (than those who read the infographics) in their ability to assess whether or not misinformation was manipulative.[71] Finally, Maertens et al. found that the inoculation effect remained stable for at least three months with regular testing. However, they found significant decay without regular testing, so the long-term inoculation effect was no longer significant.[72]

[67] Rakoen Maertens, Frederik Anseel, and Sander van der Linden, "Combatting Climate Change Misinformation: Evidence for Longevity of Inoculation and Consensus Messaging Effects," *Journal of Environmental Psychology* 70 (2020), doi: https://doi.org/10.1016/j.jenvp.2020.101455.

[68] See: Banas and Rains, "A Meta-Analysis of Research on Inoculation Theory"; Jeff Niederdeppe, Sarah E. Gollust, and Colleen L. Barry, "Inoculation in Competitive Framing: Examining Message Effects on Policy Preferences," *Public Opinion Quarterly* 78, no. 3 (2014), accessed Feb. 15, 2023, doi: 10.1093/poq/nfu026, https://doi.org/10.1093/poq/nfu026; Thomas Zerback, Florian Töpfl, and Maria Knöpfle, "The Disconcerting Potential of Online Disinformation: Persuasive Effects of Astroturfing Comments and Three Strategies for Inoculation Against Them," *New Media & Society* 23, no. 5 (2021), doi: 10.1177/1461444820908530.

[69] Briony Swire et al., "Processing Political Misinformation: Comprehending the Trump Phenomenon," *Royal Society Open Science* 4, no. 3 (2017), doi: 10.1098/rsos.160802.

[70] Zerback, Töpfl, and Knöpfle, "The Disconcerting Potential of Online Disinformation: Persuasive Effects of Astroturfing Comments and Three Strategies for Inoculation Against Them"; Niederdeppe, Gollust, and Barry, "Inoculation in Competitive Framing: Examining Message Effects on Policy Preferences."

[71] Basol et al., "Towards Psychological Herd Immunity: Cross-Cultural Evidence for Two Prebunking Interventions Against COVID-19 Misinformation."

[72] Rakoen Maertens et al., "Long-Term Effectiveness of Inoculation Against Misinformation: Three Longitudinal Experiments," *Journal of Experimental Psychology* 27, no. 1 (2021), doi: 10.1037/xap0000315, NLM.

Debunking

***Debunking** is the use of a concise correction to MDM that demonstrates that the prior message or messaging campaign was inaccurate (Table 7).*

Table 7. Debunking key findings

Debunking is an effective way to reduce belief in MDM accuracy.

- Debunking can be used to correct specific instances of inaccurate information, but it cannot be used to protect people from influence in general
- Debunking messages appear to be more effective when they:
  - cite high-credibility sources (i.e., sources that have expertise and are trustworthy)
  - contain detailed corrective information, which is more effective than simple corrections
  - express stronger corrections (e.g., those containing more information)
- The tone of the correction (e.g., uncivil, neutral, affirmational) does not appear to change the effect of the correction
- The format of the correction (e.g., truth first, myth first) does not appear to change the effect of the correction

Source: CNA.

A brief history of debunking

Debunking of some form—that is, a correction that counters false information—has been used informally as a strategy to correct inaccurate information for decades. However, our literature review indicated that formal research on debunking emerged most prominently in the 1980s and 1990s, with heightened interest beginning around 2009 and increasing in recent years.[73] Early studies in the 1980s and 1990s were largely psychological, focused on how the mind works relative to misinformation and its correction. This literature included studies of mental models, attitude changes, and deductive thinking. In the early 21st century, from roughly 2000 to 2010, these psychological studies were supplemented by research focused on specific misinformation campaigns in the political, environmental, and health spheres, touching on topics such as 9/11, the Iraq War and weapons of mass destruction, vaccines, and climate

[73] Man-Pui Sally Chan et al., "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation," *Psychological Science* PMC5673564 28, no. 11 (2017), doi: 10.1177/0956797617714579, NLM; Stephan Lewandowsky et al., "Misinformation and Its Correction: Continued Influence and Successful Debiasing," *Association for Psychological Science* (Sept. 18, 2012), https://www.psychologicalscience.org/publications/journals/pspi/misinformation1.html.

change. These studies have continued since 2010, with an increasing focus on the effectiveness of specific debunking strategies. Perhaps unsurprisingly, the issues that debunking experts focus on have expanded to include MDM about both refugees and COVID-19.

In what follows, we first provide more background on debunking, explore the state of the overall research, and end with a brief discussion of how long the effects of debunking typically last.

Definition and logic of debunking

Debunking, in simple terms, consists of issuing a corrective message establishing that a prior message was inaccurate.[74] This definition is similar to that of *fact-checking* (discussed in the next section). Indeed, the debunking experts we consulted believe the two terms refer to the same fundamental activity, and that the term used in the literature represents an academic choice. A 2021 North Atlantic Treaty Organization (NATO) best practice guide, however, explicitly distinguishes the two. It describes *fact-checking* as the "process of checking that all facts in a piece of writing, news article, speech, etc. are correct." By contrast, it defines *debunking* as "the process of exposing falseness or showing that something is less important, less good or less true than it has been made to appear."[75] Differences between the two practices, as articulated by the NATO guide, are depicted in Table 8. Even though the two practices are quite similar (and arguably the same), our literature search found distinct literature on each, so we provide separate discussions in this report.

Table 8. Differences between debunking and fact-checking

	Debunking	Fact-Checking
Conducted by	Governments and organizations	Journalists, newsrooms, political analysts
Target	Specific actors or topics associated with an MDM campaign	Specific inaccuracies across a broad range of topics
Impartiality	May be partisan or strategic	Typically impartial
Purpose	To reduce harm by asserting the truth, exposing falsehoods by a particular actor, and educating the public	To correct falsehoods

Source: NATO best practice guide, 2021.

[74] Ibid.

[75] James Pamment and Anneli Lindvall Kimber, *Fact-Checking and Debunking: A Best Practice Guide to Dealing with Disinformation*, NATO Strategic Communications Centre of Excellence, 2021, https://lup.lub.lu.se/search/publication/d5a3ed77-e218-431b-ac9b-c38a6d5a98a1.

Broadly speaking, the logic of debunking is relatively straightforward: it involves the targeted provision of correct information in response to incorrect information. In this respect, debunking is primarily a therapeutic intervention (responding directly to MDM after it has been circulated), but it can be a quasi-prophylactic intervention when the correction alerts recipients to specific bad actors or sources who are likely to spread MDM as well as the techniques they use.

Overall findings

A common finding of debunking research is that corrections succeed in reducing beliefs in MDM, or as one debunking expert we consulted put it, "Corrections are wildly effective."[76] Researchers who have studied various aspects of debunking messages generally report that reduced misperceptions is a main effect of a correction, although the degree of the effect may vary depending on various issues (summarized below).

Information source

Research has shown that the information source has a key influence on people's belief in misinformation and in subsequent debunking messages. In general, high-credibility sources are more persuasive and promote greater belief and attitude change than low-credibility sources. Some scholars conceptualize credibility as encompassing *expertise* (i.e., the extent to which the source has the knowledge and experience to provide accurate information) and *trustworthiness* (the extent to which the source is providing information that the source itself assumes to be correct).[77] Research to date indicates that corrections from credible sources can change people's beliefs about misinformation,[78] but that the topic and cultural context may influence the extent of belief change.[79] Studies also indicate that source trustworthiness is more important than source expertise.[80] Because the findings vary depending on the

[76] Interview with Dr. Briony Swire-Thompson, Dec. 5, 2022.

[77] Ullrich K. H. Ecker and Luke M. Antonio, "Can You Believe It? An Investigation into the Impact of Retraction Source Credibilty on the Continued Influence Effect," *Memory & Cognition* 49 (2021); Lewandownsky et al., *The Debunking Handbook 2020*.

[78] Swire et al., "Processing Political Misinformation: Comprehending the Trump Phenomenon"; Briony Swire-Thompson et al., "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation," *Political Psychology* 41, no. 1 (2020), doi: https://doi.org/10.1111/pops.12586.

[79] Swire-Thompson et al., "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation."

[80] Ecker and Antonio, "Can You Believe It? An Investigation into the Impact of Retraction Source Credibilty on the Continued Influence Effect."

configuration of topics, contexts, and methods of the studies, more research is needed to identify effective debunking strategies for various issues (e.g., political, environmental) and cultural contexts (i.e., US or other).

A suite of studies that examined the role of source credibility in the political arena have generally found that corrective information reduces misperceptions regardless of the source, although the degree of the belief change may vary based on partisanship. For instance, a 2018 study of climate change attitudes found that all respondents, regardless of partisanship, reported increased concern about climate change and greater agreement about the scientific consensus on the topic after corrections.[81]

In 2017, Swire et al. conducted a study that explored whether the public believes initial information spread by a polarizing source (in this case, then presidential candidate Donald Trump), whether belief in false information can be effectively corrected, and, if so, whether a change in belief leads to a shift in voting preferences. Participants were presented with actual statements, some inaccurate and some factual, made by Donald Trump on the 2015 campaign trail. The false statements were then corrected and true statements affirmed, and participants rated their belief in the statements before and after the corrective information. The study's findings indicated the following:

- *Views on source credibility influence beliefs*. The opinion of Republicans and Democrats alike about Trump's credibility influenced their assessment of the original statement's accuracy; that is, Republicans tended to believe the statement, and Democrats did not.
- *Support for the initial information source influences beliefs more than the source of the corrective or affirming information.* The source of the corrective or affirming information did not affect beliefs as much as participants' support for the original information source (Trump).
- *Corrective or affirming information influence beliefs*. Republicans and Democrats alike corrected their beliefs after they read the corrective or affirming information.
- *Change in beliefs about false information did not change voting preferences*. Although participants changed their beliefs about the false information after receiving a correction, these changed beliefs did not influence their voting intentions and feelings toward Trump. One explanation for this finding is that people expect politicians to make inaccurate statements and are not overly concerned when they do so.[82]

[81] Salil D. Benegal and Lyle Scruggs, "Correcting Misinformation About Climate Change: The Impact of Partisanship in an Experimental Setting," *Climatic Change* 148, no. 1 (2018), https://EconPapers.repec.org/RePEc:spr:climat:v:148:y:2018:i:1:d:10.1007_s10584-018-2192-4.

[82] Swire et al., "Processing Political Misinformation: Comprehending the Trump Phenomenon."

Studies of debunking in the health information context have found a stronger effect of source expertise than studies in the political arena. For example, three studies have found that corrections of health-related misinformation are more effective when they come from a credible source than from a peer or other nonexpert.

- A 2017 study with US university students explored the effectiveness of corrections of misinformation about the origins of the Zika virus. Participants read false information about the virus origins on Twitter, followed by corrections from the Centers for Disease Control and Prevention (CDC), from an unknown Twitter user, or from both. Results showed that a single correction from a reputable source (the CDC) reduced misperceptions about the causes of Zika spread, particularly if the CDC correction followed a correction from another user. In contrast, a correction from a single user did not reduce misperceptions on its own, nor when it followed the CDC correction.[83]
- A pre-COVID-19 online experiment with 700 US adults examined how the source of corrective information about the threat of a new, highly infectious Asian influenza affected beliefs. Initial misinformation indicating that the virus was not a severe threat was followed by various conditions that paired three types of corrective information (i.e., no correction, simple rebuttal, and detailed factual information) with various sources (i.e., the CDC, news media, and a social media peer). Participants who received information from the CDC and news media reported higher perceived crisis severity and more anxiety than participants who received information from a social peer source. However, researchers observed no significant differences between the groups regarding the likelihood of taking preventive actions.[84]
- A 2019 study found that a correction from the CDC was more effective in reducing vaccine misperceptions than corrections from library sources or other Facebook users.[85]

Note that the studies cited above were conducted prior to the COVID-19 pandemic, when trust in the CDC and news media in the US may have been higher than in the years following the COVID-19 outbreak. Thus, people's judgments about the credibility of information from previously trusted sources of health information may have diminished.

[83] Emily K. Vraga and Leticia Bode, "Using Expert Sources to Correct Health Misinformation in Social Media," *Science Communication* 39, no. 5 (2017).

[84] Toni G. L. A. van der Meer and Yan Jin, "Seeking Formula for Misinformation Treatment in Public Health Crises: The Effects of Corrective Information Type and Source," *Health Communication* 35, no. 5 (2020).

[85] M. Connor Sullivan, "Leveraging Library Trust to Combat Misinformation on Social Media," *Library & Information Science Research* 41, no. 1 (2019).

To summarize, research on how source credibility and trustworthiness influence belief in MDM, as well as belief in debunking messages, indicates that credible sources can influence belief change, but that people's assessment of source credibility is influenced more strongly by their perceptions of trustworthiness than expertise.

Preexisting beliefs

Building on previous studies about source credibility, an online study by Swire-Thompson and colleagues presented US adults with statements from presidential candidates—Republican Donald Trump and Democrat Bernie Sanders—to examine the relationship between an individual's political party affiliation, corrective information for misinformation, and feelings toward political figures. Participants viewed statements made by one of the politicians, and then—both before and after receiving corrective information—rated their feelings toward the politicians and their beliefs in the statements. The results were as follows:

- *Each politician's supporters reduced their belief in the MDM once it was corrected*, which replicated findings from Swire et al. (2017).
- *Participants did not change their feelings about their favored political figure* if they viewed an equal number of MDM and factual statements, but there was a slight reduction in feelings if they viewed more MDM than factual statements.
- *Participants' views on their favored politician's general truthfulness did not change*, regardless of the number of corrections they viewed. Note, however, that participants' estimates of veracity was extremely low for all politicians, suggesting that people in the US expect politicians to lie.
- Researchers observed *few differences by political affiliation in the behavior of participants*. The only difference observed was that Trump supporters, compared to non-supporters, showed greater continued belief in Trump misinformation even after viewing corrections. In contrast, supporters and non-supporters of Sanders reported equally low belief in misinformation after it had been corrected.[86]

The article cites different findings from an Australian study (conducted by some of these same authors) that used a similar methodology (i.e., presenting participants with statements from leaders of the left-wing Labor Party and right-wing Liberal Party). Although participants did not change their feelings about their favored politician when presented with equal numbers of false and accurate statements, a sizable reduction in feelings occurred for participants on both the left and right after viewing disproportionately more misinformation than factual

[86] Swire-Thompson et al., "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation."

statements. Specifically, the effect size was 10 times greater in the Australia study than in the later US study by Swire-Thompson et al., and it was 6 times greater when focusing on the disproportionate condition alone (i.e., when participants were exposed to more false statements than true statements).[87] In combination, the studies seem to suggest that preexisting beliefs have some influence on correction uptake, and that this tendency is universal (i.e., not limited to a single political party); however, the differences between the American and Australian environments highlight the importance of cultural context in people's expectations of truthfulness among politicians, and in how they respond to MDM from politicians.[88]

Content of debunking message

Research on the content of the debunking message has focused on the debunking message's level of detail, framing, strength, and tone.

Level of detail of correction

Research on the level of detail of debunking messages has generally shown that debunking messages that contain detailed corrective information are more effective than simple corrections. Although we found one study that did not see a significant difference in belief updating following a simple versus detailed correction,[89] most studies reported stronger effects for detailed messages. A meta-analysis of research published between 1995 and 2015 examined the effect of the level of detail of the debunking message (i.e., simply labeling the misinformation as incorrect versus providing new and credible information) on successful debunking and curbing the persistence of misinformation. Results showed that a detailed debunking message was associated with a stronger debunking effect than a message that simply labeled the misinformation as incorrect.[90] We summarize two illustrative studies below.

- *Corrective information can counter misinformation, and more detailed corrective information can stimulate people to take appropriate actions.* The online experiment on public health misinformation (described in the prior section) considered the effect of

[87] M. J. Aird et al., "Does Truth Matter to Voters? The Effects of Correcting Political Misinformation in an Australian Sample," *Royal Society Open Science* PMC6304148 5, no. 12 (2018), doi: 10.1098/rsos.180593, NLM, cited in Swire-Thompson et al., "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation."

[88] Swire-Thompson et al., "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation."

[89] Cameron Martel, Mohsen Mosleh, and David Gertler Rand, "You're Definitely Wrong, Maybe: Correction Style Has Minimal Effect on Corrections of Misinformation Online," *Media and Communication* 9, no. 1 (Feb. 2021), https://dspace.mit.edu/handle/1721.1/129719.

[90] Chan et al., "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation."

the level of detail of the corrective information: simple rebuttal versus factual elaboration. As a reminder, participants were exposed to misinformation that an emerging virus was not a serious threat, and then randomly assigned to one of the seven conditions that manipulated both the corrective information source and the level of detail in the correction. Results on a questionnaire found that participants who received any kind of corrective information perceived the crisis as more severe and felt more fear and anxiety. In addition, participants exposed to factual elaboration, compared to those exposed to the simple rebuttal, reported more anxiety and fear as well as greater likelihood to take preventive actions.[91]

- *Well-designed, detailed refutations are the most effective in reducing misinformation promotion.* MacFarlane and colleagues conducted an online experiment with over 600 US adults to explore whether exposure to misinformation claiming that vitamin E can prevent and cure COVID-19 would affect participants' willingness to purchase vitamin E and to share the misinformation on social media. The researchers used hypothetical behavioral measures (rather than asking directly about beliefs) to assess the influence of misinformation on participants' beliefs. Specifically, participants indicated how much they would bid in a hypothetical auction to purchase vitamin E supplements and whether they would share, like, flag as inappropriate, or decline to interact with a social media post containing the misinformation. Results showed that both refutation types substantially reduced participants' willingness to pay (for vitamin E) and willingness to share misinformation, and that the enhanced refutation was more effective than the tentative refutation in reducing misinformation promotion.[92]

Framing of the correction

Researchers have examined whether the effectiveness of debunking messages is associated with the manner in which the correction is framed. For instance, studies have examined corrections that use humor, logic, facts, or narratives; corrections that incorporate news literacy; and corrections that provide various reminders of the original MDM. For the most part, these studies found that corrections work regardless of framing, when compared to the no-correction control condition. However, the degree of effectiveness sometimes varies. We summarize key findings from this body of research below.

[91] van der Meer and Jin, "Seeking Formula for Misinformation Treatment in Public Health Crises: The Effects of Corrective Information Type and Source."

[92] Doulas MacFarlane et al., "Refuting Spurious COVID-19 Treatment Claims Reduces Demand and Misinformation Sharing," *Journal of Applied Research in Memory and Cognition* PMC7771267 10, no. 2 (2021), doi: 10.1016/j.jarmac.2020.12.005, NLM.

- *Fact-based corrections are generally more effective than corrections using other devices* (with some nuanced findings). Illustrative examples include studies demonstrating the following:
    - Factual corrections outperformed narrative corrections (that tell a story) in debunking MDM about e-cigarettes.[93]
    - Detailed factual corrections that address parental concerns about vaccines are more effective than one-dimensional corrections that do not acknowledge these concerns.[94]
    - Logic-focused corrections (that explain rhetorical techniques used to mislead) and fact-focused corrections (that counter MDM with factual information) of climate change MDM on Instagram were equally effective at reducing misperceptions when seen *after* the misinformation post, but only the logic-focused corrections reduced misperceptions when they appeared *before* the misinformation.[95]
- *Corrections that explicitly repeat the original MDM are more effective than those with subtle reminders*. A study by Ecker et al. found that any kind of correction reduced reliance on MDM, but that corrections explicitly reminding recipients of the MDM were more effective than those that merely pointed out that the earlier MDM was incorrect.[96]

Tone of the correction

Two studies that examined the tone of debunking messages found that corrections appear to be effective regardless of the tone. A study of corrections on social media found that corrections that call attention to the MDM and offer credible information to counter it reduce misperceptions regardless of whether the correction's tone is uncivil (e.g., "Don't be stupid, everybody knows that..."), affirmational ("This is superconfusing, but it is not true…"), or

[93] Yan Huang and Weirui Wang, "When a Story Contradicts: Correcting Health Misinformation on Social Media Through Different Message Formats and Mechanisms," *Information, Communication & Society* 25, no. 8 (2022), doi: 10.1080/1369118X.2020.1851390.

[94] Anat Gesser-Edelsburg et al., "Correcting Misinformation by Health Organizations During Measles Outbreaks: A Controlled Experiment," *PLOS ONE* 13, no. 12 (2018), doi: 10.1371/journal.pone.0209505.

[95] Emily K. Vraga et al., "Testing the Effectiveness of Correction Placement and Type on Instagram," *International Journal of Press/Politics* 25, no. 4 (2020).

[96] Ullrich K. H. Ecker, Joshua L. Hogan, and Stephan Lewandowsky, "Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction?" *Journal of Applied Research in Memory and Cognition* 6 (2017), doi: 10.1037/h0101809.

neutral (e.g., "This isn't true…" followed by facts).[97] Similarly, Sangalang et al. found that narrative corrections of MDM about the safety of natural tobacco products were effective at reducing misinformed beliefs and behavioral intentions regardless of whether they contained emotional content (i.e., anger about being misled) or were neutral (i.e., simply corrected the MDM).[98]

Format

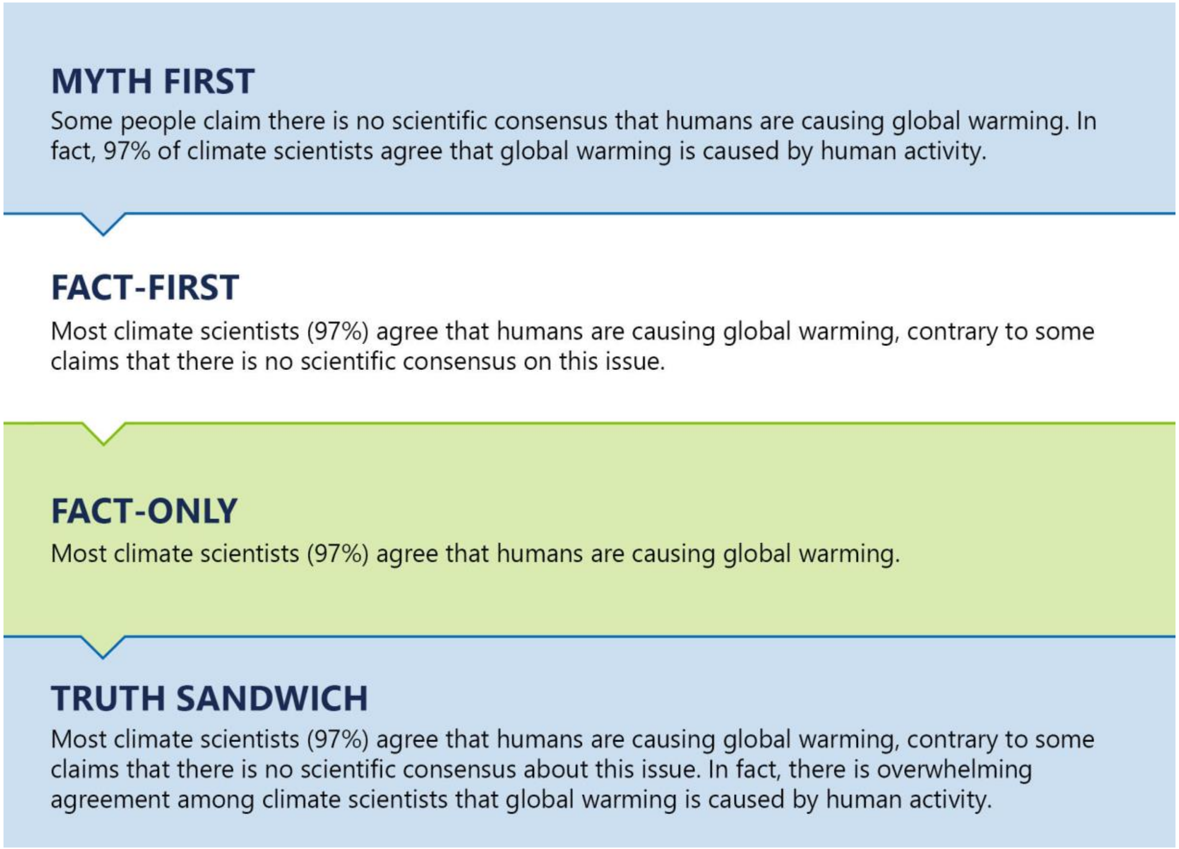
Based on earlier research suggesting that repeating the MDM within a correction may inadvertently reinforce the MDM (e.g., "You may have heard that vitamin E cures COVID-19; in fact, vitamin E does not cure COVID-19"), some scholars have suggested that corrections should be presented in a "truth sandwich" format that begins with the factual information, briefly repeats the MDM, and concludes by reinforcing the fact (e.g., "Vitamin E should not be used to prevent or treat COVID-19. Although some sources claim that vitamin E can prevent or cure COVID-19, research does not support this claim").[99] Recent research has sought to study this idea in controlled settings. Swire-Thompson and colleagues (2021) conducted four online experiments (three with US adults and university students and one with Australian students) to assess the effectiveness of several correction formats (myth first, fact first, etc.) using a range of materials, participant pools, and topics (see Figure 5).

[97] Leticia Bode, Emily. K. Vraga, and Melissa Tully, "Do the Right Thing: Tone May Not Affect Correction of Misinformation on Social Media," *Harvard Kennedy School (HKS) Misinformation Review* (2020), https://doi.org/10.37016/mr-2020-026.

[98] Angela Sangalang, Yotam Ophir, and Jospeh. N. Cappella, "The Potential for Narrative Correctives to Combat Misinformation(†)," *Journal of Communication* PMC6544903 69, no. 3 (2019), doi: 10.1093/joc/jqz014, NLM.

[99] Lewandowsky et al., *The Debunking Handbook 2020*.

Figure 5. Possible formats for debunking messages



MYTH FIRST

Some people claim there is no scientific consensus that humans are causing global warming. In fact, 97% of climate scientists agree that global warming is caused by human activity.

FACT-FIRST

Most climate scientists (97%) agree that humans are causing global warming, contrary to some claims that there is no scientific consensus on this issue.

FACT-ONLY

Most climate scientists (97%) agree that humans are causing global warming.

TRUTH SANDWICH

Most climate scientists (97%) agree that humans are causing global warming, contrary to some claims that there is no scientific consensus about this issue. In fact, there is overwhelming agreement among climate scientists that global warming is caused by human activity.

Source: Adapted from Swire-Thompson et al. 2021.

Results showed that the effectiveness of a correction in influencing beliefs and inferential reasoning was largely independent of the format. Although one experiment suggested that the myth-first approach may be superior, researchers have called for more research on whether the "truth sandwich" approach is more effective than other approaches.[100] Broadly speaking, though, the corrective message format did not seem to make a considerable difference as long as the key ingredients of a correction were presented.[101] Current research thus makes clear

[100] Briony Swire-Thompson et al., "Searching for the Backfire Effect: Measurement and Design Considerations," *Journal of Applied Research in Memory and Cognition* 9, no. 3 (2020).

[101] Briony Swire-Thompson et al., "Correction Format Has a Limited Role When Debunking Misinformation," *Cognitive Research: Principles and Implications* PMC8715407 6, no. 1 (2021), doi: 10.1186/s41235-021-00346-6, NLM.

that providing corrective information, regardless of format, is far more important than how the correction is presented.[102]

Longevity of effect

We found few studies that explored the longevity of debunking effects, but a soon-to-be published study by Swire-Thompson et al. (2023) examined the phenomenon of *belief regression,* in which people who initially believed a correction appear to re-endorse or "re-believe" in the original misinformation over time. The authors distinguish belief regression from the continued influence effect in that *belief regression* refers to the impermanence of the correction's efficacy over time, whereas *continued influence effect* refers to general continued use of corrected misinformation in memory and reasoning. Noting that prior studies have shown belief regression to be a robust phenomenon, the authors aimed to better understand the role of memory in belief regression and whether belief regression is more likely to occur with corrected misinformation than affirmed facts.

The experimental study with over 600 US adults assessed participants at three time points: pre-test, immediately post-test, and one-month delayed post-test. In the pre-test, participants rated 16 facts and 16 MDM items for their accuracy as well as how much time and thought they had given to the issues in the past. Participants in the correction condition were shown corrections of MDM and affirmations of facts. In the immediate and delayed post-tests, all participants re-rated their beliefs in each item. Those in the control condition also rated how well they remembered whether each statement was true or false. Results showed that participants with better memories were more likely to reduce their belief in the MDM, both immediately and after a one-month delay, and that memory at the one-month delay explained 66 percent of the variance in belief regression. Based on their findings, the authors suggest that repeated corrections could effectively counteract belief regression, and they call for additional research on the phenomenon.[103]

[102] Swire-Thompson et al., "Correction Format Has a Limited Role When Debunking Misinformation."

[103] Briony Swire-Thompson et al., "Memory Failure Predicts Belief Regression After the Correction of Misinformation," *Cognition* 230 (2023), doi: 10.1016/j.cognition.2022.105276, NLM.

Fact-Checking

Fact-checking *is a journalistic practice designed to reject clearly false claims with empirical evidence from neutral or unimpeachable sources (Table 9).*[104]

Table 9. Fact-checking key findings

Fact-checking is an effective way to reduce belief in MDM accuracy.

- Fact-checking can be used to correct specific instances of inaccurate information, but it cannot be used to protect people from influence in general
- Fact-checking is best when integrated into the consumption of news
- Fact-checking is a potentially powerful tool for DOD personnel with communications responsibilities

Source: CNA.

A brief history of fact-checking

Fact-checking came to contemporary prominence during the 2010s as part of evolutions in news coverage in the early social media era. According to Graves et al.,[105] journalists started engaging in fact-checking out of professional motives. Contemporary precursors to fact-checking interventions are retractions[106] and "ad watches."[107] That said, the foundations of contemporary fact-checking go back nearly two centuries, first with the Associated Press's shift to strictly publishing only "material facts" in 1854, followed by *Time* magazine instituting a research department in 1923 with the explicit task of fact-checking article drafts.[108]

[104] Interview with fact-checking expert, Dec. 1, 2022.

[105] Lucas Graves et al., "Understanding Innovations in Journalistic Practice: A Field Experiment Examining Motivations for Fact-Checking," *Journal of Communication* 66, no. 1 (2016), https://doi.org/10.1111/jcom.12198.

[106] Ullrich K. H. Ecker et al., "Correcting False Information in Memory: Manipulating the Strength of Misinformation Encoding and Its Retraction," *Psychonomic Bulletin & Review* 18, no. 3 (2011), doi: 10.3758/s13423-011-0065-1, https://doi.org/10.3758/s13423-011-0065-1.

[107] Michael Pfau and Allan Louden, "Effectiveness of Adwatch Formats in Deflecting Political Attack Ads," *Communication Research* 21, no. 3 (1994), doi: 10.1177/009365094021003005.

[108] Colin Dickey, "The Rise and Fall of Facts," *Columbia Journalism Review* (Fall 2019), accessed Nov. 4, 2022, https://www.cjr.org/special_report/rise-and-fall-of-fact-checking.php; Merrill Fabry, "Here's How the First Fact-Checkers Were Able to Do Their Jobs Before the Internet," *Time*, Aug. 24, 2017, accessed Aug. 24, 2017, https://time.com/4858683/fact-checking-history/.

Fact-checking is one of the key postulated interventions against disinformation in modern times.[109] Within the US, fact-checking is primarily done by journalists and news organizations.[110] Examples include the *Washington Post*'s Fact Check with its scale of Pinocchios, PolitiFact's Truth-O-Meter, and Snopes's spectrum-based rating system.[111] Outside the US, a growing number of nonprofit and nongovernmental organizations have been established with a mission to conduct fact-checking.[112] As Nieminen and Rapeli note in their review of the fact-checking research, "It seems that an international fact-checking movement is emerging."[113]

In what follows, we first provide more background on fact-checking, explore the state of the overall research, and end with a brief discussion of how long the benefit of a fact-checking intervention typically lasts.

Definition and logic of fact-checking

As we noted in the previous section, debunking and fact-checking are quite similar, but a distinguishing characteristic is that fact-checking is primarily employed by journalists and newsrooms, is typically impartial (to the extent that it adheres to journalistic standards of accuracy), and aims to correct all falsehoods within a given context (e.g., a political debate).

Most studies of fact-checking do not explicitly define the concept, instead jumping right into discussion of its various aspects. However, some of the broader reviews of the research on fact-checking[114] do provide a definition, though these remain overly complicated. In this review, we adopt the definition provided to us by an expert on the topic:[115] *fact-checking* is a journalistic practice designed to reject clearly false claims with empirical evidence from neutral or unimpeachable sources.

[109] Nathan Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom," *Political Communication* 37, no. 3 (2020): 360, doi: 10.1080/10584609.2019.1668894.

[110] Daniel Funke, "From Pants on Figre to Pinocchio: All the Ways That Fact-Checkers Rate Claims," Poynter, June 18, 2019, accessed Nov. 4, 2022, https://www.poynter.org/fact-checking/2019/from-pants-on-fire-to-pinocchio-all-the-ways-that-fact-checkers-rate-claims/.

[111] Ibid.

[112] Ibid.; S. Nieminen and L. Rapheli, "Fighting Misperceptions and Doubting Journalists' Objectivity: A Review of Fact-Checking Literature," *Perspectives on Psychological Science* 17, no. 3 (2019).

[113] Nieminen and Rapheli, "Fighting Misperceptions and Doubting Journalists' Objectivity: A Review of Fact-checking Literature," 2.

[114] See: Ibid.; Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom."

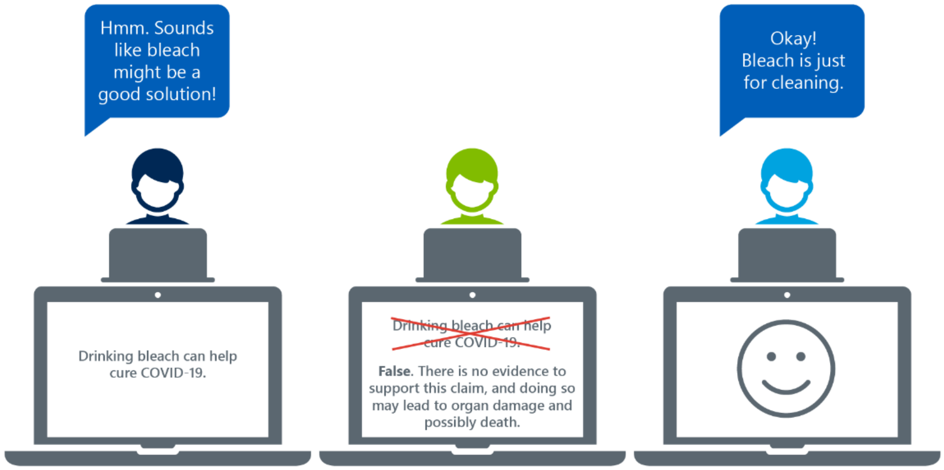
[115] Interview with fact-checking expert, Dec. 1, 2022.

The research on fact-checking is extensive, with a significant body of work published in peer-reviewed journals that leverages social science research approaches (e.g., field and survey experiments, statistical analysis) to address the key research questions. Though the research focuses on aspects of fact-checking (e.g., its format, effectiveness), the general logic of fact-checking remains relatively consistent. In an idealized sense, fact-checking works as follows:

1. An individual is presented with false information.
2. The individual is presented with subsequent information that corrects the initial information.
3. The individual updates their belief to be more aligned with the factual information.

Fact-checking is, as a result, a therapeutic intervention designed to directly respond to a specific piece of false information. As a specific example, suppose a Twitter user saw a tweet stating "Got COVID-19? Drink some bleach to kill the virus and recover faster." The intervention here would be a corrective statement subsequently presented to the user (a "fact-check") stating that no evidence shows that drinking bleach would kill the virus (and, ideally, an additional statement about how drinking chemical products like bleach could result in serious internal damage and possibly death). The result is that the user will update (or "correct") their prior belief about drinking bleach as a treatment for COVID-19 as false. Figure 6 illustrates this process.

Figure 6. Fact-checking: an example


Hmm. Sounds like bleach might be a good solution!
Drinking bleach can help cure COVID-19.
~~Drinking bleach can help cure COVID-19.~~
False. There is no evidence to support this claim, and doing so may lead to organ damage and possibly death.
Okay! Bleach is just for cleaning.

Source: CNA.

Overall findings

Unfortunately, not all studies of fact-checking use the same research design, sampling strategy, or set of covariates when testing the effectiveness of fact-checking,[116] so comparing findings across studies is challenging. A recent meta-analysis determined that the overall effectiveness of fact-checking seems to be contingent upon "various moderating variables"[117] (which we discuss below): political sophistication, the nature of the message, preexisting beliefs, and the longevity of the effect.

Political sophistication

Research suggests that political sophistication (i.e., how aware and knowledgeable an individual is about politics) may influence the effectiveness of fact-checking. The idea is that individuals who are more attentive to politics may be more skeptical of corrections, thereby reducing their effectiveness as an intervention for false information.[118] Political sophistication is distinct from the preexisting beliefs factor because it is purely about political awareness and knowledge (rather than partisan ideology). However, this dynamic is related to partisanship in that individuals who are more partisan have been found to be more motivated to reason against a fact-check, especially one that counters their political beliefs.[119] In their meta-analysis of the research on fact-checking, Walter et al. found support for this notion.[120] However, Fridkin et al., Young et al., and Velez et al. found that political sophistication (measured by knowledge and awareness of politics) has no impact on the effectiveness of fact-checking.[121]

[116] This is in contrast to, for instance, the academic literatures on civil wars and political regime transitions in which scholars often use a similar set of covariates and conduct statistical analyses and robustness checks across the same few datasets when examining the core research questions, such as civil war onset and political regime survival or collapse. Research conducted in this way lends itself much better to the aggregation and comparison of findings across dozens of studies.

[117] Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom," 360.

[118] Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom," 353.

[119] Nieminen and Rapheli, "Fighting Misperceptions and Doubting Journalists' Objectivity: A Review of Fact-Checking Literature," 7.

[120] Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom," 360.

[121] Kim Fridkin, Patrick J. Kenney, and Amanda Wintersieck, "Liar, Liar, Pants on Fire: How Fact-Checking Influences Citizens' Reactions to Negative Advertising," *Political Communication* 32, no. 1 (2015), doi: 10.1080/10584609.2014.914613; Dannagal G. Young et al., "Fact-Checking Effectiveness as a Function of Format and Tone: Evaluating FactCheck.org and FlackCheck.org," *Journalism & Mass Communication Quarterly* 95, no. 1 (2018), doi: 10.1177/1077699017710453; Yamil R. Velez, Ethan Porter, and Thomas J. Wood, "Latino-Targeted Misinformation and the Power of Factual Corrections," *Journal of Politics* (published online Feb. 14, 2023).

Thus, among the research we examined that studies the influence of political sophistication on fact-checking, the findings are mixed.

Nature of the fact-checking message

Presentation

How a fact-check is presented may influence its effectiveness, though the findings are again mixed across studies. To start, the length and complexity of a presented fact-check can influence its effectiveness.[122] Longer and more complex written fact-checks are not as effective as shorter and more concise statements.[123] Representation also matters, since visual rating scales or "truth scales" that indicate the degree of accuracy have been found to be more effective than simple corrective statements.[124] An example is the *Washington Post*'s "Pinocchio scale" (included, along with a variety of truth scales, in Figure 7). The scale is ordinal and indicates how true a statement is from "mostly true" (one Pinocchio) to "whoppers" (four Pinocchios).[125] The scale also includes the "Bottomless Pinocchio" for false claims that have received three or four Pinocchios and have been repeated at least 20 times.[126]

[122] Oscar Barrera et al., "Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics," *Journal of Public Economics* 182 (2020), https://doi.org/10.1016/j.jpubeco.2019.104123; Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom," 355 and 365.

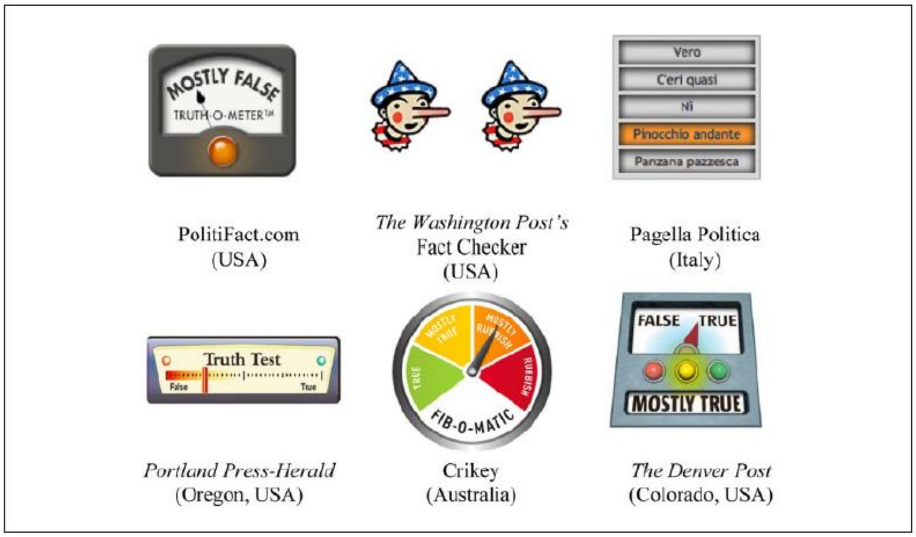
[123] Ibid., 355.

[124] Brendan Nyhan and Jason Reifler, *Misinformation and Fact-Checking: Research Findings from Social Science*, New America Foundation (Feb. 2012); Brendan Nyhan and Jason Reifler, "The Roles of Information Deficits and Identity Threat in the Prevalence of Misperceptions," *Journal of Elections, Public Opinion and Parties* 29, no. 2 (2019); Michelle A. Amazeen et al., "Correcting Political and Consumer Misperceptions: The Effectiveness and Effects of Rating Scale Versus Contextual Correction Formats," *Journalism & Mass Communication Quarterly* 95, no. 1 (2018), doi: 10.1177/1077699016678186.

[125] See: Glenn Kessler, "About the Fact Checker," *Washington Post*, Jan. 1, 2017, https://www.washingtonpost.com/politics/2019/01/07/about-fact-checker/.

[126] Ibid.

Figure 7. A selection of truth scales



Source: Amazeen, Michelle A., Emily Thorson, Ashley Muddiman, and Lucas Graves. "Correcting Political and Consumer Misperceptions: The Effectiveness and Effects of Rating Scale Versus Contextual Correction Formats," *Journalism & Mass Communication Quarterly* 95, no. 1 (2018): 28-48.

However, as with the discussion on political sophistication, other researchers have made opposite findings. In their meta-analysis of the research on fact-checking, Walter et al. found that including graphical elements with a fact-check diminishes the effectiveness of the misinformation correction.[127] Walter et al. also found that message length does not influence fact-checking effectiveness, but the complexity of a message does, with more complex corrections negatively correlated with fact-checking effectiveness.

Format

In the same vein as presentation, the format of fact-checking (e.g., print versus video versus photograph) may influence its effectiveness. For instance, Young et al. found that fact-checking via a video (whether humorous or non-humorous) was more effective in correcting beliefs than fact-checking via a written statement.[128] That said, Garrett et al. found that including additional

[127] Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom," 364.

[128] Young et al., "Fact-Checking Effectiveness as a Function of Format and Tone: Evaluating FactCheck.org and FlackCheck.org."

contextual information and cues (such as photographs) may undermine the effectiveness of fact-checking on correcting disinformation.[129] Again, the overall findings appear to be mixed.

Preexisting beliefs

Like other MDM interventions, the baseline effectiveness of fact-checking is shaped by the human tendency to engage in motivated reasoning: "Preexisting beliefs play a major role in determining the way information is processed even in the face of concrete evidence and mounting facts…citizens are more accepting of (mis)information that match their preexisting worldview."[130] In addition, the natural human desire to avoid cognitive dissonance will lead people to prioritize coherent information, even if it is inaccurate.[131] More plainly: people like information that fits neatly with what they already believe, and they don't like information that conflicts with what they already believe. This reality creates a challenge for fact-checking efforts because the correct information will, in some cases, need to overcome these inherent human tendencies.

As with research on both inoculation theory and debunking, research on fact-checking has identified some partisan-related differences in the American political context. Specifically, research findings indicate that conservatives may be less receptive to fact-checking[132] for a variety of reasons.[133] Per Gallup polls on the issue, these reasons include the aforementioned human tendency to engage in motivated reasoning to avoid cognitive dissonance, coupled with

[129] Garrett, Nisbet, and Lynch, "Undermining the Corrective Effects of Media-Based Political Fact Checking? The Role of Contextual Cues and Naïve Theory."

[130] Walter et al., 2020, 352-353, partially from D. J. Flynn, Brendan Nyhan, and Jason Reifler, "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics," *Political Psychology* 38, no. S1 (2017): 127–150, doi: https://doi.org/10.1111/pops.12394.

[131] Lewandowsky et al., "Misinformation and Its Correction: Continued Influence and Successful Debiasing."

[132] Nyhan and Reifler, "When Corrections Fail: The Persistence of Political Misperceptions"; Fridkin, Kenney, and Wintersieck, "Liar, Liar, Pants on Fire: How Fact-Checking Influences Citizens' Reactions to Negative Advertising"; Jeffrey W. Jarman, "Influence of Political Affiliation and Criticism on the Effectiveness of Political Fact-Checking," *Communication Research Reports* 33, no. 1 (2016): 9-15, doi: 10.1080/08824096.2015.1117436; Ethan Porter et al., "Can Presidential Misinformation on Climate Change Be Corrected? Evidence from Internet and Phone Experiments," *Research and Politics* 6, no 3. (2019), https://doi.org/10.1177/2053168019864784; Brendan Nyhan et al., "Taking Fact-Checks Literally but Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability," *Political Behavior* 42 (2020): 939–960, doi: 10.1007/s11109-019-09528-x; Swire-Thompson et al., "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation," 21-34; N. Walter et al., "Evaluating the Impact of Attempts to Correct Health Misinformation on Social Media: A Meta-Analysis," *Health Communication* 36, no. 13 (2021): 1776-1784, doi: 10.1080/10410236.2020.1794553, NLM.

[133] Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom," 354 and 364.

a lower likelihood to seek out information that runs counter to their beliefs[134] and a greater distrust of the media.[135] These findings concerning conservatives are not a product of the present era; rather, they were established in fact-checking studies dating back to the first decade of the 2000s.[136] However, there is significant debate about whether these findings reflect reality or are a product of specific research designs and publication bias.[137]

Longevity of effect

Finally, most studies do not examine the longevity of fact-checking as an intervention, instead focusing on whether it works in an immediate sense. As a result, it is unclear how long-lasting fact-checking is as an intervention against disinformation. A few studies have explicitly examined the longevity of fact-checking with mixed results. Swire et al. found that the efficacy of fact-checking declines over time, with fact-checks sustaining belief change one week out but then declining three weeks out.[138] In contrast, Porter and Wood used a second-wave survey of participants two weeks after the initial fact-check and found that it was still effective two weeks later.[139] As noted, the research that explicitly examines this aspect of fact-checking is limited, so whether fact-checking works as a lasting intervention against disinformation is unclear.

[134] Pablo Barberá et al., "Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber?" *Psychological Science* 26, no. 10 (2015): 1531–1542, doi: 10.1177/0956797615594620, https://journals.sagepub.com/doi/abs/10.1177/0956797615594620.

[135] Art Swift, "Americans' Trust in Mass Media Sinks to New Low," Gallup, Sept. 14, 2016, accessed Nov. 4, 2022, https://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx; Megan Brenan, "Americans' Trust in Media Remains Near Record Low," Gallup, Oct. 18, 2022, accessed Nov. 4, 2022, https://news.gallup.com/poll/403166/americans-trust-media-remains-near-record-low.aspx.

[136] See: Nyhan and Reifler, "When Corrections Fail: The Persistence of Political Misperceptions."

[137] Interview with fact-checking expert, Dec. 1, 2022.

[138] B. Swire, U. K. H. Ecker, and S. Lewandowsky, "The Role of Familiarity in Correcting Inaccurate Information," *Journal of Experimental Psychology. Learning, Memory, and Cognition* 43, no. 12 (2017): 1948-1961, doi: 10.1037/xlm0000422, NLM.

[139] Ethan Porter and Thomas J. Wood, "The Global Effectiveness of Fact-Checking: Evidence from Simultaneous Experiments in Argentina, Nigeria, South Africa, and the UK," *Proceedings of the National Academy of Sciences* 188, no. 37 (2021), https://doi.org/10.1073/pnas.2104235118.

Media Literacy

Media literacy *describes an individual's ability to critically assess a piece of content. It includes the skills required to evaluate a piece of content, as well as an understanding of the structures that produced that content (Table 10).*

Table 10. Media literacy key findings

Media literacy is an effective way to increase resistance to persuasion and manipulation

- In-person media literacy training has been found to be effective across a range of topics, behaviors, and outcomes
- Online media literacy training has been shown to positively affect media use in multiple ways:
  - increase trust in media
  - increase the ability to differentiate real from fake headlines
  - lower people's belief that MDM is accurate
- Online *news* media literacy training may be limited in its ability to counter MDM, but it has been shown to:
  - improve self-perceptions of media literacy
  - effectively reinforce lessons learned from in-person trainings
  - improve the quality of the news that people share online

Source: CNA.

A brief history of media literacy

Renee Hobbs, an American scholar who is today widely cited as a leading media literacy theorist, notes that media literacy draws upon philosophy, sociology, and even literature. She identifies a diverse range of individuals as the "grandparents" of contemporary media literacy, ranging from Gordon Allport, one of the founders of media psychology in the 1930s, to Theodor Adorno, who offered critiques of what he called "the culture industry" in the 1940s.[140] Another major figure in the field is Canadian communication theorist Marshall McLuhan. Writing in the 1960s, McLuhan's work became a cornerstone of media theory. He famously wrote, "The medium is the message," encouraging the study of the communications mediums, rather than the content included within them.[141] Other scholars have emphasized that media literacy

[140] Renee Hobbs, "Grandparents of Media Literacy," *Media Educatation Lab*, 2017, accessed Nov. 20, 2022, https://grandparentsofmedialiteracy.com/.

[141] Der-Thanq Victor Chen, Jing Wu, and Yu-Mei Wang, "Unpacking New Media Literacy," *Journal on Systemics, Cybernetics and Informatics* 9 (2011): 84.

draws inspiration from the ideals of Socratic questioning, critical thinking, and civic engagement, citing philosophers such as Aristotle and Plato.[142]

Research into media literacy, and the development of relevant curricula, began in the US around the 1970s and was led by educators and activists concerned about the negative effects of the media on young people, both in the US and in countries like the UK, Australia, and Canada.[143] The issue began to gain national attention with the 1992 Conference for Media Literacy[144] and the founding of organizations such as the National Association on Media Literacy Education, the Center for Media Literacy, the Alliance for a Media Literate America, and the National Telemedia Council.

Starting in the early 2010s, amid a rise in political polarization and partisan news, researchers began considering ways to promote media literacy outside of K–12 classrooms and undergraduate programs. Early research found that media literacy materials (taking the form of presentations or videos) were effective, and researchers soon began exploring ways that media literacy messages could be integrated into daily news consumption.[145]

The interest in media literacy and its subfields as a tool to counter MDM showed up in literature around 2015, a timeline that corresponds with the rise in MDM in the US ahead of the 2016 elections. Writing in early 2017, Mihailidis and Viotty argued that faced with "the spread of misinformation, the appropriation of cultural iconography, and the willing engagement of mainstream media to perpetuate partisan and polarizing information," media literacy was a valuable response mechanism to help cultivate more critical consumers of media.[146] Across a wide range of partisan and nonpartisan issues, scholars began to call for training or instruction

[142] Bulger and Davison, "The Promises, Challenges, and Futures of Media Literacy."

[143] Ibid.

[144] David M. Considine, "Medita Literacy: National Developments and International Origins," *Journal of Popular Film and Television* 30, no. 1 (2002): 7, doi: 10.1080/01956050209605554.

[145] Donna Chu and Alice Y. L. Lee, "Media Education Initiatives by Media Organizations: The Uses of Media Literacy in Hong Kong Media," *Journalism and Mass Communication Educator* 69 (2014), doi: 10.1177/1077695813517884; Renee Hobbs, *Digital and Media Literacy: A Plan of Action*, Aspen Institute, 2010, https://mediaeducationlab.com/sites/default/files/Hobbs%2520Digital%2520and%2520Media%2520Literacy%2520Plan%2520of%2520Action_0_0.pdf; Emily K. Vraga, Melissa Tully, H. Akin, and H. Rojas, "Modifying Perceptions of Hostility and Credibility of News Coverage of an Environmental Controversy Through Media Literacy," *Journalism* 13, no. 7 (2012): 942–959, https://doi.org/10.1177/1464884912455906.

[146] Paul Mihailidis and Samantha Viotty, "Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in 'Post-Fact' Society," *American Behavioral Scientist* 61, no. 4 (2017): 441–454, doi: 10.1177/0002764217701217, https://journals.sagepub.com/doi/abs/10.1177/0002764217701217.

to help the public navigate online spaces, assess information more critically, and discern fact from fiction.[147]

In what follows, we first provide more background on media literacy, explore the state of the overall research, and end with a brief discussion of how long the benefits of media literacy interventions typically last.

The definition and logic of media literacy

Scholars disagree over the definition of *media literacy*, and most scholars make note of these definitional disagreements in the introductions to their papers. Many scholars cite the National Association on Media Literacy Education's definition: "the ability to access, analyze, create, and act using all forms of communication,"[148] which is closely based on a definition developed by Hobbs.[149] Others highlight the five core concepts that the Center for Media Literacy considers central to media literacy:

> All media messages are constructed; messages are constructed using a creative language with its own rules; different people experience the same message differently; media has embedded values and points of view; and most media messages are organized to gain profit and/or power.[150]

This disagreement about the definitions of media literacy extends to the goals and intended outcomes of media literacy interventions. Bulger and Davison capture some of the ambiguity of these goals by asking: "Is it about instilling confidence? Promoting behavioral change? Or creating new practices of media creation?"[151] A perhaps better representation of this disagreement is provided in a 2012 meta-analysis by Jeong et al., in which the authors are compelled to analyze the effects of media literacy interventions across nine possible outcomes because there is no agreement in the field.[152]

[147] Wei Peng, Sue Lim, and Jingbo Meng, "Persuasive Strategies in Online Health Misinformation: A Systematic Review," *Information, Communication & Society* (2022), doi: 10.1080/1369118X.2022.2085615.

[148] "Media Literacy Defined," National Association of Media Literacy Education, https://namle.net/resources/media-literacy-defined/.

[149] Seth Ashley, Mark Poepsel, and Erin Willis, "Media Literacy and News Credibility: Does Knowledge of Media Ownership Increase Skepticism in News Consumers?" *Journal of Media Literacy Education* 2 (2010), doi: 10.23860/jmle-2-1-4.

[150] "Five Key Questions of Media Literacy Education," *Center for Media Literacy*, 2005, https://www.medialit.org/sites/default/files/14B_CCKQPoster+5essays.pdf.

[151] Bulger and Davison, "The Promises, Challenges, and Futures of Media Literacy," 1-21.

[152] S. H. Jeong, H. Cho, and Y. Hwang, "Media Literacy Interventions: A Meta-Analytic Review," *Journal of Communication* PMC3377317 62, no. 3 (2012): 454-472, doi: 10.1111/j.1460-2466.2012.01643.x, NLM.

Further complicating the landscape, multiple subfields or specialties exist within the field of media literacy, including digital media literacy, news literacy, health literacy, information literacy, new media literacy, and others.[153] Many of these subfields now have their own theories and best practices; for example, news literacy scholars cite four key tenets that need to be communicated in a news literacy program to ensure its efficacy.[154] These tenets, however, are unique to news literacy and do not overlap perfectly with the definitions or principles of media literacy cited above.

Though at times these distinctions feel confusing or arbitrary (e.g., some authors make a point of distinguishing news literacy from news media literacy), the argument for such distinctions is that the skills needed to analyze and contextualize a tweet are different from those needed to analyze and contextualize a *Bloomberg* article. Karadjov and Fleming write, "The thinking behind multiliteracies is that *content is created with different tools and techniques*; hence, different pedagogies are needed to engage with different media" (emphasis added).[155] Hobbs argues that this fragmentation is due in part to diverse stakeholders coming to understand the importance of media literacy in their own disciplines (Figure 8).[156] As one example, doctors have realized the value of teaching individuals how to assess health information online, which has led to the development of science or health media literacy.

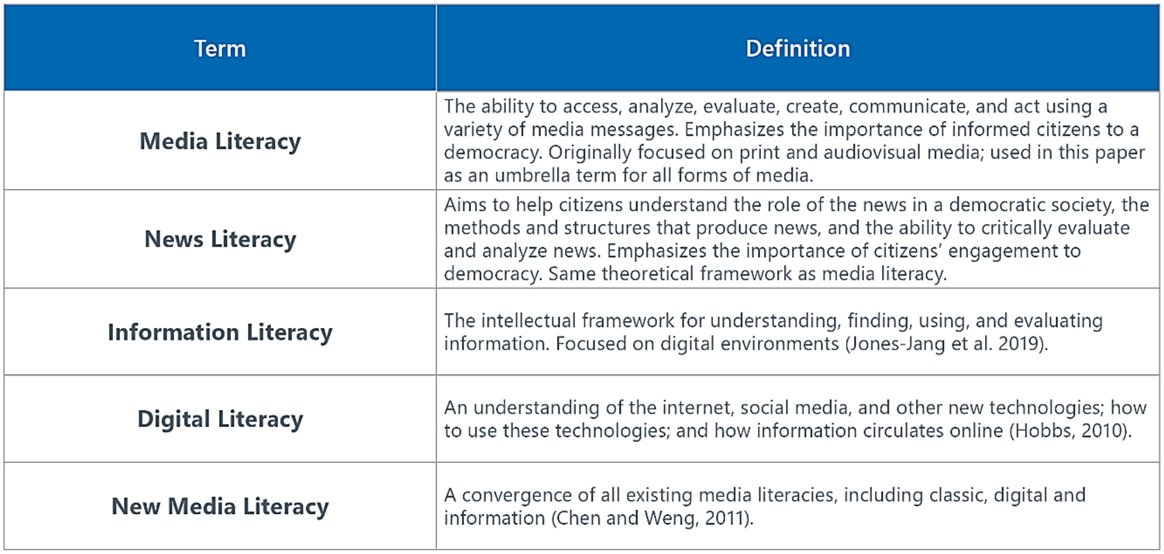
[153] Jennifer Fleming and Christopher Karadjov, "Focusing on Facts: Media and News Literacy Education in the Age of Misinformation," in *Media Literacy in a Disruptive Media Environment*, (Milton Park, UK: Routledge, 2020), 77-93.

[154] Melissa Tully, Emily K. Vraga, and Leticia Bode, "Designing and Testing News Literacy Messages for Social Media," *Mass Communication and Society* 23, no. 1 (2020): 22-46, doi: 10.1080/15205436.2019.1604970.

[155] Fleming and Karadjov, "Focusing on Facts: Media and News Literacy Education in the Age of Misinformation," 80.

[156] Interview with Dr. Renee Hobbs, Dec. 9, 2022.

Figure 8. Different types of media literacy relevant to MDM



Term	Definition
Media Literacy	The ability to access, analyze, evaluate, create, communicate, and act using a variety of media messages. Emphasizes the importance of informed citizens to a democracy. Originally focused on print and audiovisual media; used in this paper as an umbrella term for all forms of media.
News Literacy	Aims to help citizens understand the role of the news in a democratic society, the methods and structures that produce news, and the ability to critically evaluate and analyze news. Emphasizes the importance of citizens' engagement to democracy. Same theoretical framework as media literacy.
Information Literacy	The intellectual framework for understanding, finding, using, and evaluating information. Focused on digital environments (Jones-Jang et al. 2019).
Digital Literacy	An understanding of the internet, social media, and other new technologies; how to use these technologies; and how information circulates online (Hobbs, 2010).
New Media Literacy	A convergence of all existing media literacies, including classic, digital and information (Chen and Weng, 2011).

Source: Some sources combine online, digital, and new media literacy (Jones-Jang et al., 2019).

Despite these disagreements about definitions, goals, and subgoals, media literacy can be thought of as a process or set of skills based on critical thinking.[157] It focuses on equipping individuals with the tools and abilities necessary to critically evaluate a piece of "media," whether that be a tweet, an article, a TV show, or other content. It teaches skills including, but not limited to, asking questions, analyzing sources, assessing bias, and valuing the role of an independent media. In this sense, media literacy is content neutral: it does not advance or counter specific ideas but rather teaches wholly nonpartisan skills.

Given these core similarities and the overlaps between different subfields, this analysis does not focus on one subfield, but instead examines each of the subfields' potential relevance to countering MDM. This focus aligns with scholars like Potter and McDougall, who have proposed the term *dynamic literacies* to bring together the shifting and dynamic definitions of literacy,[158] and Hobbs, who has diagrammed the ways that media literacy and digital literacy overlap (Figure 9).[159] The core question at the heart of this literature review is whether

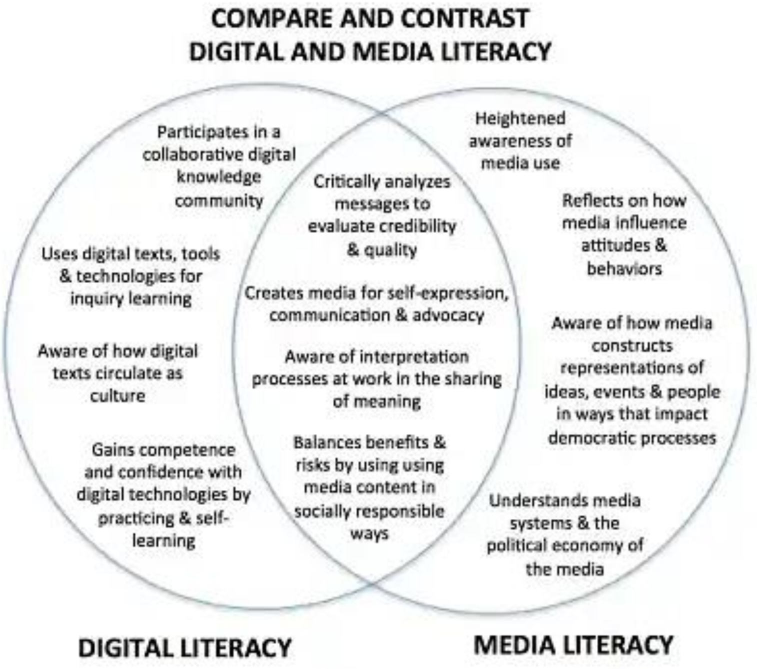
[157] Bulger and Davison, "The Promises, Challenges, and Futures of Media Literacy."

[158] J. Potter and J. McDougall, *Digital Media, Culture and Education*, (London, UK: Palgrave Macmillan/Springer, 2017).

[159] Hobbs, "Digital and Media Literacy: A Plan of Action."

improved critical thinking, analytical skills, and media knowledge improve individuals' resilience to MDM.

Figure 9. An example of the overlap between fields: digital literacy vs. media literacy


COMPARE AND CONTRAST
DIGITAL AND MEDIA LITERACY
Participates in a collaborative digital knowledge community
Uses digital texts, tools & technologies for inquiry learning
Aware of how digital texts circulate as culture
Gains competence and confidence with digital technologies by practicing & self-learning
Critically analyzes messages to evaluate credibility & quality
Creates media for self-expression, communication & advocacy
Aware of interpretation processes at work in the sharing of meaning
Balances benefits & risks by using using media content in socially responsible ways
Heightened awareness of media use
Reflects on how media influence attitudes & behaviors
Aware of how media constructs representations of ideas, events & people in ways that impact democratic processes
Understands media systems & the political economy of the media
DIGITAL LITERACY
MEDIA LITERACY

Source: Renee Hobbs, "Defining Digital Literacy" *MediaLab,* Feb. 10, 2019, https://mediaedlab.com/2019/02/10/defining-digital-literacy-2/.

Finally, scholars highlight two theoretical rationales for how media literacy interventions are effective, which are also largely applicable across subfields. The first is inoculation theory (explored earlier in this paper and expanded upon further below), which suggests that prior exposure helps protect the audience against future attacks. In the context of fake news, media literacy education offers the knowledge and skills to resist or critically interpret fake news stories, helping "inoculate" individuals against MDM's harmful influence. The second is the message interpretation process, which draws upon social cognitive theory and argues that proper interventions or educations can mediate the relationship between exposure to harmful messages and subsequent decision-making. For example, an intervention prior to or after an individual views an MDM post might make them less likely to believe or share the post.[160]

Media literacy and inoculation theory

Media literacy, like inoculation theory, is a preventative or prophylactic intervention, and its original goal was to help protect young people against negative media exposure effects, such as the negative effects media exposure can have on perceptions of body image in adolescent girls. Fleming and Karadjov wrote in 2020 that "the protectionist paradigm has been consistent throughout the history of media literacy education."[161]

The success of inoculation theory, along with that of other pre-emptive refutations, is partly why scholars began applying media literacy theories to MDM. According to Vraga, Kim, and Cook, inoculation theory provides the "theoretical premise" for media literacy interventions' applicability to misinformation. As described by some scholars, media literacy trainings and interventions are examples of "logic-based inoculations."[162] In these framings, media literacy education and critical-thinking interventions almost become types of inoculation.

Not all scholars go as far as Vraga, Kim, and Cook in calling their media literacy programs inoculations, but many cite well-known scholars and specific tenets from the inoculation theory literature. For example, Vraga and Tully note that a media literacy public service announcement may serve a similar role as a reinforcement message in an inoculation campaign, reminding people of the things they learned in earlier media literacy education.[163]

Because media literacy interventions were found to be increasingly effective outside of classrooms, against both inaccurate headlines and biased media, they emerged as a promising preemptive intervention that might (similar to, but distinct from, inoculation efforts) be used to combat MDM.

[160] S. Mo Jones-Jang, Tara Mortensen, and Jingjing Liu, "Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't," *American Behavioral Scientist* 65, no. 2 (2021), doi: 10.1177/0002764219869406.

[161] Fleming and Karadjov, "Focusing on Facts: Media and News Literacy Education in the Age of Misinformation," 79.

[162] Vraga, Kim, and Cook, "Testing Logic-Based and Humor-Based Corrections for Science, Health, and Political Misinformation on Social Media*."*

[163] Vraga and Tully, "Media Literacy Messages and Hostile Media Perceptions: Processing of Nonpartisan Versus Partisan Political Information."

Overall findings

Media literacy interventions can be broken down into two primary categories: (1) in-person educational programs, which take the shape of in-person education, trainings, and presentations, and (2) remote interventions, such as video messages during TV programming, online tips, online training, and tweets promoting media literacy. Typically, the two intervention types are studied separately, although scholars may reference relevant research from the other type of intervention, and some are beginning to analyze the relationship between the two.[164]

Some scholars argue that the broad field of media literacy lacks applicability to the issue of MDM, and that both scholars and trainings should focus on specific sub-literacies. Fleming and Karadjov argue that news media literacy is the only effective subfield,[165] as do Vraga and Tully.[166] In a 2021 study, Jones-Jang et al. found that only information literacy accurately predicted an individual's ability to identify fake news, with news literacy, media literacy, and digital literacy failing to make accurate predictions.[167] In the same year, Sirlin et al. argued that the only subfield of media literacy that improved sharing discernment (i.e., how many true headlines someone shared relative to the number of false headlines they shared) was what they call "procedural news knowledge" (i.e., a clear understanding of professional news operations and procedures).[168] These conflicting findings may explain some of the mixed or inconclusive results cited below; namely, it may be true that specific subtypes of media literacy are necessary to counter specific subtypes of MDM (e.g., health media literacy may be the only, or the most effective, way to counter health-related MDM). Although the field lacks consensus on whether this is the case, these findings indicate potential avenues for further research given some promising early findings about the applicability of these subfields to counteracting misinformation.

[164] Emily K. Vraga, Melissa Tully, and Hernando Rojas, "Media Literacy Training Reduces Perception of Bias," *Newspaper Research Journal* 30, no. 4 (2009): 68-81, doi: 10.1177/073953290903000406.

[165] Fleming and Karadjov, "Focusing on Facts: Media and News Literacy Education in the Age of Misinformation," 80.

[166] Tully, Vraga, and Bode, "Designing and Testing News Literacy Messages for Social Media."

[167] Jones-Jang, Mortensen, and Liu, "Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't."

[168] N. Sirlin et al., "Digital Literacy Is Associated with More Discerning Accuracy Judgments but Not Sharing Intentions," *Harvard Kennedy School (HKS) Misinformation Review* (2021); M.A. Amazeen and Erik Bucy, "Conferring Resistance to Digital Disinformation: The Innoculating Influence of Procedural News Knowledge," *Journal of Broadcasting and Electronic Media* 63, no. 3 (2019): 424.

Though increased interest in the utility of media literacy training in countering MDM can be traced to the same time period (~2016) as the other interventions explored in this literature review, there appears to be less scholarship on this particular issue. As a result, below we include major relevant findings in studies specific and not specific to the problem of MDM, as well as findings from a variety of relevant subtypes of media literacy.

In-person training

Researchers have found in-person media literacy education and training to be effective across a range of topics, behaviors, and outcomes. In a 2012 meta-analysis, Jeong et al. analyzed 51 quantitative studies of in-person educational interventions and found that in-person interventions are effective across a wide range of topics and audiences. Media literacy interventions increased audiences' knowledge of the media, criticism of the media, and awareness of the influence of the media, while reducing "media realism," or the belief that the media's portrayal of events corresponds with events in the real world. Interventions were found to reduce what they described as "risky or antisocial behaviors," increase negative beliefs and negative attitudes toward such behaviors, and increase self-efficacy to avoid such behaviors. They also found that only two moderating variables affected the efficacy of the program: the length of the intervention and the content of the intervention. Longer interventions with greater numbers of sessions were more effective; interventions with more distinct components (e.g., content, medium, grammar, and structure literacy) were less effective.[169] The finding that the age of the audience did not influence the program's efficacy is particularly significant, confirming that media literacy training isn't solely effective with school-age populations.

One lingering area of disagreement relates to the question of who should deliver media literacy training. Durantini, Albarracin, Mitchell, Earl, & Gillette (2006) found that experts were more effective, citing their knowledge, experience, and authority; however, Webel, Okonsky, Trompeta, & Holzemer (2010) found that peers were more effective than nonpeers, potentially because of perceived similarity and identification.[170] This remains an open debate, and one that is being explored primarily in education literature. As one example, a 2017 study found that university students who were given a rationale for why learning is important from their peers (actors posing as young professionals) wrote more effective essays and got significantly better

[169] Jeong, Cho, and Hwang, "Media Literacy Interventions: A Meta-Analytic Review."

[170] Ibid.

final grades than students who were given the same rationale from the course instructor.[171] Additional research is needed on the topic to specifically focus on media literacy training.

Little research has explored the efficacy of in-person MDM-focused media literacy interventions, even though various media literacy trainings have specifically sought to increase participants' resilience to MDM. Hobbs notes that the practice of media literacy is mostly not conducted by scholars but rather by educators who have incorporated media literacy into their classrooms and curricula, which is one explanation for the limited number of scholarly articles exploring in-person interventions and training.[172]

As an example of such a training, the International Research and Exchanges Board (IREX) has developed and implemented a half-day curriculum called "Learn to Discern." The program is aimed at adults and has three units: understanding media, recognizing misinformation and manipulation, and fighting misinformation. Although no peer-reviewed academic literature speaks to the IREX programming's efficacy, the program self-reports results of its efforts. Participants in a Learn to Discern training in Jordan were 44 percent better able to identify and analyze false or manipulative information, had a 14 percent greater sense of control of how they responded to information, and had 65 percent more knowledge about the news industry.[173] Participants in the Learn to Discern training in Ukraine reported they were more likely to cross-check the news, were more confident in their ability to analyze the truthfulness of media content, and were better able to distinguish true from false news. In a survey of participants a month after the Ukrainian training, 80–90 percent of respondents reported using the media literacy behaviors taught in the training, including cross-checking news, looking for facts, and checking sources.[174] In 2017, IREX analyzed the long-term benefits of their training—one of the only attempts to quantitively analyze the long-term outcomes of media literacy training. IREX administered an assessment and survey to a random sample of 207 Ukrainian individuals who had participated in the program 16 months earlier. Compared to a control group matched for gender, age, region, and education level, participants from the Learn to Discern training were 13 percent better at identifying a fake news story and 28 percent more likely to demonstrate sophisticated knowledge of the news industry.[175] Though

[171] Tae S. Shin, John Ranellucci, and Cary J. Roseth, "Effects of Peer and Instructor Rationales on Online Students' Motivation and Achievement," *International Journal of Educational Research* 82 (2017): 184, doi: https://doi.org/10.1016/j.ijer.2017.02.001.

[172] Interview with Dr. Renee Hobbs, Dec. 9, 2022.

[173] Case Study: Learn to Discern in Jordan," IREX, accessed Nov. 18, 2022, https://www.irex.org/project/learn-discern-l2d-media-literacy-training#component-id-783.

[174] Erin Murrock et al., "Winning the War on State-Sponsored Propaganda," *IREX* (2018): 5.

[175] Ibid.

promising, the strength of these findings on the success of the IREX training is undermined by a number of issues (e.g., the fact that the program evaluated itself and the small sample sizes).

Remote interventions

Most scholars agree that remote media literacy interventions, such as tweets, videos, and short online presentations, have some effect on the populations they are seeking to influence, although the content of the interventions and what the interventions seek to affect (behavior, attitudes, or some combination of the two) vary.

- *Online media literacy interventions increased participants' trust in media.* As early as 2009, scholars found that online media literacy interventions could influence individuals' perceptions of the media, decreasing their belief that a headline on a highly political issue (in this case, the Iraq War) was biased.[176] Later studies found that media literacy training successfully reduced hostile interpretations of media content. In one experiment, scholars manipulated subjects' exposure to media literacy training and then presented them with news coverage on biofuels. Exposure to a media literacy video led individuals to rate the news coverage as more credible, while also increasing participants' trust in the news more broadly.[177]
- *News media literacy tweets improve individuals' perceptions of their own media literacy* (sometimes referred to in the literature as self-perceived media literacy (SPML)). This in turn can increase individuals' sense of political efficacy, or the belief that one can understand and effectively participate in the political process.[178]
- *Online interventions can be effective at reinforcing lessons and ideas from in-person interventions*. In one study, scholars manipulated whether individuals were exposed to a news media literacy public service announcement (PSA) immediately before viewing a political program. They looked at two groups: students enrolled in media education courses and students enrolled in a non-media education course. Their findings suggest that the ability of media literacy messages to influence individuals is conditioned by their preexisting media literacy education; specifically, media literacy interventions led

[176] Vraga, Tully, and Rojas, "Media Literacy Training Reduces Perception of Bias."

[177] Vraga et al., "Modifying Perceptions of Hostility and Credibility of News Coverage of an Environmental Controversy Through Media Literacy."

[178] Melissa Tully and Emily K. Vraga, "A Mixed Methods Approach to Examining the Relationship Between News Media Literacy and Political Efficacy," *International Journal of Communication* 12 (2018), 766–787.

to more effective processing of political programs only if students had been exposed to preexisting media literacy education.[179]

We found significant disagreement within the literature about how remote media literacy interventions affect an individual's susceptibility to misinformation. Although some studies found that media or news literacy training improved individuals' resistance or resilience to MDM, others produced mixed or inconclusive results on the efficacy of media and news literacy interventions for countering MDM. This variation is partly due to the diversity of intervention types, the various outcomes being analyzed, and the range of moderating variables that must be considered.

- *Media literacy training can increase participants' ability to distinguish real headlines from fake news headlines.* For example, Guess et al. conducted a 2019 study in which 9,190 people[180] participated in an online survey that included an embedded media literacy intervention; they found that participants' ability to distinguish real headlines from fake news headlines improved by 26.5 percent.[181] The intervention was a replica of the training "Tips to Spot Fake News," developed by Facebook in collaboration with the nonprofit First Draft and previously promoted on Facebook news feeds in 14 countries (see Figure 10 for a screenshot of the training). The intervention improved participants' ability to distinguish real headlines from fake news headlines, and participants still had an improved ability to distinguish real from fake headlines three weeks later, albeit with decreased effectiveness.[182] A second study demonstrating effective media literacy training was developed in 2022 by the Poynter Institute, a nonprofit journalism organization that has developed various digital literacy trainings that claim to have reached 21 million people online.[183] Poynter Institute analyzed the efficacy of a 1-hour online digital literacy program called MediaWise for Seniors, a program specifically designed for Americans over 65 who may be more susceptible to misinformation, due in part to lower levels of digital literacy. The study found that participants' ability to accurately identify a headline as either true or false increased

[179] Emily K. Vraga and Melissa Tully, "Effectiveness of a Non-Classroom News Media Literacy Intervention Among Different Undergraduate Populations," *Journalism & Mass Communication Educator* 71, no. 4 (2016): 440–452, doi: 10.1177/1077695815623399.

[180] Selected to match the demographic and political attributes of the US population.

[181] Andrew M. Guess et al., "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* 117, no. 27 (2020): 15536-15545, doi: doi:10.1073/pnas.1920498117.

[182] Note that this study occurred during a period of high political interest, taking place shortly after the 2018 US midterm elections, which may have increased participants' interest in the topic and made them more attentive to the intervention, thereby increasing its efficacy.

[183] "Digital Media Literacy for All," *Poynter*, accessed Nov. 19, 2022, https://www.poynter.org/mediawise/.

20 percent after the training, and their understanding of digital literacy, their skill levels, and their likelihood of doing research to confirm a headline's veracity also significantly increased following the intervention.[184] Finally, a third study found that an online game titled Fakey increased individuals' ability to identify mainstream from low-credibility news, made them more likely to engage with mainstream sources, and increased their skepticism of low-credibility sources.[185] Fakey was released free to the public in 2018 as both a web platform and app,[186] and its design mimics popular social media platforms such as Facebook and Twitter. It displays a variety of current articles randomly selected from mainstream and low-credibility sources representing both moderate, liberal, and conservative views. Micallef et al. analyzed the results of 8,608 players from around the world who chose to voluntarily play the game.[187] They found that those who played more than one round of the game were better at recognizing whether an article was from a mainstream or low-credibility source, and players continued to improve their performance across multiple rounds of game play. Unfortunately, players who did not perform well in the initial rounds of the game often stopped playing, meaning that those who could benefit the most from receiving help with their media literacy are not receiving it via the game.[188]

[184] Ryan C. Moore and Jeffrey T. Hancock, "A Digital Media Literacy Intervention for Older Adults Improves Resilience to Fake News," *Scientific Reports* 12, no. 1 (2022), doi: 10.1038/s41598-022-08437-0.

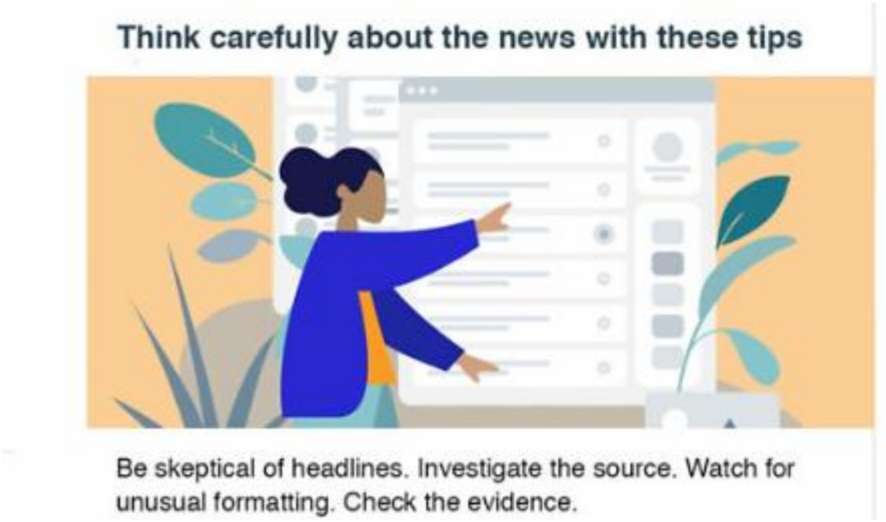
[185] One reason that Fakey and other games may be effective is that they use "system 2 thinking," requiring audiences to think and engage. In general, this active participation is more effective than quick, uninvolved engagement. See: McDougall, Julian, Lee Edwards, and Karen Fowler-Watt, "Media Literacy in the Time of COVID," *Sociologia Della Comunicazione* 62, no. 2. (2021).

[186] Other examples of free online games designed to teach media literacy skills include BBC iReporter, Factitious, and Newsfeed Defender, but none of them report how the games influence users' ability to identify misinformation. See: Daniel Funke and Susan Benkleman, "Factually: Games to Teach Media Literacy," July 18, 2019, American Press Institute, https://www.americanpressinstitute.org/fact-checking-project/factually-newsletter/factually-games-to-teach-media-literacy/.

[187] Nicholas Micallef, Mihai Avram, Filippo Menczer, and Sameer Patil, "Fakey: A Game Intervention to Improve News Literacy on Social Media," *Proceedings of the ACM on Human-Computer Interaction* 5 (2021): 1-27, doi: 10.1145/3449080.

[188] Julian McDougall, Lee Edwards, and Karen Fowler-Watt, "Media Literacy in the Time of COVID," *Sociologia Della Comunicazione* 62, no. 2 (2021).

Figure 10. An example of a tip from the Facebook "Tips to Spot Fake News" program



Think carefully about the news with these tips

Be skeptical of headlines. Investigate the source. Watch for unusual formatting. Check the evidence.

Source: Z. Epstein et al., "Developing an Accuracy-Prompt Toolkit to Reduce COVID-19 Misinformation Online," *Harvard Kennedy School (HKS) Misinformation Review* (2021).

- *Viewing news literacy tips can improve the quality of news that Americans share online.* In a 2021 study, Epstein et al. found that media literacy tips increased sharing discernment by roughly 50 percent.[189] They did so by assessing how 9,000 participants responded to eight experimental treatments, one of which was called the "Tips" treatment and consisted of tips from the same Facebook training that Guess et al. used. Some participants were asked to rank the accuracy of the headlines, while others were asked to rank whether they would share the headline. The study also found increased discernment in sharing headlines.[190]
- *Exposure to a media literacy message significantly lowers the perceived accuracy of misinformation but fails to change individuals' agreement with the misinformation.* A 2022 study by Michael Hameleers found that exposure to a media literacy intervention significantly lowered the perceived accuracy of misinformation, for both evidence-based and fact-free misinformation.[191] However, exposure to the media literacy

[189] Z. Epstein et al., "Developing an Accuracy-Prompt Toolkit to Reduce COVID-19 Misinformation Online," *Harvard Kennedy School (HKS) Misinformation Review* (2021).

[190] Ibid.

[191] Michael Hameleers, "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 & Society* 25, no. 1 (2022): 110-126, doi: 10.1080/1369118X.2020.1764603.

message did not result in lower levels of agreement with the misinformation, raising the question of whether media literacy is failing to tackle the root of the problem: the belief in inaccurate information. Whether interventions change news consumption behavior is also unclear.[192] This gap in knowledge is significant although not unique to media literacy, since linking interventions to behavioral change is challenging in many fields.

- *News literacy interventions have different effects if they are broadly or narrowly constructed, but there is no consensus on which is most effective. As a result, tailoring media literacy messages to the topic or type of misinformation one is hoping to counter may be necessary.* In 2019, a pair of experiments embedded in an online survey tested the efficacy of news literacy tweets at reducing the impact of exposure to misinformation, boosting people's perceptions of their own media literacy, and increasing people's perceptions of media literacy's value in a democratic society. The first experiment exposed participants to broad statements about news literacy, and the second exposed participants to specific tips on how to spot fake news. The comparative results indicate that the effectiveness of broader media literacy interventions significantly differs from narrower concrete steps for countering misinformation; the results were not uniformly successful or unsuccessful across broad or narrow interventions, and the interventions differently affected skepticism of MDM, self-perceived media literacy, and the perceived value of societal media literacy. Understanding these differences is thus critical to tailoring interventions to meet specific goals.[193]
- *News literacy interventions may be limited in their ability to counter MDM.* For example, a 2021 study that sought to analyze the efficacy of news literacy warnings (as compared to real-time corrections from other users) found that a news literacy video—which drew upon best practices in the field, emphasizing individuals' responsibility to evaluate information and providing specific tips on how to best do so—did not have a clear effect on users, neither helping people resist the misinformation nor helping them accept the corrections.[194] Similarly, two additional experiments—which tested three news literacy tweets, each leveraging different images and tactics to stand out in users' Twitter feeds—suggest that even though expert organizations can successfully correct

[192] Guess et al., "A Digital Media Literacy Intervention Increases Discernment Between Mainstream and False News in the United States and India."

[193] Tully, Vraga, and Bode, "Designing and Testing News Literacy Messages for Social Media."

[194] E. K. Vraga, L. Bode, and M. Tully, "The Effects of a News Literacy Video and Real-Time Corrections to Video Misinformation Related to Sunscreen and Skin Cancer," *Health Communication* 37, no. 13 (2020), doi: 10.1080/10410236.2021.1910165, NLM.

misinformation with a single tweet, news literacy tweets may not improve the efficacy of these corrections.[195] This study highlights the difficulty of creating news literacy messages that can effectively break through the clutter on social media.

Preexisting beliefs

Individuals' preexisting beliefs have been found to affect the processing and reception of news literacy messages,[196] but the literature hasn't yet reached consensus about how media literacy interventions influence individuals with different partisan views. More research is needed on this topic to identify whether the effects across different political ideologies are significant enough to limit the feasibility of media literacy interventions for diverse populations.

- *Both conservatives and liberals were more significantly affected by concrete and focused news media literacy programming.* Tully and Vraga demonstrated that different types of news media literacy messages would produce different outcomes. In the study, participants were shown one of four 30-second media literacy PSAs, which each had a tailored focus on a specific component of news media literacy. In a post-test survey, participants responded differently depending on which PSA they had been shown. The worst performing PSA addressed macro-level concerns about news and society and included few concrete examples. Interventions that emphasize the role of the individual and include concrete tips were found to be more effective.[197]

- *Media literacy interventions appear to have unique outcomes for political conservatives and liberals.* For example, a 2009 study by Tully, Vraga, and Rojas that had participants watch a 3-minute media literacy presentation before viewing a neutral Associated Press story on the Fox News website found that liberals rated the content as less biased, while conservatives' ratings did not change. In other words, conservatives demonstrated less responsiveness to the media literacy intervention.[198] The scholars theorized that the liberals may have become aware of their potential bias against Fox News and overcorrected by rating the story as more neutral, while the conservatives

[195] Emily K. Vraga, Leticia Bode, and Melissa Tully, "Creating News Literacy Messages to Enhance Expert Corrections of Misinformation on Twitter," *Communication Research* 49, no. 2 (2020), doi: 10.1177/0093650219898094.

[196] Tully, Vraga, and Bode, "Designing and Testing News Literacy Messages for Social Media."

[197] Emily K. Vraga and Melissa Tully, "Effective Messaging to Communicate News Media Literacy Concepts to Diverse Publics," *Communication and the Public* 1, no. 3 (2016): 305–322, doi: 10.1177/2057047316670409.

[198] Vraga, Tully, and Rojas, "Media Literacy Training Reduces Perception of Bias."

had little reaction because Fox News stories generally align with their beliefs.[199] A second study found that news literacy programming decreased conservatives' perception of the credibility and accuracy of liberal TV hosts, while increasing their perception of both neutral and conservative hosts.[200] And a third study, by van der Meer and Hameleers, found that news literacy messages relying on descriptive norms (as opposed to injunctive norms) were unsuccessful. It made those who opposed immigration (conservatives) even *less* likely to engage with news sources that disagreed with their views. Exposure to the intervention with injunctive norms was effective only among supporters of immigration (liberals).[201]

- *Tailoring news media literacy messages based on party affiliation and prior beliefs can increase the efficacy of a news literacy intervention.* For example, in a second study building on their prior findings, van der Meer and Hameleers adjusted the messages of news literacy interventions to match people's group identifications and prior beliefs. They found that by tailoring interventions depending on partisan beliefs, interventions could be effective for members of all groups. They theorized that targeted messages make individuals feel like they are part of the majority group, less likely to feel attacked, and more likely to see the relevance of the intervention. The outcome is less reactive or defensive behavior.[202] This study highlights the opportunities to reach additional groups, particularly those who may react negatively to some news literacy interventions, by tailoring news media literacy messages around partisan beliefs.

In short, individuals' preexisting beliefs have been found to affect the processing and reception of news literacy messages,[203] but the literature hasn't yet reached consensus about how media literacy interventions affect individuals with different partisan views.

[199] V. Otero, "Media Bias Chart: Version 4.0 - Ad Fontes Media," Ad Fontes Media, 2019, accessed Nov. 2, 2022, https://www.adfontesmedia.com.

[200] Vraga and Tully, "Media Literacy Messages and Hostile Media Perceptions: Processing of Nonpartisan Versus Partisan Political Information."

[201] Toni G. L. A. van der Meer and Michael Hameleers, "Fighting Biased News Diets: Using News Media Literacy Interventions to Stimulate Online Cross-Cutting Media Exposure Patterns," *New Media & Society* 23, no. 11 (2021), doi: 10.1177/1461444820946455.

[202] Ibid.

[203] Tully, Vraga, and Bode, "Designing and Testing News Literacy Messages for Social Media."

Longevity of effect

Unfortunately, few studies look at the longevity of the interventions' effects, and experts note that this area is understudied.[204] This is particularly true of remote interventions, with multiple scholars noting that their study does not provide any information about durability (Moore and Hancock, 2022) or arguing that analyzing the longevity of the remote intervention's effect would be impossible given confounding variables. There are exceptions, most notably Hameleers (2022), which found that participants in Facebook's "Tips to Spot Fake News" online training had an improved ability to tell accurate headlines from inaccurate headlines three weeks later. IREX's 2017 study found that participants had improved abilities to identify fake news two years later—which is promising, but that study had notable weaknesses.[205] It seems that broad, in-person media literacy training also has some lasting effects, given findings that those who had participated in prior training were more receptive to later remote interventions.[206]

[204] Interview with Dr. Renee Hobbs, Dec. 9, 2022.

[205] Murrock et al., "Winning the War on State-Sponsored Propaganda."

[206] Vraga and Tully, "Effectiveness of a Non-Classroom News Media Literacy Intervention Among Different Undergraduate Populations."

Potential Concerns

In reviewing the literature on MDM interventions, we identified three concerns that scholars repeatedly raised: the backfire effect, the continued influence effect, and news cynicism.[207] The *backfire effect* is a worry that counter-MDM interventions and trainings will result in unforeseen consequences (i.e., a backfire). However, the research we reviewed gives little reason to believe that the backfire effect is a major issue. The *continued influence effect* is a worry that MDM cannot be truly eliminated but will continue to exert an influence even after an intervention or training. Research suggests that this concern is legitimate, but given the nature of the continued influence effect (i.e., a failure to *fully* eliminate the influence of MDM), this issue is not a reason to avoid counter-MDM trainings or interventions. The third concern, news cynicism, is slightly more complicated. As the research summarized below highlights, counter-MDM interventions and training may increase skepticism or cynicism of real news. And yet, numerous experts have pointed out that this outcome may not necessarily be bad. Certainly, people doubting the veracity of all information would be a negative outcome, but it may be socially healthy for people to approach all headlines (those from partisan and nonpartisan sites) with a critical eye.

Backfire effects

Backfire effects are thought to occur when a corrective message inadvertently increases belief in, or reliance on, misinformation. Although earlier studies detected backfire effects,[208] recent research has found little evidence of these effects, and studies have been unable to show that they occur under only certain conditions.[209] Moreover, a 2022 study by Swire-Thompson et al. found that the specific language researchers used in earlier experiments may account for the

[207] As noted in the body of this report, the lines between debunking and fact-checking are somewhat porous. As a result, we treated the two in tandem for the purposes of this section on potential concerns. In other words, if a concern appeared in the literature for one of these (e.g., fact-checking), we counted it as a concern for both (e.g., debunking and fact-checking).

[208] Nyhan and Reifler, “When Corrections Fail: The Persistence of Political Misperceptions”; Nyhan, Reifler, and Ubel, “The Hazards of Correcting Myths About Health Care Reform,” 127-132; Christenson, Kreps, and Kriner, “Contemporary Presidency: Going Public in an Era of Social Media: Tweets, Corrections, and Public Opinion.”

[209] Garrett, Nisbet, and Lynch, “Undermining the Corrective Effects of Media-Based Political Fact Checking? The Role of Contextual Cues and Naïve Theory”; Wood and Porter, “The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence”; Porter and Wood, “Political Misinformation and Factual Corrections on the Facebook News Feed: Experimental Evidence.”

backfire effects that were previously documented.[210] That said, concern about backfire effects persists, with researchers focusing on three distinct types: the familiarity backfire effect, overkill backfire effect, and worldview backfire effect.

Familiarity backfire effect

Mentioned in literature on debunking and fact-checking.

The *familiarity backfire effect* refers to people increasing their belief in MDM due to seeing inaccurate information repeated *within* a correction, which increases their propensity to make inaccurate associations.[211] For example, the correction "childhood vaccines do NOT cause autism" increases the likelihood that "vaccines" and "autism" will be linked in the recipient's mind. In the same vein, an experiment conducted during the 2017 French presidential election by Barrera et al. found that fact-checking Marine Le Pen's false statements about immigration may have raised the salience of the immigration issue for voters, thereby serving the purpose of the original disinformation communicated by Le Pen.[212] A literature review on backfire effects notes that the familiarity backfire effect has often been conflated with the more well-established *illusory truth effect*, which refers to increasing belief in MDM due to hearing the MDM repeatedly in the absence of a correction. Whereas the illusory truth effect has robust empirical data support, the authors state that the familiarity backfire effect has little to no empirical support.[213] In fact, recent studies have indicated that repeating the misinformation while refuting it is not only safe but may also make the correction more effective.[214]

Specifically, a 2017 experiment by Ecker and colleagues randomly assigned Australian undergraduate students to one of four conditions, all of which began with reading fictional news reports that contained misinformation about the causes of a fire. The conditions were a no-retraction control condition, a retraction that updated the information without referencing the earlier misinformation (i.e., "After a full investigation, authorities concluded that the fire

[210] Briony Swire-Thompson et al., "The Backfire Effect After Correcting Misinformation Is Strongly Associated with Reliability," *Journal of Experimental Psychology: General* 151 (2022), doi: 10.1037/xge0001131.

[211] Swire-Thompson et al., "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation."

[212] Barrera et al., "Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics." Other studies have found the same with regard to increasing salience and undermining effectiveness (e.g., Nyhan and Reifler, "When Corrections Fail: The Persistence of Political Misperceptions"; Katherine Clayton et al., "Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media," *Political Behavior* 42 (2020), https://doi.org/10.1007/s11109-019-09533-0).

[213] Swire-Thompson et al., "Searching for the Backfire Effect: Measurement and Design Considerations."

[214] Lewandowsky et al., *The Debunking Handbook 2020*; Swire-Thompson et al., "Searching for the Backfire Effect: Measurement and Design Considerations."

was caused by…"), a retraction with a subtle reminder of the false information that simply noted it was incorrect and provided an accurate explanation, and a retraction that explicitly repeated the earlier misinformation followed by the correction. Results showed that corrections were more effective when they explicitly repeated the misinformation. The authors suggest that these sorts of explicit-reminder retractions can make the falsity of the misinformation more salient, leading participants to update their mental models immediately.[215]

In short, repeated exposure to misinformation can increase people's belief in its veracity when no correction is presented, but this sort of backfire effect rarely (if ever) occurs when the misinformation is followed by a clear, well-crafted correction.[216]

Overkill backfire effect

Mentioned in literature on debunking and fact-checking.

The *overkill backfire effect* refers to the idea that providing "too many" counterarguments against a false claim might produce unintended effects or even backfire. Although few studies have directly examined this effect, a recent study found no evidence for this effect and instead concluded that a greater number of relevant counterarguments generally leads to greater reduction of misconceptions.[217] Specifically, Ecker and colleagues conducted three laboratory experiments with Australian undergraduate students to explore whether a greater number of counterarguments would result in stronger reduction of belief in equivocal claims, or whether an overkill backfire effect might occur. The researchers defined *equivocal claims* as claims deemed to be false or likely false based on available evidence but sufficiently plausible about content that would be unfamiliar to many participants (i.e., the impact of brain training on intelligence). The experiments used varying numbers of counterarguments, as well as various combinations of arguments deemed to be strong, weak, or irrelevant. Results showed that as long as the counterarguments were relevant, using more counterarguments led to stronger reduction in the belief in the misinformation. The authors note that future research should investigate whether more counterarguments are effective in refuting claims that are strongly congruent with a person's worldview (a topic that we discuss below).[218]

[215] Ecker, Hogan, and Lewandowsky, "Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction?"

[216] Swire-Thompson et al., "Searching for the Backfire Effect: Measurement and Design Considerations."

[217] Lewandowsky et al., *The Debunking Handbook 2020*.

[218] Ullrich Ecker et al., "Refutations of Equivocal Claims: No Evidence for an Ironic Effect of Counterargument Number," *Journal of Applied Research in Memory and Cognition* 8, no. 1 (2019), https://doi.org/10.1016/j.jarmac.2018.07.005.

Worldview backfire effect

Mentioned in literature on inoculation, debunking, and fact-checking.

The notion of a worldview backfire effect derives from the motivated reasoning literature, which asserts that a person's ideology influences how they process information, and that they will evaluate information that counters their existing beliefs more critically than confirming information.[219] Following this logic, researchers have speculated that information that counters a person's preexisting beliefs may lead the individual to generate counterarguments consistent with their preexisting views, resulting in *stronger* belief in the original misinformation after receiving a retraction.[220] Research on the influence of preexisting beliefs on correction acceptance and retention is related to worldview backfire effect, but the backfire effect theorizes a specific type of response (i.e., generation of a counterargument and recommitment to preexisting beliefs). Early evidence supported the worldview backfire effect, but recent research suggests that—although worldview backfire effects may emerge under specific conditions—concern about this phenomenon may be overblown.[221]

Initial support for a worldview backfire effect was derived from a 2010 landmark study by Nyhan and Reifler (2010) that examined misperceptions about the presence of weapons of mass destruction (WMD) in Iraq. The study found strong evidence that when presented with facts correcting the misinformation, subjects (especially self-identified conservatives) doubled down on their misperceptions. Subsequent research corroborated this finding, including studies showing that Republican subjects became *more* opposed to environmental regulation after seeing evidence of scientific consensus on climate change, that parents who were wary of vaccinating their children were less willing to do so after receiving information on vaccine safety, and that recipients who feared death panels under the Affordable Care Act became more entrenched in their views after receiving a correction.[222]

[219] Briony Swire-Thompson et al., "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation."

[220] Swire et al., "Processing Political Misinformation: Comprehending the Trump Phenomenon"; Swire-Thompson et al., "Searching for the Backfire Effect: Measurement and Design Considerations."

[221] Lewandowsky et al., *The Debunking Handbook 2020*.

[222] Studies cited in Wood and Porter, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence," 135-163. Studies include Nyhan and Reifler, "When Corrections Fail: The Persistence of Political Misperceptions"; P. S. Hart and E. C. Nisbet, "Boomerang Effects in Science Communication: How Motivated Reasoning and Identity Cues Amplify Opinion Polarization About Climate Mitigation Policies," *Communication Research* 39, no. 6 (2012): 701–23, doi:10.1177/0093650211416646; Brenda Nyhan, J. Reifler, S. Richey, and G. L. Freed, "Effective Messages in Vaccine Promotion: A Randomized Trial," *Pediatrics* 133, no. 4 (2014): e835–42, doi:10.1542/peds.2013-2365; Adam J. Berinsky, "Rumors and Health Care Reform: Experiments in Political Misinformation," *British Journal of Political Science* (June 2015): 1–22, doi:10.1017/S0007123415000186; Nyhan, Reifler, and Ubel, "The Hazards of Correcting Myths About Health Care Reform."

Subsequent research, however, has shown that people are capable of changing their views even when the corrections conflict with their partisan political views. For example, the Swire et al. (2017) study discussed earlier, which explored whether belief in misinformation or factual information depended on whether or not it stemmed from a politically polarizing source (e.g., Donald Trump), found no evidence of a worldview backfire effect. The authors conjectured that the kinds of misinformation presented in the study did not resonate strongly enough to create a notable backfire effect, referencing prior studies that suggest that topics create more of a backfire effect if participants feel strongly about them.[223]

In a similar vein, Wood and Porter explored whether the salience of a given topic, as well as recipients' partisan affiliation, influences the worldview backfire effect. Four separate online studies with over 8,000 US subjects provided information and corrections on 36 commonly misunderstood policy topics (e.g., crime rates, taxes, immigration, abortion, climate change). The experiments used actual statements made by Democratic and Republican political leaders, involving a variety of information and correction types and sources (e.g., newspaper article excerpts, fictitious news articles, neutral corrective data from governmental sources). Results showed a backfire effect for only one issue—the presence of WMD in Iraq, using the same corrective information that had been used in the earlier Nyhan and Reifler study. However, when subjects were presented with a less elaborate survey item to assess their beliefs, no backfire was detected.[224] Wood and Porter conjectured that a worldview backfire effect may be triggered by certain circumstances, including contentious ideological issues or survey question wording.[225] Whether Wood and Porter's report was peer-reviewed is unclear, although both authors served on the expert panel that developed *The Debunking Handbook*, and this study was cited as providing evidence against a worldview backfire effect in that book. Recent research found that the worldview backfire effect may be primarily a function of item reliability,[226] confirming Wood and Porter's views that this effect may not be a "real thing" or may occur only under certain conditions. Specifically, the study showed that less reliable items backfire at a substantially higher rate than more reliable items, which the authors note is consistent with previous work showing that backfire effects are more often elicited with less reliable single-item measures compared to more reliable multi-item measures. These findings

[223] Swire et al., "Processing Political Misinformation: Comprehending the Trump Phenomenon."

[224] The Nyhan & Reifler (2010) original survey question read: "Immediately before the US invasion, Iraq had an active weapons of mass destruction program, the ability to produce these weapons, and large stockpiles of WMD, but Saddam Hussein was able to hide or destroy these weapons right before US forces arrived." The less elaborate survey item used by Wood & Porter read: "Following the US invasion of Iraq in 2003, US forces did not find weapons of mass destruction."

[225] Wood and Porter, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence."

[226] Swire-Thompson et al., "The Backfire Effect After Correcting Misinformation Is Strongly Associated with Reliability."

indicate that future research on the backfire effect should use multi-item measures to boost reliability.[227]

Attempting to shed light on the issue, a 2020 literature review on the worldview backfire effect notes that such effects have been found almost exclusively in political or attitudinal subgroups. Indeed, worldview backfire effects have been found in US studies of partisan issues such as vaccine safety and climate change. Unfortunately, these findings have often been over-generalized. In reality, numerous studies have failed to observe a worldview backfire effect, and current scholarship suggests that it does not exist, is difficult to elicit on the larger group level, or is extremely item, situation, or individual specific.[228]

In inoculation theory, the backfire effect occurs when an inoculation inadvertently increases belief in, or reliance on, the very MDM that the inoculation was intended to counter. For example, after receiving information about the scientific consensus on global warming, participants with strong support for unregulated markets (a proxy for conservativism) became less accepting of climate change."[229] In theory, inoculation interventions are meant to avoid the backfire effect by exposing people to logical fallacies inherent in misleading communications before they form strong views,[230] and some inoculation theory scholars hold that no consistent evidence shows that inoculation theory increases misbeliefs.[231] However, the literature is unresolved on how people respond to inoculation techniques if they already hold strong polarizing views. Recent evidence suggests that inoculation is an effective intervention for controversial topics, and potentially even for individuals who already hold strong beliefs. However, if further research demonstrates that inoculation interventions are the most effective in populations that do not already have entrenched beliefs, or that those with entrenched beliefs are likely to double down on these beliefs *because* of interventions, then this finding would clearly limit the potential applications of inoculation interventions.

In short, research indicates that backfire effects are not common and may, in fact, be a function of study design. MDM scholars emphasize the importance of issuing corrections, which have been shown to influence beliefs.[232]

[227] Ibid.

[228] Swire-Thompson et al., "Searching for the Backfire Effect: Measurement and Design Considerations"; Swire-Thompson et al., "The Backfire Effect After Correcting Misinformation Is Strongly Associated with Reliability."

[229] Cook, Lewandowsky, and Ecker, "Neutralizing Misinformation Through Inoculation: Exposing Misleading Argumentation Techniques Reduces Their Influence."

[230] Toby Bolsen and James N. Druckman, "Counteracting the Politicization of Science," *Journal of Communication* 65, no. 5 (2015): 745-769, doi: https://doi.org/10.1111/jcom.12171.

[231] Interview with Dr. Jon Roozenbeek, Nov. 22, 2022.

[232] Swire-Thompson et al., "Searching for the Backfire Effect: Measurement and Design Considerations."

Continued influence effect

Mentioned in literature on debunking and fact-checking.

Although corrections can reduce people's belief in false information, the misinformation often continues to influence their thinking even after credible corrections are processed. This observed pattern is known as the *continued influence effect*. When this effect occurs, people report the correction accurately and state that they no longer believe the original misinformation when asked directly, but the misinformation may resurface when answering inferential questions.[233] In an example adapted from one study, participants may receive information about a city water system shutdown after numerous fish deaths in the waterway, which were believed to be caused by contaminants from a nearby mining company. The story then reports that no industrial contaminants were found. Although respondents may indicate their belief in the retraction, they may later respond to inferential questions (e.g., "How could such incidents be prevented in the future?") by indicating that the mining company played a role in the incident.[234] Studies have shown that misinformation typically leaves small lingering effects like these, even though various debunking strategies and messages reduce the belief in misinformation.

A 2019 meta-analysis of 32 experiments confirmed that some level of the continued influence effect was present, even after participants received a correction. The analysis explored the extent to which a correction reverts participants' attitudes and beliefs "back to baseline." The assumption was that a fully effective corrective message would be evidenced by ***no*** significant difference between the beliefs of those exposed to the correction and those who were never exposed to the misinformation in the first place. Results revealed that, overall, correction of misinformation does not entirely revert people's attitudes and beliefs to their baseline levels. Rather, misinformation continues to have a small, although significant, effect, thus affirming prior research showing the continued influence effect.[235]

[233] Ecker and Antonio, "Can You Believe It? An Investigation into the Impact of Retraction Source Credibilty on the Continued Influence Effect"; Ecker, Hogan, and Lewandowsky, "Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction?"; Lewandowsky et al., *The Debunking Handbook 2020*; Swire et al., "Processing Political Misinformation: Comprehending the Trump Phenomenon."

[234] Ecker and Antonio, "Can You Believe It? An Investigation into the Impact of Retraction Source Credibilty on the Continued Influence Effect."

[235] Nathan Walter and Riva Tukachinsky, "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, no. 2 (2020): 155-177, doi: 10.1177/0093650219854600.

Some of the studies summarized earlier in this report included findings related to the continued influence effect, as noted below:

- The 2017 Ecker et al. study of misinformation about causes of a fire found that corrections were more effective when they explicitly repeated the misinformation. These results may suggest that this approach made the falsity of the misinformation more salient, leading participants to update their mental models of the event immediately upon processing the retraction.[236]
- The 2021 Ecker and Antonio study sought to determine the relative influence of perceived trustworthiness versus credibility of the retraction source, as well as the person's lack of belief in the retraction, on the continued influence effect. Their studies used five conditions for the retraction source representing various combinations of trustworthiness and credibility (low expertise/low trustworthiness, low expertise/high trustworthiness, and so on). The results indicated that the credibility of the retraction source had a small but significant influence on the continued influence effect, but belief in the retraction was significantly lower than belief in the original misinformation. The authors suggest that this effect may be due to the brevity of the retractions, or the fact that retractions are more closely scrutinized because they contradict the original misinformation.[237]
- The Swire et al. study, which explored whether belief in misinformation or factual information depended on whether it stemmed from a politically polarizing source, found that members of both parties changed their beliefs post-explanation, but the belief change was not sustained.[238]
- Chan et al.'s meta-analysis showed that detailed debunking messages are generally associated with a stronger immediate debunking effect, but these debunking effects have not always translated into reduced continued influence effect across studies.[239]

[236] Ecker, Hogan, and Lewandowsky, "Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction?"

[237] Ecker and Antonio, "Can You Believe It? An Investigation into the Impact of Retraction Source Credibilty on the Continued Influence Effect."

[238] Swire et al., "Processing Political Misinformation: Comprehending the Trump Phenomenon."

[239] Chan et al., "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation."

News cynicism

Mentioned in literature on inoculation, debunking, fact-checking, and media literacy.

At present, researchers have not reached consensus over whether MDM interventions make people more cynical of the news, despite the clearly significant implications of such a finding. Some scholars, however, are concerned that such interventions may increase citizens' skepticism and cynicism about the veracity of *all* information online, not just false information. Importantly, research on this topic is complicated by the language used to describe the phenomena. For example, some literature refers to this issue as "real news skepticism," a worry that people will demonstrate increased doubt in the accuracy of "real news." But this language raises a range of issues. To begin, one expert in the field noted that clarity is lacking regarding what constitutes "real news."[240] Additionally, increased discerning and thoughtful skepticism of factually accurate news may be a social good, though outright rejection of factually accurate news would be a problem. In fact, the development of a healthy skepticism that people learn to apply after an MDM intervention may be broadly beneficial. Given the lack of specificity about what "real news" means and the possibility that skepticism might be good, we have chosen to discuss this issue under the heading "news cynicism."

Media literacy's most notable and heavily cited critic is Danah Boyd. She argues that media literacy is counterproductive because it encourages individuals to be skeptical of information and the news when the general public is *already* skeptical.[241] Though her work is cited frequently by other media literacy scholars (including Tully, Vraga, and Bode (2020), Bulger and Davison (2018), and many others), it is not grounded in a quantitative study. Hameleers had similarly theoretical (i.e., not grounded in a quantitative study) concerns, worrying that media literacy interventions might increase consumers' skepticism and cynicism, especially if they already overestimate the presence of fake news.[242] Two of the leading experts in the field, Renee Hobbs and Emily Vraga, highlight an important difference between skepticism and cynicism. Both argue that skepticism is valuable, even of mainstream sources such as the *New York Times* or CNN.[243] An excellent distinction between the two is provided in the 2007 book

[240] Interview with Dr. Jon Roozenbeek, Nov. 22, 2022.

[241] Danah Boyd, "Did Media Literacy backfire?" *Journal of Applied Youth Studies* 1, no. 4 (2017), 83-89.

[242] Hameleers, "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."

[243] Interview with Dr. Emily Vraga, Dec. 6, 2022; Interview with Dr. Renee Hobbs, Dec. 9, 2022.

UnSpun: Finding Facts in a World of Disinformation: "The skeptic demands evidence, and rightly so. The cynic assumes that what he or she is being told is false."[244]

This concern about news cynicism is only partly supported by empirical research. A 2008 multi-year study by Mihailidis of 239 undergraduate students enrolled in a media literacy course found that the study increased news cynicism. The course increased students' ability to comprehend, evaluate, and analyze media messages in print, video, and audio, with students noting that the education helped them look more deeply into the media and made them feel more informed in general. However, in focus group discussions on media relevance and credibility, the students expressed considerable negativity about media's role in society. Their new skills made them "cynical and defensive."[245] Importantly, Mihailidis claims that this problem is not inherent in media literacy but is instead born from the way this training was carried out. He argues that only teaching critical analysis skills is inadequate and will lead to cynical consumers. Instead, courses need to be "civically and democratically-oriented," providing participants with an understanding of the value of free and diverse media to a democratic society. Hobbs noted that 15 years later, this study is still cited as an example of how *not* to do media literacy training,[246] so even though this study proved that media literacy can indeed produce news skepticism, it may be related to how interventions are carried out and not media literacy in its entirety. Tully and Vraga conducted interviews with various survey participants for a 2018 study and note in their conclusion: "At worst, news media literacy could promote cynicism and apathy toward news and politics, making people less likely to engage with news or politics."[247]

A mixed result comes from a 2010 study by Ashley, Poespel, and Willis, which specifically sought to study the linkages between media literacy and news credibility and found that the effects may not always be negative. They designed a simple between-subjects experiment to determine how increased knowledge affects judgments of message credibility. Participants received either a print article discussing media ownership or the control (a set of nature poems), and they then reviewed four articles from different mainstream news outlets (ABC, MSNBC, the *Wall Street Journal*, and the *New York Times*). The articles covered a range of topics and were chosen to demonstrate objectivity, balance, and independence. Participants rated the articles for truthfulness, superficiality, general accuracy, and completeness. Simply reading

[244] Brooks Jackson and Kathleen Hall Jamieson, *UnSpun: Finding Facts in a World of Disinformation*, (Toronto, Ontario: Random House Trade Paperbacks, 2007).

[245] Paul Mihailidis, "Beyond Cynicism: How Media Literacy Can Make Students More Engaged Citizens," (PhD diss., University of Maryland, College Park, 2008), 182.

[246] Interview with Dr. Renee Hobbs, Dec. 9, 2022.

[247] Tully and Vraga, "A Mixed Methods Approach to Examining the Relationship Between News Media Literacy and Political Efficacy," 17.

about media ownership lowered participants' perceptions of superficiality and general accuracy of the news, although not their ratings of its truthfulness or completeness.[248] Although this experiment was narrow and had a limited sample, the study found that educational approaches can affect judgments of the credibility of accurate headlines, reinforcing some of the concerns highlighted by other scholars.

Guess et al. found that the "Tips to Detect Fake News" intervention successfully increased participants' ability to identify MDM, but it also negatively affected participants' belief in mainstream news. Specifically, it reduced the perceived accuracy of mainstream news headlines. The decreased belief in mainstream headlines was significantly smaller than the decreased belief in false news headlines, however, showing that increasing skepticism primarily affected perceptions of false news. Similarly, Guess et al. found in 2020 that exposure to an intervention had a small negative effect on participants' belief in mainstream news.[249]

In an experiment testing inoculation with the Go Viral! game, Basol et al. found no significant difference (in overall pre- and post-manipulativeness scores) for two out of three real news items, and a small but significant increase in the perceived manipulativeness of one real item.[250] In other words, in two out of three cases, people were not more skeptical of real news after going through the inoculation process, but in one case, people were more skeptical of a real news item after they were inoculated against misinformation. This finding suggests that inoculation could *potentially* make people more suspicious of real news, but the effect appears to be limited. If proven, this finding could have negative implications for the post-truth era because it would make people more skeptical of news in general (as opposed to generating the skills to identify and reject misinformation).

Interestingly, Roozenbeek et al. did not concur with this finding; they found that intervention does not increase general skepticism for real news.[251] Looking deeper at this finding, Roozenbeek et al. did find some skepticism of real news, but concluded that these effects were due to an interaction with the specific item set in the experiment, and not to a negative effect of inoculation on real news.

[248] Ashley, Poepsel, and Willis, "Media Literacy and News Credibility: Does Knowledge of Media Ownership Increase Skepticism in News Consumers?"

[249] Guess et al., "A Digital Media Literacy Intervention Increases Discernment Between Mainstream and False News in the United States and India."

[250] Basol et al., "Towards Psychological Herd Immunity: Cross-Cultural Evidence for Two Prebunking Interventions Against COVID-19 Misinformation."

[251] Jon Roozenbeek et al., "Disentangling Item and Testing Effects in Inoculation Research on Online Misinformation: Solomon Revisited," *Educational and Psychological Measurement* 81, no. 2 (2021): 340-362, doi: 10.1177/0013164420940378.

Finally, a 2022 study by Moore and Hancock produced mixed results that further complicate the landscape. In their pre-test survey of individuals, they expected to see individuals identify most headlines as accurate, in line with truth-default theory. This theory argues that people are truth-biased (i.e., they interpret most messages as true), so they are more likely to accurately detect true than false messages (an empirical phenomena known as the "veracity effect"). Instead, participants in this 2022 study identified the majority of headlines as inaccurate, validating other studies that have found individuals may not be truth-biased about the news. The digital media intervention in this study actually had a more significant influence on individuals' ability to accurately identify *real news*, with participants' ability to detect true news increasing 36 percent and their ability to detect fake news increasing only 7 percent. In other words, the intervention helped individuals correctly identify headlines as accurate, which could help *decrease* their cynicism. Moore and Hancock explicitly highlight the potential trade-offs of digital media literacy in their conclusion:

> Given that most of the news individuals encounter in daily life is not false, it may be undesirable to make people more accurate at judging the accuracy of content that constitutes a small fraction of their news diet (false news) at the expense of content that makes up a substantially larger portion (true news).[252]

That said, not all findings have been negative. In addition to the inconclusive or mixed effects highlighted in the Ashley, Poepsel, and Willis and Guess et al. studies above, Vraga, Tully, and Rojas found that news literacy trainings increased individuals' trust in the news and increased their sense that media covered contentious issues fairly.[253]

[252] Moore and Hancock, "A Digital Media Literacy Intervention for Older Adults Improves Resilience to Fake News."

[253] Vraga et al., "Modifying Perceptions of Hostility and Credibility of News Coverage of an Environmental Controversy Through Media Literacy."

Conclusion

The threat of foreign adversary messaging—and particularly foreign adversary MDM—is a pressing national security concern, and the US government must act decisively to protect US servicemembers from this malign influence.

In support of that goal, this ONR-sponsored paper has reviewed the evidence-based literature on counter-MDM interventions. We have offered a plain-language explanation of four types of counter-MDM interventions:

- *Inoculation*: The practice of exposing individuals to persuasive messages containing weakened arguments that threaten an attitude or belief in order to "inoculate" them against stronger persuasive messages and attacks on this attitude or belief in the future.
- *Debunking*: The use of a concise correction to MDM that demonstrates that the prior message or messaging campaign was inaccurate.
- *Fact-checking*: A journalistic practice designed to reject clearly false claims with empirical evidence from neutral or unimpeachable sources.[254]
- *Media literacy*: An individual's ability to critically assess a piece of content, including the skills required to evaluate a piece of content and an understanding of the structures that produced that content.

In this paper, we discussed the origins and logic of each intervention, summarized overall research findings, identified issues of ongoing analysis, and briefly explored how long each type of intervention can be expected to last.

Though this review was completed in support of a broader project—whose goal is to recommend a single intervention (or suite of interventions) that the US government might adopt to protect servicemembers from malign foreign influence—it stands alone as a useful primer for those hoping to understand the state of research on these issues.

More specific guidance—in the form of best practices, a more systematized assessment of applicability to military populations, and recommendations for near-term implementation—can be found in the companion report: *Protecting Servicemembers from Foreign Influence: A Counter-MDM Toolkit*.

[254] Interview with fact-checking subject matter expert, Dec. 1, 2022.

Figures

Tables

Abbreviations

AGW	anthropogenic global warming
CDC	Centers for Disease Control and Prevention
CIA	Central Intelligence Agency
DOD	Department of Defense
DROG	Disinformation Intervention Model
IREX	International Research and Exchanges Board
MDM	mis-/dis-/mal-information
NATO	North Atlantic Treaty Organization
ONR	Office of Naval Research
PSA	public service announcement
SPML	self-perceived media literacy
UK	United Kingdom
UN	United Nations
US	United States
WHO	World Health Organization
WMD	weapons of mass destruction

References

Aird, M. J., U. K. H. Ecker, B. Swire, A. J. Berinsky, and S. Lewandowsky. "Does Truth Matter to Voters? The Effects of Correcting Political Misinformation in an Australian Sample." *Royal Society Open Science* PMC6304148 5, no. 12 (2018): 180593. doi: 10.1098/rsos.180593. NLM.

Amazeen, Michelle A, and Erik Bucy. "Conferring Resistance to Digital Disinformation: The Innoculating Influence of Procedural News Knowledge." *Journal of Broadcasting and Electronic Media* 63, no. 3, (2019): 415-432.

Amazeen, Michelle A., Emily Thorson, Ashley Muddiman, and Lucas Graves. "Correcting Political and Consumer Misperceptions: The Effectiveness and Effects of Rating Scale Versus Contextual Correction Formats." *Journalism & Mass Communication Quarterly* 95, no. 1 (2018): 28-48. doi: 10.1177/1077699016678186.

Ashley, Seth, Mark Poepsel, and Erin Willis. "Media Literacy and News Credibility: Does Knowledge of Media Ownership Increase Skepticism in News Consumers?" *Journal of Media Literacy Education* 2 (2010). doi: 10.23860/jmle-2-1-4.

Banas, John A., and Stephen A. Rains. "A Meta-Analysis of Research on Inoculation Theory." *Communication Monographs* 77, no. 3 (2010): 281-311.

Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. "Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber?" *Psychological Science* 26, no. 10 (2015): 1531-1542. doi: 10.1177/0956797615594620. https://journals.sagepub.com/doi/abs/10.1177/0956797615594620.

Barrera, Oscar et al. "Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics." *Journal of Public Economics* 182 (2020). https://doi.org/10.1016/j.jpubeco.2019.104123.

Basol, Melisa, Jon Roozenbeek, Manon Berriche, Fatih Uenal, William P. McClanahan, and Sander van der Linden. "Towards Pychological Herd Immunity: Cross-Cultural Evidence for Two Prebunking Interventions Against COVID-19 Misinformation." *Big Data & Society* 8, no. 1 (2021): 1.

Benegal, Salil D., and Lyle Scruggs. "Correcting Misinformation About Climate Change: The Impact of Partisanship in an Experimental Setting." *Climatic Change* 148, no. 1 (2018): 61-80. https://EconPapers.repec.org/RePEc:spr:climat:v:148:y:2018:i:1:d:10.1007_s10584-018-2192-4.

Berinsky, Adam J. "Rumors and Health Care Reform: Experiments in Political Misinformation." *British Journal of Political Science*, (June 2015): 1–22. doi:10.1017/S0007123415000186.

Bode, Leticia, Emily. K. Vraga, and Melissa Tully. "Do the Right Thing: Tone May Not Affect Correction of Misinformation on Social Media." *Harvard Kennedy School (HKS) Misinformation Review* (2020). https://doi.org/10.37016/mr-2020-026.

Bolsen, Toby, and James N. Druckman. "Counteracting the Politicization of Science." *Journal of Communication* 65, no. 5 (2015): 745-769. doi: https://doi.org/10.1111/jcom.12171.

Boyd, Danah. "Did Media Literacy Backfire?" *Journal of Applied Youth Studies* 1, no. 4 (2017).

Brenan, Megan. "Americans' Trust in Media Remains Near Record Low." Gallup. Oct. 18, 2022. Accessed Nov. 4, 2022. https://news.gallup.com/poll/403166/americans-trust-media-remains-near-record-low.aspx.

Bulger, Monica, and Patricia Davison. "The Promises, Challenges, and Futures of Media Literacy." *Journal of Media Literacy Education* 10 (2018): 1-21.

"Case Study: Learn to Discern in Jordan." IREX. Accessed Nov. 18, 2022. https://www.irex.org/project/learn-discern-l2d-media-literacy-training#component-id-783.

Chan, Man-Pui Sally, Christopher R. Jones, Kathleen Hall Jamieson, and Dolores Albarracín. "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation." *Psychological Science* PMC5673564 28, no. 11 (2017): 1531-1546. doi: 10.1177/0956797617714579. NLM.

Chen, Der-Thanq Victor, Jing Wu, and Yu-Mei Wang. "Unpacking New Media Literacy." *Journal on Systemics, Cybernetics and Informatics* 9 (2011): 84-88.

Christenson, Dino P., Sarah E. Kreps, and Douglas L. Kriner. "Contemporary Presidency: Going Public in an Era of Social Media: Tweets, Corrections, and Public Opinion." *Presidential Studies Quarterly* 51, no. 1 (2021): 151-165. doi: https://doi.org/10.1111/psq.12687.

Chu, Donna and Alice Y. L. Lee. "Media Education Initiatives by Media Organizations: The Uses of Media Literacy in Hong Kong Media." *Journalism and Mass Communication Educator* 69 (2014), doi:10.1177/1077695813517884.

Clayton, Katherine et al. "Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media." *Political Behavior* 42 (2020), https://doi.org/10.1007/s11109-019-09533-0.

Compton, Josh, Sander van der Linden, John Cook, and Melisa Basol. "Inoculation Theory in the Post-Truth Era: Extant Findings and New Frontiers for Contested Science, Misinformation, and Conspiracy Theories." *Social and Personality Psychology Compass* 15, no. 6 (2021).

Considine, David M. "Medita Literacy: National Developments and International Origins." *Journal of Popular Film and Television* 30, no. 1 (2002): 7-15. doi: 10.1080/01956050209605554.

Cook, John, Stephan Lewandowsky, and Ullrich K. H. Ecker. "Neutralizing Misinformation Through Inoculation: Exposing Misleading Argumentation Techniques Reduces Their Influence." *PLOS ONE* 12, no. 5 (2017): e0175799. doi: 10.1371/journal.pone.0175799. https://doi.org/10.1371/journal.pone.0175799.

Dai, Yue Nancy, Wufan Jia, Lunrui Fu, Mengru Sun, and Li Crystal Jiang. "The Effects of Self-Generated and Other-Generated eWOM in Inoculating Against Misinformation." *Telematics and Informatics* 71 (2022). doi: 101835.

Deryugina, Tatyana, and Olga Shurchkov. "The Effect of Information Provision on Public Consensus About Climate Change." *PLOS ONE* 11, no. 4 (2016): e0151469. doi: 10.1371/journal.pone.0151469. https://doi.org/10.1371/journal.pone.0151469.

Dickey, Colin. "The Rise and Fall of Facts." *Columbia Journalism Review* (fall 2019). Accessed Nov. 4, 2022. https://www.cjr.org/special_report/rise-and-fall-of-fact-checking.php.

"Digital Media Literacy for All." *Poynter*. Accessed Nov. 19, 2022. https://www.poynter.org/mediawise/.

Ecker, Ullrich K. H. "Why Rebuttals May Not Work: The Psychology of Misinformation." *Media Asia* 44, no. 2 (2017): 79-87.

Ecker, Ullrich K. H. and Luke M. Antonio. "Can You Believe It? An Investigation into the Impact of Retraction Source Credibility on the Continued Influence Effect." *Memory & Cognition* 49 (2021): 631-644.

Ecker, Ullrich K. H., Joshua L. Hogan, and Stephan Lewandowsky. "Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction?" *Journal of Applied Research in Memory and Cognition* 6 (2017): 185-192. doi: 10.1037/h0101809.

Ecker, Ullrich, Stephan Lewandowsky, Kalpana Jayawardana, and Alexander Mladenovic. "Refutations of Equivocal Claims: No Evidence for an Ironic Effect of Counterargument Number." *Journal of Applied Research in Memory and Cognition* 8, no. 1 (2019): 98-107. https://doi.org/10.1016/j.jarmac.2018.07.005.

Ecker, Ullrich K. H., Stephan Lewandowsky, Briony Swire, and Darren Chang. "Correcting False Information in Memory: Manipulating the Strength of Misinformation Encoding and Its Retraction." *Psychonomic Bulletin & Review* 18, no. 3 (2011): 570-578. doi: 10.3758/s13423-011-0065-1. https://doi.org/10.3758/s13423-011-0065-1.

Epstein, Z., A. J. Berinsky, R. Cole, A. Gully, G. Pennycook, and D.G. Rand. "Developing an Accuracy-Prompt Toolkit to Reduce COVID-19 Misinformation Online." *Harvard Kennedy School (HKS) Misinformation Review* (2021).

Fabry, Merrill. "Here's How the First Fact-Checkers Were Able to Do Their Jobs Before the Internet." *Time*. Aug. 24, 2017. Accessed Aug. 24, 2017. https://time.com/4858683/fact-checking-history/.

"Five Key Questions of Media Literacy Education." *Center for Media Literacy*. 2005. https://www.medialit.org/sites/default/files/14B_CCKQPoster+5essays.pdf.

Fleming, Jennifer, and Christopher Karadjov. "Focusing on Facts: Media and News Literacy Education in the Age of Misinformation." *Media Literacy in a Disruptive Media Environment*, 77-93. Routledge, 2020.

Flynn, D.J., Brendan Nyhan, and Jason Reifler. "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics." *Political Psychology* 38, no. S1 (2017): 127-150. doi: https://doi.org/10.1111/pops.12394.

Fridkin, Kim, Patrick J. Kenney, and Amanda Wintersieck. "Liar, Liar, Pants on Fire: How Fact-Checking Influences Citizens' Reactions to Negative Advertising." *Political Communication* 32, no. 1 (2015): 127-151. doi: 10.1080/10584609.2014.914613.

Funke, Daniel. "From Pants on Figre to Pinocchio: All the Ways That Fact-Checkers Rate Claims." Poynter. June 18, 2019. Accessed Nov. 4, 2022. https://www.poynter.org/fact-checking/2019/from-pants-on-fire-to-pinocchio-all-the-ways-that-fact-checkers-rate-claims/.

Funke, Daniel, and Susan Benkleman. "Factually: Games to Teach Media Literacy." July 18, 2019. American Press Institute. https://www.americanpressinstitute.org/fact-checking-project/factually-newsletter/factually-games-to-teach-media-literacy/.

Garrett, R. Kelly, Erik C. Nisbet, and Emily K. Lynch. "Undermining the Corrective Effects of Media-Based Political Fact Checking? The Role of Contextual Cues and Naïve Theory." *Journal of Communication* 63, no. 4 (2013): 617-637. doi: https://doi.org/10.1111/jcom.12038.

Gesser-Edelsburg, Anat, Alon Diamant, Rana Hijazi, and Gustavo S. Mesch. "Correcting Misinformation by Health Organizations During Measles Outbreaks: A Controlled Experiment." *PLOS ONE* 13, no. 12 (2018): e0209505. doi: 10.1371/journal.pone.0209505.

Goodwin, Cara. "The Benefits of In-Person School vs. Remote Learning." *Psychology Today* (Aug. 20, 2021). https://www.psychologytoday.com/us/blog/parenting-translator/202108/the-benefits-in-person-school-vs-remote-learning.

Graves, Lucas et al. "Understanding Innovations in Journalistic Practice: A Field Experiment Examining Motivations for Fact-Checking." *Journal of Communication* 66, no. 1 (2016). https://doi.org/10.1111/jcom.12198.

Guess, Andrew M., Michael Lerner, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. "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* 117, no. 27 (2020): 15536-15545. doi: doi:10.1073/pnas.1920498117.

Hameleers, Michael. "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 & Society* 25, no. 1 (2022): 110-126. doi: 10.1080/1369118X.2020.1764603.

Hart, P. S., and E. C. Nisbet. "Boomerang Effects in Science Communication: How Motivated Reasoning and Identity Cues Amplify Opinion Polarization About Climate Mitigation Policies." *Communication Research* 39, no. 6 (2012): 701–23. doi: 10.1177/0093650211416646.

Hobbs, Renee. *Digital and Media Literacy: A Plan of Action*. Aspen Institute. 2010. https://mediaeducationlab.com/sites/default/files/Hobbs%2520Digital%2520and%2520Media%2520Literacy%2520Plan%2520of%2520Action_0_0.pdf.

Hobbs, Renee. "Grandparents of Media Literacy." Media Educatation Lab. 2017. https://grandparentsofmedialiteracy.com/.

Huang, Yan, and Weirui Wang. "When a Story Contradicts: Correcting Health Misinformation on Social Media Through Different Message Formats and Mechanisms." *Information, Communication & Society* 25, no. 8 (2022): 1192-1209. doi: 10.1080/1369118X.2020.1851390.

Interview with Dr. Briony Swire-Thompson, Dec. 5, 2022.

Interview with Dr. Emily Vraga, Dec. 6, 2022.

Interview with fact-checking subject matter expert, Dec. 1, 2022.

Interview with Dr. Jon Roozenbeek, Nov. 11, 2022.

Interview with Dr. Renee Hobbs, Dec. 9, 2022.

Jackson, Brooks, and Kathleen Hall Jamieson. *UnSpun: Finding Facts in a World of Disinformation*. Toronto, Ontario: Random House Trade Paperbacks, 2007.

Jarman, Jeffrey W. “Influence of Political Affiliation and Criticism on the Effectiveness of Political Fact-Checking.” *Communication Research Reports* 33, no. 1 (2016): 9-15. doi: 10.1080/08824096.2015.1117436.

Jeong, S. H., H. Cho, and Y. Hwang. “Media Literacy Interventions: A Meta-Analytic Review.” *Journal of Communication* PMC3377317 62, no. 3 (2012): 454-472. doi: 10.1111/j.1460-2466.2012.01643.x. NLM.

Jones-Jang, S. Mo, Tara Mortensen, and Jingjing Liu. “Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don’t.” *American Behavioral Scientist* 65, no. 2 (2021): 371-388. doi: 10.1177/0002764219869406.

Kessler, Glenn. “About the Fact Checker.” *Washington Post*. Jan. 1, 2017. https://www.washingtonpost.com/politics/2019/01/07/about-fact-checker/.

“Learn to Discern (L2D): Media Literacy Training.” *IREX*. Accessed Nov. 18, 2021. https://www.irex.org/project/learn-discern-l2d-media-literacy-training#component-id-783.

Lewandowsky, Stephan, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. “Misinformation and Its Correction: Continued Influence and Successful Debiasing.” *Association for Psychological Science* (Sept. 18, 2012). https://www.psychologicalscience.org/publications/journals/pspi/misinformation1.html.

Lewandowsky, Stephan et al. *The Debunking Handbook 2020*. 2020.

Lewandowsky, Stephan, and Sander Van Der Linden. “Countering Misinformation and Fake News Through Inoculation and Prebunking.” *European Review of Social Psychology* 32, no. 2 (2021): 348-384.

Li Qian Tay et al. “A Comparison of Prebunking and Debunking Interventions for Implied versus Explicit Misinformation.” *British Journal of Psychology* 113, no. 3 (2022). doi: 10.1111/bjop.12551. NLM.

Linden, Sander van der. “Misinformation: Susceptibility, Spread, and Interventions to Immunize the Public.” *Nature Medicine* 28, no. 3 (2022): 460-467.

MacFarlane, Doulas, Lian Qian Tay, Mark J. Hurlstone, and Ullrich. K. H. Ecker. “Refuting Spurious COVID-19 Treatment Claims Reduces Demand and Misinformation Sharing.” *Journal of Applied Research in Memory and Cognition* PMC7771267 10, no. 2 (2021): 248-258. doi: 10.1016/j.jarmac.2020.12.005. NLM.

Mackinnon, Amy. “US Army Failed to Warn Troops About COVID-19 Disinformation.” *Foreign Policy*. Oct. 21, 2021. https://foreignpolicy.com/2021/10/21/us-army-covid-19-disinformation-russia-china/.

Maertens, Rakoen, Frederik Anseel, and Sander van der Linden. “Combatting Climate Change Misinformation: Evidence for Longevity of Inoculation and Consensus Messaging Effects.” *Journal of Environmental Psychology* 70 (2020): 101455. doi: https://doi.org/10.1016/j.jenvp.2020.101455.

Maertens, Rakoen, Jon Roozenbeek, Melisa Basol, and Sander van der Linden. “Long-Term Effectiveness of Inoculation Against Misinformation: Three Longitudinal Experiments.” *Journal of Experimental Psychology* 27, no. 1 (2021): 1-16. doi: 10.1037/xap0000315. NLM.

Martel, Cameron, Mohsen Mosleh, and David Gertler Rand. "You're Definitely Wrong, Maybe: Correction Style Has Minimal Effect on Corrections of Misinformation Online." *Media and Communication* 9, no. 1 (Feb. 2021). https://dspace.mit.edu/handle/1721.1/129719.

McDougall, Julian, Lee Edwards, and Karen Fowler-Watt. "Media Literacy in the Time of COVID." *Sociologia Della Comunicazione* (2021).

McGuire, William J. "Inducing Resistance to Persuasion. Some Contemporary Approaches." in *Self and Society. An Anthology of Readings*, edited by C. C. Haaland, and W. O. Kaelber, 192-230. Lexington, MA: Ginn Custom Publishing, 1964.

"Media Literacy Defined." National Association of Media Literacy Education. https://namle.net/resources/media-literacy-defined/.

Micallef, Nicholas, Mihai Avram, Filippo Menczer, and Sameer Patil. "Fakey: A Game Intervention to Improve News Literacy on Social Media." *Proceedings of the ACM on Human-Computer Interaction* 5 (2021): 1-27. doi: 10.1145/3449080.

Mihailidis, Paul. "Beyond Cynicism: How Media Literacy Can Make Students More Engaged Citizens." PhD diss., University of Maryland, College Park, 2008. https://drum.lib.umd.edu/handle/1903/8301.

Mihailidis, Paul, and Samantha Viotty. "Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in 'Post-Fact' Society." *American Behavioral Scientist* 61, no. 4 (2017): 441-454. doi: 10.1177/0002764217701217. https://journals.sagepub.com/doi/abs/10.1177/0002764217701217.

Moore, Ryan C., and Jeffrey T. Hancock. "A Digital Media Literacy Intervention for Older Adults Improves Resilience to Fake News." *Scientific Reports* 12, no. 1 (2022): 6008. doi: 10.1038/s41598-022-08437-0.

Murrock, Erin, Joy Amulya, Mehri Druckman, and Tetiana Libyva. "Winning the War on State-Sponsored Propaganda." *IREX* (2018).

Niederdeppe, Jeff, Sarah E. Gollust, and Colleen L. Barry. "Inoculation in Competitive Framing: Examining Message Effects on Policy Preferences." *Public Opinion Quarterly* 78, no. 3 (2014): 634-655. Accessed Feb. 15, 2023. doi: 10.1093/poq/nfu026. https://doi.org/10.1093/poq/nfu026.

Nieminen, Sakari, and Lauri Rapheli. "Fighting Misperceptions and Doubting Journalists' Objectivity: A Review of Fact-Checking Literature." *Perspectives on Psychological Science* 17, no. 3 (2019).

Nyhan, Brendan, Ethan Porter, Jason Reifler, and Thomas J. Wood. "Taking Fact-Checks Literally but Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability." *Political Behavior* 42 (2020): 939-960. doi: 10.1007/s11109-019-09528-x.

Nyhan, Brendan, and Jason Reifler. *Misinformation and Fact-Checking: Research Findings from Social Science*. New America Foundation. Feb. 2012.

Nyhan, Brendan, and Jason Reifler. "The Roles of Information Deficits and Identity Threat in the Prevalence of Misperceptions." *Journal of Elections, Public Opinion and Parties* 29, no. 2 (2019): 222-244.

Nyhan, Brendan, and Jason Reifler. “When Corrections Fail: The Persistence of Political Misperceptions.” *Political Behavior* 32, no. 2 (2010): 303-330.

Nyhan, Brenda., J. Reifler, S. Richey, and G. L. Freed. “Effective Messages in Vaccine Promotion: A Randomized Trial.” *Pediatrics* 133, no. 4 (2014): e835–42. doi: 10.1542/peds.2013-2365.

Nyhan, B., J. Reifler, and P. A. Ubel. “The Hazards of Correcting Myths About Health Care Reform.” *Medical Care* 51, no. 2 (2013): 127-132. doi: 10.1097/MLR.0b013e318279486b. NLM.

Otero, V. “Media Bias Chart: Version 4.0 - Ad Fontes Media.” Ad Fontes Media. 2019. Accessed Nov. 2, 2022. https://www.adfontesmedia.com.

Pamment, James, and Anneli Lindvall Kimber. *Fact-Checking and Debunking: A Best Practice Guide to Dealing with Disinformation.* NATO Strategic Communications Centre of Excellence. 2021. https://lup.lub.lu.se/search/publication/d5a3ed77-e218-431b-ac9b-c38a6d5a98a1.

Parker, Kimberly A., Stephen A. Rains, and Bobi Ivanov. “Examining the ‘Blanket of Protection’ Conferred by Inoculation: The Effects of Inoculation Messages on the Cross-Protection of Related Attitudes.” *Communication Monographs* 83, no. 1 (2016): 49-68. doi: 10.1080/03637751.2015.1030681.

Peng, Wei, Sue Lim, and Jingbo Meng. “Persuasive Strategies in Online Health Misinformation: A Systematic Review.” *Information, Communication & Society* (2022): 1-18. doi: 10.1080/1369118X.2022.2085615.

Pennycook, Gordon, Adam Bear, Evan T. Collins, and David G. Rand. “The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines without Warnings.” *Management Science* 66, no. 11 (2020): 4944-4957. doi: 10.1287/mnsc.2019.3478.

Pfau, Michael. “Designing Messages for Behavioral Inoculation.” In *Designing Health Messages: Approaches from Communication Theory and Public Health Practice*. Thousand Oaks, CA: Sage Publications, 1995, 99-113. doi: 10.4135/9781452233451.n6.

Pfau, Michael, and Allan Louden. “Effectiveness of Adwatch Formats in Deflecting Political Attack Ads.” *Communication Research* 21, no. 3 (1994): 325-341. doi: 10.1177/009365094021003005.

Porter, Ethan and Thomas J. Wood. “The Global Effectiveness of Fact-Checking: Evidence from Simultaneous Experiments in Argentina, Nigeria, South Africa, and the UK.” *Proceedings of the National Academy of Sciences* 188, no. 37 (2021). https://doi.org/10.1073/pnas.2104235118.

Porter, Ethan, and Thomas J. Wood. “Political Misinformation and Factual Corrections on the Facebook News Feed: Experimental Evidence.” *Journal of Politics* 84, no. 3 (2022): 1812-1817. doi: 10.1086/719271.

Porter, Ethan, Thomas J. Wood, and Babak Bahador. ”Can Presidential Misinformation on Climate Change Be Corrected? Evidence from Internet and Phone Experiments.” *Research and Politics* 6, no. 3. (2019). https://doi.org/10.1177/2053168019864784.

Potter, J., and J. McDougall. *Digital Media, Culture and Education*. London: Palgrave Macmillan/Springer, 2017.

Ritchie, Elspeth Cameron. “Psychiatry in the Korean War: Perils, PIES, and Prisoners of War.” *Military Medicine* 167, no. 11 (2002): 898-903.

Roozenbeek, Jon, and Sander Van der Linden. "Fake News Game Confers Psychological Resistance Against Online Misinformation." *Palgrave Communications* 5, no. 1 (2019): 1-10.

Roozenbeek, Jon, and Sander van der Linden. "How to Combat Health Misinformation: A Psychological Approach." *American Journal of Health Promotion* 36, no. 3 (2022): 569-575.

Roozenbeek, Jon, Rakoen Maertens, William McClanahan, and Sander van der Linden. "Disentangling Item and Testing Effects in Inoculation Research on Online Misinformation: Solomon Revisited." *Educational and Psychological Measurement* 81, no. 2 (2021): 340-362. doi: 10.1177/0013164420940378.

Roozenbeek, Jon, Cecilie S. Traberg, and Sander van der Linden. "Technique-Based Inoculation Against Real-World Misinformation." *Royal Society Open Science* 9, no. 5 (2022): 211719. doi: doi:10.1098/rsos.211719. https://royalsocietypublishing.org/doi/abs/10.1098/rsos.211719.

Sangalang, Angela, Yotam Ophir, and Jospeh. N. Cappella. "The Potential for Narrative Correctives to Combat Misinformation(†)." *Journal of Communication* PMC6544903 69, no. 3 (2019): 298-319. doi: 10.1093/joc/jqz014. NLM.

Shin, Tae S., John Ranellucci, and Cary J. Roseth. "Effects of Peer and Instructor Rationales on Online Students' Motivation and Achievement." *International Journal of Educational Research* 82 (2017): 184-199. doi: https://doi.org/10.1016/j.ijer.2017.02.001.

Sirlin, N., Z. Epstein, A. A. Arechar, and D. G. Rand. "Digital Literacy Is Associated with More Discerning Accuracy Judgments but Not Sharing Intentions." *Harvard Kennedy School (HKS) Misinformation Review* (2021).

Sullivan, M. Connor. "Leveraging Library Trust to Combat Misinformation on Social Media." *Library & Information Science Research* 41, no. 1 (2019): 2-10.

Swift, Art. "Americans' Trust in Mass Media Sinks to New Low." Gallup. Sept. 14, 2016. Accessed Nov. 4, 2022. https://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx.

Swire, Briony, Adam J. Berinsky, Stephan Lewandowsky, and Ullrich K. H. Ecker. "Processing Political Misinformation: Comprehending the Trump Phenomenon." *Royal Society Open Science* 4, no. 3 (2017): 160802. doi: 10.1098/rsos.160802.

Swire-Thompson, Briony, John Cook, Lucy H. Butler, Jasmyne A. Sanderson, Stephan Lewandowsky, and Ullrich K. H. Ecker. "Correction Format Has a Limited Role When Debunking Misinformation." *Cognitive Research: Principles and Implications* PMC8715407 6, no. 1 (2021): 83. doi: 10.1186/s41235-021-00346-6. NLM.

Swire-Thompson, Briony, Joseph DeGutis, and David Lazer. "Searching for the Backfire Effect: Measurement and Design Considerations." *Journal of Applied Research in Memory and Cognition* 9, no. 3 (2020): 286-299.

Swire-Thompson, Briony, Mitch Dobbs, Ayanna Thomas, and Joseph DeGutis. "Memory Failure Predicts Belief Regression After the Correction of Misinformation." *Cognition* 230 (2023): 105276. doi: 10.1016/j.cognition.2022.105276. NLM.

Swire, B., U. K. H. Ecker, and S. Lewandowsky. "The Role of Familiarity in Correcting Inaccurate Information." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 43, no. 12 (2017): 1948-1961. doi: 10.1037/xlm0000422. NLM.

Swire-Thompson, Briony, Ullrich K. H. Ecker, Stephan Lewandowsky, and Adam J. Berinsky. "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation." *Political Psychology* 41, no. 1 (2020): 21-34. doi: https://doi.org/10.1111/pops.12586.

Swire-Thompson, Briony, Nicholas Miklaucic, John P. Wihbey, David Lazer, and Joseph DeGutis. "The Backfire Effect After Correcting Misinformation Is Strongly Associated with Reliability." *Journal of Experimental Psychology: General* 151 (2022): 1655-1665. doi: 10.1037/xge0001131.

Tay, Li Qian, Mark J. Hurlstone, Tim Kurz, and Ullrich K. H. Ecker. "A Comparison of Prebunking and Debunking Interventions for Implied Versus Explicit Misinformation." *British Journal of Psychology* 113, no. 3 (2022): 591-607. doi: 10.1111/bjop.12551. NLM.

"The True Story of Brainwashing and How It Shaped America." *Smithsonian Magazine*. May 22. 2017. https://www.smithsonianmag.com/history/true-story-brainwashing-and-how-it-shaped-america-180963400/.

Tully, Melissa, and Emily K. Vraga. "A Mixed Methods Approach to Examining the Relationship Between News Media Literacy and Political Efficacy." *International Journal of Communication* 12 (2018): 22.

Tully, Melissa, Emily K. Vraga, and Leticia Bode. "Designing and Testing News Literacy Messages for Social Media." *Mass Communication and Society* 23, no. 1 (2020): 22-46. doi: 10.1080/15205436.2019.1604970.

US Senate. The Select Committee on Intelligence and the Subcommittee on Health and Scientific Research of the Committee on Human Resources. *Joint Hearing on Project MKULTRA, the CIA'S Program of Research in Behavioral Modification*. 95 Cong., 1st sess., Aug. 3, 1977. https://www.intelligence.senate.gov/sites/default/files/hearings/95mkultra.pdf.

van der Linden, Sander, Anthony Leiserowitz, Seth Rosenthal, and Edward Maibach. "Inoculating the Public Against Misinformation About Climate Change." *Global Challenges* 1, no. 2 (2017): 1600008. doi: https://doi.org/10.1002/gch2.201600008. https://onlinelibrary.wiley.com/doi/abs/10.1002/gch2.201600008.

van der Linden, Sander, and Jon Roozenbeek. "Psychological Inoculation Against Fake News." *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation.* New York, NY: Routledge/Taylor & Francis Group, 2021, 147-169. doi: 10.4324/9780429295379-11.

van der Meer, Toni G. L. A., and Michael Hameleers. "Fighting Biased News Diets: Using News Media Literacy Interventions to Stimulate Online Cross-Cutting Media Exposure Patterns." *New Media & Society* 23, no. 11 (2021): 3156-3178. doi: 10.1177/1461444820946455.

van der Meer, Toni G. L. A., and Yan Jin. "Seeking Formula for Misinformation Treatment in Public Health Crises: The Effects of Corrective Information Type and Source." *Health Communication* 35, no. 5 (2020): 560-575.

Velez, Yamil R., Ethan Porter, and Thomas J. Wood. "Latino-Targeted Misinformation and the Power of Factual Corrections." *Journal of Politics* (published online Feb. 14, 2023).

Vraga, Emily K., and Leticia Bode. "Using Expert Sources to Correct Health Misinformation in Social Media." *Science Communication* 39, no. 5 (2017): 621-645.

Vraga, Emily K., Leticia Bode, and Melissa Tully. "Creating News Literacy Messages to Enhance Expert Corrections of Misinformation on Twitter." *Communication Research* 49, no. 2 (2020): 245-267. doi: 10.1177/0093650219898094.

Vraga, E. K., L. Bode, and M. Tully. "The Effects of a News Literacy Video and Real-Time Corrections to Video Misinformation Related to Sunscreen and Skin Cancer." *Health Communication* 37, no. 13 (2020): 1622-1630. doi: 10.1080/10410236.2021.1910165. NLM.

Vraga, Emily K., Sojung Claire Kim, and John Cook. "Testing Logic-Based and Humor-Based Corrections for Science, Health, and Political Misinformation on Social Media." *Journal of Broadcasting & Electronic Media* 63, no. 3 (2019), 393-414. doi: 10.1080/08838151.2019.1653102.

Vraga, Emily K., Sojung Claire Kim, John Cook, and Leticia Bode. "Testing the Effectiveness of Correction Placement and Type on Instagram." *International Journal of Press/Politics* 25, no. 4 (2020): 632-652.

Vraga, Emily K., and Melissa Tully. "Effective Messaging to Communicate News Media Literacy Concepts to Diverse Publics." *Communication and the Public* 1, no. 3 (2016): 305-322. doi: 10.1177/2057047316670409.

Vraga, Emily K., Melissa Tully, Heather Akin, and Hernando Rojas. "Modifying Perceptions of Hostility and Credibility of News Coverage of an Environmental Controversy Through Media Literacy." *Journalism* 13, no. 7 (2012): 942-959. doi: 10.1177/1464884912455906.

Vraga, Emily K., and Melissa Tully. "Effectiveness of a Non-Classroom News Media Literacy Intervention Among Different Undergraduate Populations." *Journalism & Mass Communication Educator* 71, no. 4 (2016): 440-452. doi: 10.1177/1077695815623399.

Vraga, Emily K., and Melissa Tully. "Media Literacy Messages and Hostile Media Perceptions: Processing of Nonpartisan Versus Partisan Political Information." *Mass Communication and Society* 18, no. 4 (2015): 422-448. doi: 10.1080/15205436.2014.1001910.

Vraga, Emily K., Melissa Tully, and Hernando Rojas. "Media Literacy Training Reduces Perception of Bias." *Newspaper Research Journal* 30, no. 4 (2009): 68-81. doi: 10.1177/073953290903000406.

Walter, N., J. J. Brooks, C. J. Saucier, and S. Suresh. "Evaluating the Impact of Attempts to Correct Health Misinformation on Social Media: A Meta-Analysis." *Health Communication* 36, no. 13 (2021): 1776-1784. doi: 10.1080/10410236.2020.1794553. NLM.

Walter, Nathan, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. "Fact-Checking: A Meta-Analysis of What Works and for Whom." *Political Communication* 37, no. 3 (2020): 350-375. doi: 10.1080/10584609.2019.1668894.

Walter, Nathan, and Riva Tukachinsky. "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, no. 2 (2020): 155-177. doi: 10.1177/0093650219854600.

Williams, Matt N., and Christina M. C. Bond. "A Preregistered Replication of 'Inoculating the Public Against Misinformation About Climate Change.'" *Journal of Environmental Psychology* 70 (2020). doi: 101456.

Wolters, Heather, Kasey Stricklin, Neil Carey, and Megan K. McBride. *The Psychology of (Dis) Information: A Primer on Key Psychological Mechanisms*. CNA. 2021.

Wood, Thomas, and Ethan Porter. “The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence.” *Political Behavior* 41 (2019). doi: 10.1007/s11109-018-9443-y.

Young, Dannagal G., Kathleen Hall Jamieson, Shannon Poulsen, and Abigail Goldring. “Fact-Checking Effectiveness as a Function of Format and Tone: Evaluating FactCheck.org and FlackCheck.org.” *Journalism & Mass Communication Quarterly* 95, no. 1 (2018): 49-75. doi: 10.1177/1077699017710453.

Zerback, Thomas, Florian Töpfl, and Maria Knöpfle. “The Disconcerting Potential of Online Disinformation: Persuasive Effects of Astroturfing Comments and Three Strategies for Inoculation Against Them.” *New Media & Society* 23, no. 5 (2021): 1080-1098. doi: 10.1177/1461444820908530.

This page intentionally left blank.

This report was written by CNA's Strategy, Policy, Plans, and Programs Division (SP3).

SP3 provides strategic and political-military analysis informed by regional expertise to support operational and policy-level decision-makers across the Department of the Navy, the Office of the Secretary of Defense, the unified combatant commands, the intelligence community, and domestic agencies. The division leverages social science research methods, field research, regional expertise, primary language skills, Track 1.5 partnerships, and policy and operational experience to support senior decision-makers.

This report may contain hyperlinks to websites and servers maintained by third parties. CNA does not control, evaluate, endorse, or guarantee content found in those sites. We do not assume any responsibility or liability for the actions, products, services, and content of those sites or the parties that operate them.

Dedicated to the Safety and Security of the Nation

DIM-2023-U-035081-Final

3003 Washington Boulevard, Arlington, VA 22201
www.cna.org 703-824-2000