Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 9, 2299–2313
https://doi.org/10.1037/xge0001627

Discrediting Health Disinformation Sources: Advantages of Highlighting Low Expertise

Briony Swire-Thompson[1, 2, 3], Kristen Kilgallen[1, 3], Mitch Dobbs[1, 3], Jacob Bodenger[1], John Wihbey[4], and Skyler Johnson[5]

[1] Network Science Institute, Northeastern University
[2] Department of Political Science, Northeastern University
[3] Department of Psychology, Northeastern University
[4] Department of Journalism, Northeastern University
[5] Department of Radiation Oncology, Huntsman Cancer Hospital, University of Utah

Disinformation is false information spread intentionally, and it is particularly harmful for public health. We conducted three preregistered experiments ($N = 1{,}568$) investigating how to discredit dubious health sources and disinformation attributed to them. Experiments 1 and 2 used cancer information and recruited representative U.S. samples. Participants read a vignette about a seemingly reputable source and rated their credibility. Participants were randomly assigned to a control condition or interventions that (a) corrected the source's disinformation, (b) highlighted the source's low expertise, or (c) corrected disinformation *and* highlighted low expertise (Experiment 2). Next, participants rated their belief in the source's disinformation claims and rerated their credibility. We found that highlighting low expertise was equivalent to (or more effective than) other interventions for reducing belief in disinformation. Highlighting low expertise was also more effective than correcting disinformation for reducing source credibility, although combining it with correcting disinformation outperformed low expertise alone (Experiment 2). Experiment 3 extended this paradigm to vaccine information in vaccinated and unvaccinated subgroups. A conflict-of-interest intervention and 1 week retention interval were also added. Highlighting low expertise was the most effective intervention in both vaccinated and unvaccinated participants for reducing belief in disinformation and source credibility. It was also the only condition where belief change was sustained over 1 week, but only in the vaccinated subgroup. In sum, highlighting a source's lack of expertise is a promising option for fact-checkers and health practitioners to reduce belief in disinformation and perceived credibility.

Public Significance Statement

Inaccurate information can be extremely harmful to people's health. This is particularly true for cancer care, where the use of unproven therapies is associated with decreased survival, as well as vaccine uptake for diseases such as COVID-19. The present study suggests that highlighting a source's low expertise is the best strategy for reducing people's belief in disinformation and for reducing perceived credibility of dubious sources. These findings can inform health practitioners, science communicators, fact-checkers, and social media platforms.

Keywords: misinformation, disinformation, source credibility, cancer, vaccines

Supplemental materials: https://doi.org/10.1037/xge0001627.supp

Sharda Umanath served as action editor.
Briony Swire-Thompson https://orcid.org/0000-0001-7464-8940
Data are available at the Open Science Framework. All participants in this study were anonymously collected through Prolific. Preregistration data of Experiment 1 (https://osf.io/kd9sh/), Experiment 2 (https://osf.io/tq7ma/), and Experiment 3 (https://osf.io/7bjre) are publicly available. Prior dissemination of the ideas and data appearing in the article includes presentations by Briony Swire-Thompson at the Society of Research in Memory and Cognition conference in 2023, a National Academy of Sciences presentation in 2023, and National Academy of Sciences Kavli Frontiers of Science Symposium.

The authors declare no conflicts of interests. This work was primarily funded by the National Institutes of Health National Cancer Institute awarded to Briony Swire-Thompson "Cancer misinformation on social media and its correction" (Grant R00CA248720). Experiment 3 participant recruitment was funded by Northeastern University's COVID-19 crisis seed funding to Briony Swire-Thompson and John Wihbey.

Briony Swire-Thompson played a lead role in conceptualization, data curation, formal analysis, funding acquisition, methodology, supervision, and writing–original draft and an equal role in writing–review and editing. Kristen Kilgallen played a supporting role in visualization, formal analysis, and data curation and an equal role in validation and writing–review and editing. Mitch Dobbs played a lead role in visualization, a supporting role in data curation and formal analysis, and an equal role in validation and writing-review and editing.

continued

Inaccurate information can be extremely harmful to people's health. This is particularly true for topics such as cancer, where the use of unproven therapies is associated with decreased survival (Johnson et al., 2018b), and vaccine uptake for diseases such as COVID-19 (Loomba et al., 2021). In the current article, we use the term *misinformation* as an umbrella term for any information contrary to clear scientific evidence or consensus (Vraga & Bode, 2020) and *disinformation* for false information that has been spread intentionally in order to gain money, power, or reputation. Although "snake oil salesmen" who spread health disinformation are not new, the internet has enabled dubious sources to appear legitimate at scale. It is relatively easy for a person to claim to have a PhD from a prestigious university, have an impressive h-index on Google Scholar due to predatory journal publications, or profit from selling books about cancer or vaccines on Amazon (Swire-Thompson & Lazer, 2022). There is a substantial lack of knowledge available to health practitioners for how to respond to patients' inaccurate beliefs or challenge disreputable sources that patients have found online (Chou et al., 2018). The question remains: How do legitimate health communicators best discredit these sources and steer people back toward evidence-based medicine?

There are several interventions available to practitioners to discredit health sources, such as correcting the sources' prior inaccuracies or highlighting their lack of expertise. Research has shown that correcting misinformation is extremely effective at reducing belief in those exact misconceptions (Ecker et al., 2022; Vraga et al., 2022), although it is also important that people disregard *future* statements from a disreputable source. There is a growing body of literature suggesting that there is a bidirectional relationship between source credibility and the content of the claim (Collins et al., 2018; Wertgen & Richter, 2020), such that correcting a source's disinformation should decrease their perceived credibility (Westbrook et al., 2023) and highlighting that a source is not credible should influence how truthful future claims are thought to be (Hughes et al., 2014; Nadarevic et al., 2020). However, calling attention to a source's track record of incorrect statements does not always seem to reduce source credibility evaluations. For example, several studies have shown that correcting a politician's inaccuracies does not reduce participants' positive feelings toward them, their candidate favorability, or even voting intentions (Aird et al., 2018; Nyhan et al., 2020; Swire-Thompson et al., 2020).

Highlighting that the source has a conflict of interest could also be effective at discrediting disinformation sources (Eagly et al., 1978; Kelman & Hovland, 1953). Source credibility is often thought to depend on both perceived expertise (the extent to which a source is *capable* or qualified to provide accurate information) and perceived trustworthiness (the extent to which a source is *willing* to provide accurate information; Pornpitakpan, 2004). A common method for reducing trustworthiness is showing that the source has a conflict of interest. Barnes et al. (2018) found that highlighting a scientist's financial conflict of interest was equally as effective as showing their work was critically flawed at reducing how favorably participants viewed the scientists' claims. Indeed, when correcting misinformation, several studies have found that perceived trustworthiness is far more important than expertise for the correction's efficacy (Ecker & Antonio, 2021; Guillory & Geraci, 2013; though, see Susmann & Wegener, 2023).

Finally, it is unclear how complementary and alternative medicine (CAM) endorsement might influence the efficacy of discrediting interventions. While complementary therapies are used *in concert with* evidence-based treatment, alternative medicines are used *instead of* evidence-based treatment (Peterson et al., 2020). Some complementary medicines can improve symptoms and quality of life (such as yoga), although others can have adverse reactions with treatment (such as supplements affecting chemotherapy; Ambrosone et al., 2020). On the one hand, people who endorse CAM could be equally as receptive to discreditation interventions as people who do not. Trust in CAM often arises from a desire to take an active role in addressing one's health rather than an ignorance or rejection of science (Johnson et al., 2018a; Lobera & Rogero-García, 2021). On the other hand, it is also possible that people who endorse CAM are less receptive to discreditation interventions than people who do not. Scherer et al. (2021) found that people who had positive attitudes toward CAM were more susceptible to believing misinformation about health, including cancer and the HPV vaccine. Furthermore, CAM users may be less sensitive to source discreditation prompts if they identify with the disinformation source (Briñol & Petty, 2009; Lindeman, 2011), given that people are more likely to trust sources that share their worldviews. For example, Green et al. (2023) found that participants with higher conspiracy beliefs rated health professionals who shared conspiracies as "braver," while participants with weaker conspiracy beliefs had more negative impressions of these health professionals.

The goal of the present study was to examine strategies for discrediting dubious health sources and reducing belief in their disinformation. Experiments 1 and 2 used cancer information (see Swire-Thompson & Johnson, 2024, for a review of why cancer is a topic well-suited to studying mis- and disinformation). Experiment 3 extended this paradigm to vaccine information for generalizability and because the public had been saturated with corrective information about this topic throughout the COVID-19 pandemic. The topic of vaccines also allowed us to focus on an unvaccinated subgroup, a behavioral indicator that participants may be more likely to endorse vaccine disinformation. We specifically created the fictitious sources used in our experiments ("Dr. Smith" and "Dr. Davis") to have nefarious intentions and, therefore, frame this article as a study of *disinformation*. However, the current findings certainly generalize to misinformation more broadly, given that the participants were unaware of the fictitious sources' intent. The discreditation interventions revealed to participants that the source had spread inaccurate information previously, had low expertise, had both spread inaccurate information *and* low expertise, or had a conflict of interest. Of additional interest was whether disinformation belief and evaluations of credibility were associated with being "close to cancer" (i.e., those who report that cancer had impacted their lives: Experiments 1 and 2), vaccine hesitancy (Experiment 3), or belief in CAM.

Jacob Bodenger played a supporting role in writing–review and editing. John Wihbey played a supporting role in funding acquisition and writing–review and editing. Skyler Johnson played a supporting role in supervision and writing–review and editing.

Correspondence concerning this article should be addressed to Briony Swire-Thompson, Network Science Institute, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, United States. Email: b.swire-thompson@northeastern.edu

Experiment 1

Method

This study was conducted in February 2022. We hypothesized that (a) all interventions would reduce perceived source credibility and feelings more than the control condition and (b) correcting disinformation would have a larger impact on reducing disinformation belief than highlighting a source's low expertise. This hypothesis was based on prior research that found trustworthiness—in our case, a track record of accuracy—to be more important than expertise for correcting misinformation (Ecker & Antonio, 2021; Guillory & Geraci, 2013). In this first experiment, the association between CAM beliefs and disinformation belief, source credibility, and feelings was exploratory. Please see Supplemental Material A, Table S1 for whether preregistered hypotheses in all experiments were supported or rejected.

Design

Experiment 1 was a parallel-arm prospective longitudinal intervention design. The intervention arms included correcting the source's disinformation, highlighting the source's low expertise, and a control condition.

Participants

Participants were a representative sample of 301 U.S. citizens from Prolific over the age of 18. Our preregistered exclusion criteria omitted participants who showed uniform responding across all items ($SD < 0.5$; $N = 0$), participants who completed the task too quickly (i.e., outliers with regard to completion time; $N = 0$), and participants who indicated that their data should be excluded due to lack of effort ($N = 0$). Our final sample included 301 people, with an age range between 19 and 81 years ($M = 45.07$, $SD = 16.23$). Participants in all experiments were asked to self-report their gender (male, female, nonbinary, prefer to self-describe, prefer not to say). Participants also self-reported their race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, prefer to self-describe) and ethnicity (Hispanic or Latino, not Hispanic or Latino). See Supplemental Material B for tables that include counts for gender, race, and ethnicity. The final sample in Experiment 1 included 146 males, 151 females, and four individuals who were nonbinary, preferred to self-describe, or who chose not to disclose their gender. Participants were paid $1.40 for a 7-min task.

Procedure and Stimuli

See Figure 1 for a Consolidated Standards of Reporting Trials flow diagram for all experiments. Participants first read a Northeastern University approved consent form (No. 20-12-16). Next, participants were presented with a short paragraph regarding a fictitious self-proclaimed cancer expert, Dr. Sarah Smith. Modeled off real-world cancer disinformation sources, Dr. Smith presented herself as a researcher in integrative oncology and holding a PhD from University of California, Berkeley (see Supplemental Material C for full text). Participants were asked to rate Dr. Smith's credibility and their feelings toward her. Source credibility was measured using an 11-item composite, all on a 1–7 scale (1 = *not at all*, 7 = *extremely*). This included one item regarding general credibility, ("How credible do you find Dr. Smith?") and a 10-item Credibility Semantic scale (Ohanian, 1990). This Semantic scale had five pairs of adjectives for both expertise (expert, experienced, knowledgeable, qualified, and skilled) and trustworthiness (dependable, honest, reliable, sincere, and trustworthy; Cronbach's $\alpha = .96$; Hess et al., 2022). Although this scale was created to be applied to celebrities, it has often been applied to health (e.g., Lee & Sundar, 2013). Feelings were measured using a feelings thermometer (as presented in the American National Election Studies, n.d.), where participants rated Dr. Smith on a 0°–100° scale. Ratings between 0° and 50° indicated that participants felt unfavorable and "cold" toward Dr. Smith, whereas ratings between 50° and 100° indicated that they felt favorable or "warm" toward her. Ratings of exactly 50° indicated that participants felt neither warm nor cold toward Dr. Smith.

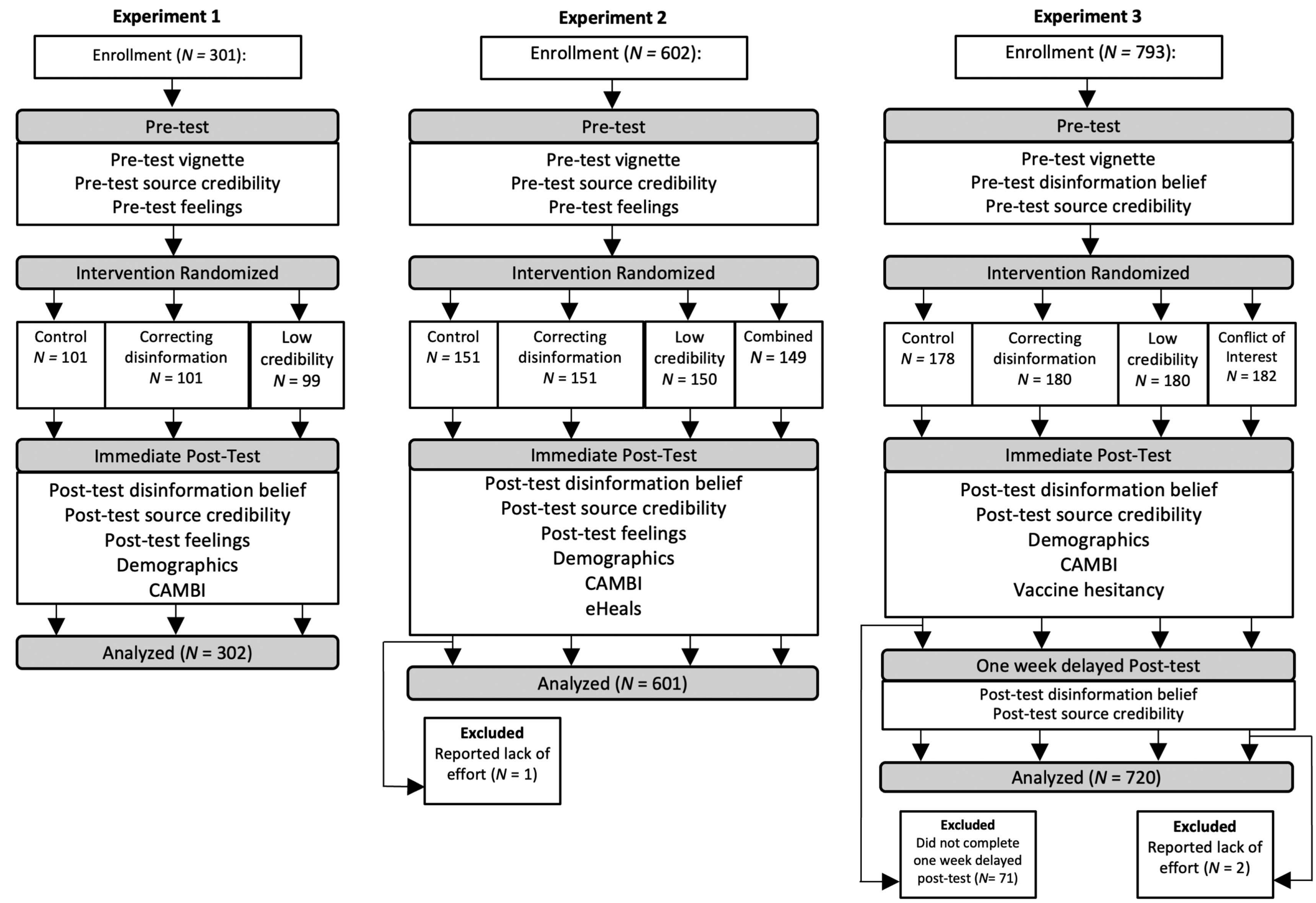
Next, participants were randomly assigned to the correcting disinformation condition, the low-expertise condition, or the control condition. In the correcting disinformation condition, participants saw four disinformation statements corrected, for example, "Dr. Smith has said that fluoride in water can increase your likelihood of developing cancer. *This is false*." Each correction provided an explanation as to why the claim was false and cited a reputable source (e.g., National Institutes of Health). A pilot study was conducted to select these items (see Supplemental Material D for details). The low-expertise condition described how Dr. Smith was not qualified to give medical advice or advice regarding cancer treatment and stated that Dr. Smith's PhD was in anthropology rather than medicine or oncology. In the control condition, participants were presented with details about Dr. Smith's hobbies such as photography and running.

In the posttest, all participants rated four new cancer claims made by Dr. Smith, all of which were disinformation. Note that these items were different from those in the correcting disinformation condition. Participants subsequently rerated source credibility and their feelings toward Dr. Smith. They next completed demographic questions and the Complementary and Alternative Medicine Beliefs Inventory (CAMBI; Bishop et al., 2005). The CAMBI includes 17 items with three subscales: Participation in Treatment, Natural Treatment, and Holistic Health. Finally, participants were asked, "In your opinion, how much has cancer has impacted your life?" (i.e., their "closeness to cancer"), which they answered on a 1–7 scale (1 = *not at all* to 7 = *very much so*). At the end of the study, all participants were given a printable debriefing sheet with information about why the statements were false with supporting evidence linking to reputable sources for further information.

Sample Size Justification

Effect sizes when comparing belief in corrected misinformation to a control are often large (e.g., $\eta_p^2 = .53$, Swire-Thompson et al., 2020), although it is unlikely that they remain as large when extended to new statements by the source. When discrediting sources compared to control, effect sizes are often small to medium (Hughes et al., 2014). Considering the smallest effect size found of $f = .20$ (Flanagin et al., 2020), G*Power3 recommended a sample size of 246 participants (with $\alpha = .05$, power = .80) in order to conduct a one-way analysis of variance (ANOVA) with three levels (Faul et al., 2007). We boosted the sample size to 300 given that our exact effect size remained unknown. This sample should also be sufficient to calculate pre/post

Figure 1
CONSORT Flow Diagram of Experiments 1, 2, and 3



Note. CAMBI = Complementary and Alternative Medicine Beliefs Inventory; CONSORT = Consolidated Standards of Reporting Trials.

differences on our source credibility and feelings measures using an analysis of covariance (ANCOVA; recommended $N = 244$).

Transparency and Openness

See Swire-Thompson et al. (2021) for preregistration (https://osf.io/kd9sh/), data, and analysis scripts.

Results

We first sought to determine if our intervention and control groups were well-balanced across demographic groups. There were no significant differences between groups for age (p = .765), education (p = .973), gender (p = .974), or partisanship (p = .668). See Supplemental Material B for demographic tables.

Belief in Disinformation

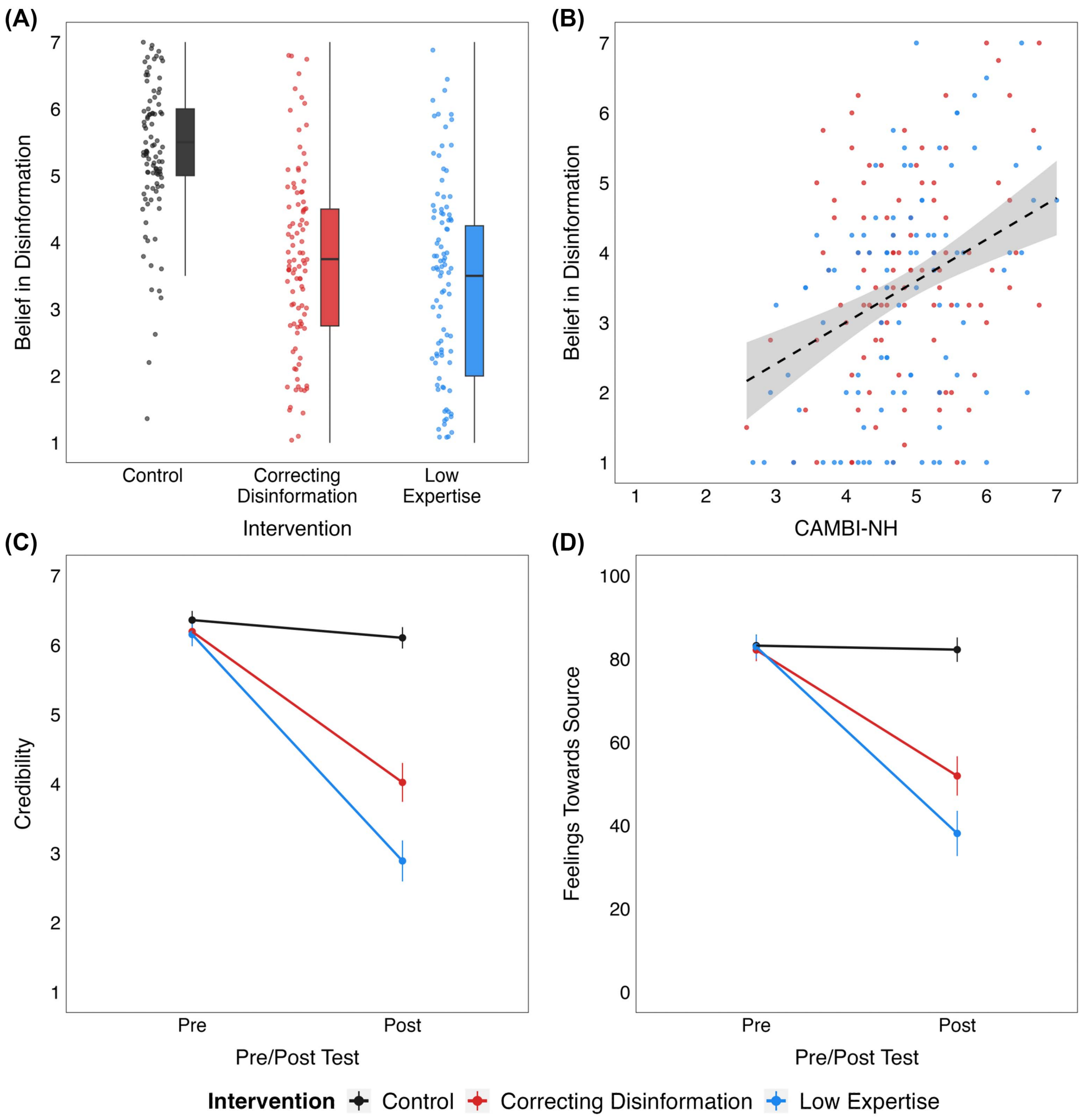
To examine participants' posttest belief in Dr. Smith's disinformation statements, a one-way ANOVA with the interventions (control vs. correction vs. low expertise) was conducted. Note that we used Welch's ANOVA, given that the homogeneity of variance assumption was violated. We found a significant difference between interventions, $F(2, 196.09) = 62.98$, $p < .001$, $MSE = 3.15$, $\eta_p^2 = .27$, with correcting disinformation and low expertise both significantly lower than control, $t(195.36) = 8.88$, $p < .001$, Cohen's $d = 1.18$, and $t(185.39) = 9.86$, $p < .001$, Cohen's $d = 1.39$, respectively. However, as can be seen from Figure 2 (Panel A), there was no difference between the correcting disinformation condition and the low-expertise condition (p = .338), with the interventions equally reducing participants' belief in disinformation. Please note that all post hoc t tests in this article are two-tailed and corrected for multiple comparisons (Tukey's honestly significant difference test or Games–Howell post hoc tests after Welch's ANOVAs), with adjusted p values reported.

Source Credibility and Feelings

Next, we investigated how the interventions impacted perceptions of credibility and feelings toward the source. We ran a one-way ANCOVA with between-subjects factors intervention (control vs. correction vs. low expertise) on the source credibility composite, entering the pretest credibility rating as a covariate. We found a main effect of intervention on posttest credibility, $F(2, 297) = 161.37$; $p < .001$; $MSE = 1.54$; $\eta_p^2 = .52$, indicating that credibility ratings differed between conditions (see Figure 2, Panel C). Both correcting

Figure 2

Experiment 1 (A) Disinformation Belief, (B) the Association Between CAMBI-NH and Postintervention Disinformation Belief (ρ = .29), (C) Source Credibility, and (D) Feelings Toward the Source



Note. Error bars are 95% confidence intervals. Black = control; red = correcting disinformation; blue = low expertise; dashed black = all interventions excluding control; CAMBI-NH = Complementary and Alternative Medicine Beliefs Inventory with only Natural Treatment and Holistic Health subscale. See the online article for the color version of this figure.

disinformation and highlighting low expertise significantly reduced source credibility compared to the control, $t(297) = 11.49$, $p < .001$, Cohen's $d = 1.62$, and $t(297) = 17.71$, $p < .001$, Cohen's $d = 2.52$, respectively. However, highlighting low expertise reduced credibility ratings more than correcting disinformation, $t(297) = 6.35$, $p < .001$, Cohen's $d = .90$.[1]

We next examined the dimensions of trust and expertise from Ohanian's (1990) Semantic scale independently (see Supplemental Material E for full analyses). We found that both trust and expertise were reduced in comparison to control for correcting disinformation, $t(298) = 10.11$, $p < .001$; $t(298) = 9.60$, $p < .001$, respectively, and low expertise $t(298) = 14.39$, $p < .001$; $t(298) = 16.88$, $p < .001$, respectively. However, highlighting low expertise lowered both trust *and* expertise more than correcting disinformation, $t(298) = 4.33$, $p < .001$, $t(298) = 7.33$, $p < .001$, respectively.

Participants' feelings toward Dr. Smith mirrored the source credibility ratings. The one-way ANCOVA on feelings found a main effect of intervention, $F(2, 297) = 106.17$; $p < .001$; $MSE = 476.13$; $\eta_p^2 = .42$. As can be seen in Figure 2 (Panel D), both correcting disinformation and highlighting low expertise reduced feelings compared to control, $t(297) = 9.71$, $p < .001$, Cohen's $d = 1.37$;

[1] We preregistered that we would conduct a 2 × 3 repeated measures ANOVA, but the ANCOVA is more appropriate.

$t(297) = 14.26$, $p < .001$, Cohen's $d = 2.02$, respectively, with the low-expertise condition reducing feelings more than correcting disinformation, $t(297) = 4.59$, $p < .001$, Cohen's $d = .65$.

Individual Differences in Response to Interventions: CAM Beliefs

We next investigated whether participants' CAM beliefs were associated with their belief in disinformation postintervention (i.e., excluding the control condition). We found that participants' CAMBI scores were correlated with disinformation belief ($\rho = .23$, $p < .001$), such that people who endorsed CAM were more likely to believe in disinformation. Note that we report Spearman's correlations as some of the data were not normally distributed. Next, we correlated belief in disinformation with the CAMBI's subscales (similar to Brewer et al., 2019). We found that the Natural Treatments and Holistic Health subscales were significantly associated with belief in disinformation ($\rho = .23$; $p < .001$; $\rho = .29$, $p < .001$), but the Participation in Health Treatments subscale was not ($p = .685$). As shown by Supplemental Material F, Figure S4, this may be due to a ceiling effect. The restricted range shows that a large portion of our sample prefer an active and collaborative role in treatment decisions, limiting the ability for participation in treatment to correlate with other dimensions. For all subsequent analyses, we therefore created a new composite, the CAM belief, specifically the Natural Treatment and Holistic Health subscales (CAMBI–Natural and Holistic; CAMBI-NH), that only included the Natural Treatment and Holistic Health subscales. As can be seen in Figure 2 (Panel B), the CAMBI-NH was correlated with belief in disinformation in the intervention conditions ($\rho = .29$, $p < .001$). However, it is important to note that when we focused on participants who scored in the top tertile of the CAMBI-NH ($N = 100$), both the correcting disinformation and low-expertise interventions still significantly reduced belief in disinformation ($p < .001$; $p < .001$, respectively).

Next, we examined whether CAM beliefs were associated with changes in credibility and feelings toward the source. Instead of creating pre/post subtraction scores as per our preregistration, we calculated residuals due to their increased reliability (DeGutis et al., 2013; see Supplemental Material G for subtraction score analyses for all experiments). We found that both credibility and feelings residuals were correlated with CAMBI-NH ($\rho = .26$, $p < .001$; $\rho = .33$, $p < .001$, respectively), showing that people who endorsed CAM were less likely to reduce their credibility and feelings ratings postintervention. However, once again, participants who scored in the top tertile of CAMBI-NH still significantly reduced their perceived credibility ($p < .001$) and feelings toward the source ($p < .001$).

Finally, we investigated whether closeness to cancer was associated with our outcome measures. We found that closeness to cancer was not associated with credibility residuals ($p = .146$) or belief in disinformation ($p = .105$) but was associated with feelings residuals ($\rho = .17$, $p = .017$). This shows that people who reported a greater closeness to cancer were less likely to reduce their feelings toward the source postintervention.

Discussion

Although we hypothesized that correcting disinformation would result in lower disinformation belief than highlighting low expertise, we found that both interventions were equally effective. We also found that both intervention conditions reduced perceptions of credibility and feelings more than control, but that highlighting low expertise was more effective than correcting disinformation. Regarding CAM belief, we found that people who endorsed CAM were more likely to believe in disinformation and less likely to reduce evaluations of source credibility and feelings postintervention. However, both interventions were still effective for participants in the top tertile CAM belief. Finally, people who reported a closeness to cancer were less likely to reduce their feelings toward the source postintervention, although the small effect size should be noted ($\rho = .17$). We next sought to replicate and extend Experiment 1 by examining whether the combined efforts of correcting disinformation and highlighting low expertise could create a more effective intervention. We were additionally interested in whether e-health literacy was associated with belief in disinformation or credibility evaluations, given that people with lower health literacy have been found to be more susceptible to believing in health misinformation (Scherer et al., 2021).

Experiment 2

Method

Experiment 2 was conducted in May 2022. We hypothesized that (a) all interventions would reduce disinformation belief, source credibility, and feelings more than the control condition; (b) the combined condition would reduce belief in disinformation, source credibility, and feelings more than the other interventions; (c) CAM belief, specifically the Natural Treatment and Holistic Health subscales (CAMBI-NH), would correlate with higher belief in disinformation, higher perceived source credibility, and warmer feelings toward the source postintervention; and finally (d) higher e-health literacy scores would be associated with lower levels of belief in disinformation, source credibility ratings, and feelings about the information source.

Design

Experiment 2 was also a parallel-arm prospective longitudinal intervention design, with intervention arms correcting the source's disinformation, highlighting the source's low expertise, a control condition, and an added combined condition (correcting disinformation plus highlighting low expertise).

Participants

Experiment 2 was a representative sample of 602 U.S. citizens from Prolific. We planned to omit participants who showed uniform responding ($SD < 0.5$; $N = 0$), completed the task too quickly ($N = 0$), and indicated that their data should be excluded due to lack of effort ($N = 1$). Our final sample included 601 people with an age range between 18 and 93 years ($M = 45.23$, $SD = 16.25$), with 284 males, 302 females, and 15 individuals who were nonbinary, preferred to self-describe, or who chose not to disclose their gender. Participants were paid $1.80 for a 9-min task.

Procedure and Stimuli

The procedure and stimuli were similar to Experiment 1, with an added combined condition and eight-item e-Health Literacy Scale

(Norman & Skinner, 2006). The combined condition corrected two inaccuracies from the disinformation condition and half the content from the low expertise condition. The abbreviated interventions were to control for length effects, given that previous research has found that longer messages are more persuasive (Shen et al., 2015). We added another item "Do you think Dr. Smith is a credible source when it comes to cancer?" to the source credibility composite, such that it had 12 items. We also added another item "To what extent have you spent thinking about cancer in your lifetime?" (1 = *not at all* to 7 = *very much so*), such that our closeness to cancer measure was a two-item composite.

Sample Size Justification

Although Experiment 1 found large effect sizes for the omnibus ANOVA on disinformation belief ($\eta_p^2 = .27$), we boosted the sample to 600 to ensure that we could detect small differences between the new combined condition and the other interventions. According to G*Power, 600 participants should also be sufficient to conduct the one-way ANOVAs on disinformation belief with four levels (with $\alpha = .05$, power = .80, $f = .20$, recommended $N = 280$) and ANCOVA for the source credibility and feelings residuals (recommended $N = 277$).

Transparency and Openness

See Swire-Thompson et al. (2022) for preregistration (https://osf.io/tq7ma/), data, and analysis scripts.

Results

Our intervention and control groups were well-balanced across demographics, and there were no significant differences between groups in age ($p = .261$), education ($p = .303$), gender ($p = .468$), or partisanship ($p = .965$).

Belief in Disinformation

A one-way Welch's ANOVA found that disinformation belief differed between conditions, $F(3, 328.97) = 71.68$; $p < .001$, $MSE = 4.14$, $\eta_p^2 = .23$, with all intervention conditions lower than control, correcting disinformation: $t(298.24) = 11.43$, $p < .001$, Cohen's $d = 1.15$; low expertise: $t(270.84) = 10.23$, $p < .001$, Cohen's $d = 1.20$; and combined: $t(280.55) = 12.13$, $p < .001$, Cohen's $d = 1.34$. When excluding the control, a one-way Welch's ANOVA confirmed that there were no differences between the intervention conditions ($p = .219$), as can be seen in Figure 3 (Panel A).

Source Credibility and Feelings

We next conducted a one-way ANCOVA on the posttest credibility ratings, entering pretest credibility as a covariate. This revealed a significant main effect of intervention, $F(3, 596) = 170.55$; $p < .001$; $MSE = 1.94$; $\eta_p^2 = .46$. As can be seen in Figure 3 (Panel C), we found that highlighting low expertise reduced credibility more than the correcting disinformation condition, $t(596) = 5.42$, $p < .001$, Cohen's $d = .63$, but the combined condition reduced this further than the low-expertise condition alone, $t(596) = 2.69$, $p = .037$, Cohen's $d = .31$. When examining the dimensions of trust and expertise independently from Ohanian's (1990) Semantic scale, we found that all conditions reduced both dimensions in comparison to control. Highlighting low expertise reduced expertise more than correcting disinformation and reduced trust to the same extent as correcting disinformation. However, the combined condition reduced trust and expertise more than only highlighting low expertise (see Supplemental Material E for full analyses and details).

We next conducted a one-way ANCOVA on posttest feelings ratings, entering pretest feelings as a covariate. This revealed a main effect of intervention, $F(3, 596) = 114.89$; $p < .001$; $MSE = 576.92$; $\eta_p^2 = .37$. This highlighted that low expertise reduced feelings more than correcting disinformation, $t(596) = 3.32$, $p = .005$, Cohen's $d = .38$, and the combined condition reduced feelings more than highlighting low expertise alone, $t(596) = 2.96$, $p = .017$, Cohen's $d = .34$.

Individual Differences in Response to Intervention: CAM Beliefs

We found that participants' CAMBI scores were correlated with disinformation belief postintervention ($\rho = .30$, $p < .001$). Replicating Experiment 1, the Natural Treatment and Holistic Health subscales were associated with disinformation belief ($\rho = .33$, $p < .001$ and $\rho = .29$, $p < .001$, respectively), while the Participation in Health Treatments subscale was not ($p = .322$). As can be seen in Figure 3 (Panel B), our new composite, the CAMBI-NH, was correlated with disinformation belief ($\rho = .36$, $p < .001$). When focusing on participants who scored in the top tertile of CAMBI-NH ($N = 200$), correcting disinformation and the combined condition were still significantly lower in disinformation belief than control ($p < .001$ and $p < .001$, respectively), but low expertise does not cross the threshold of significance, $t(79.4) = 2.58$, $p = .056$, 95% CI [−1.55, .014]. We next investigated whether CAM beliefs were associated with credibility ratings or feelings toward the source. Indeed, CAMBI-NH scores were correlated with both source credibility residuals ($\rho = .32$, $p < .001$) and feelings residuals ($\rho = .33$, $p < .001$), showing that people who endorsed CAM were less likely to reduce their credibility and feelings ratings postintervention. However, participants who scored in the top tertile of CAMBI-NH again significantly reduced their credibility and feelings ratings postintervention, $F(3, 195) = 37.98$; $p < .001$; $MSE = 2.12$; $\eta_p^2 = .37$; $F(3, 195) = 23.75$; $p < .001$; $MSE = 651.03$; $\eta_p^2 = .27$, respectively.

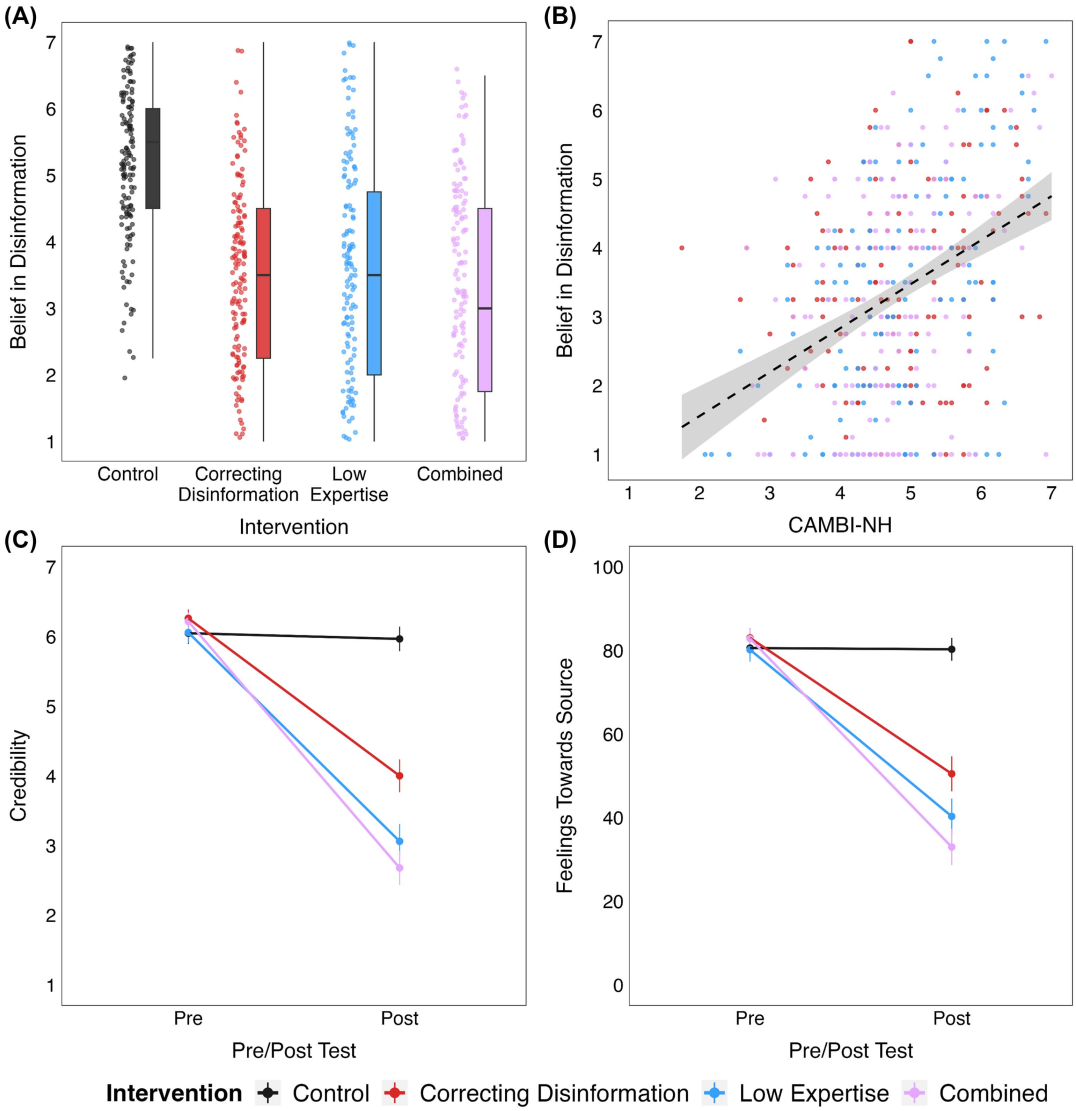
Finally, we investigated whether closeness to cancer or e-health literacy were associated with our outcome measures. We found that closeness to cancer was not correlated with belief in disinformation ($p = .075$), credibility residuals, ($p = .255$), or feelings residuals ($p = .281$). We also found that e-health literacy was not associated with credibility residuals ($p = .079$) nor feelings residuals ($p = .157$) but was related to disinformation belief ($\rho = .11$, $p = .022$). This reflects that people with higher self-reported e-health literacy were *more* likely to believe in disinformation.

Discussion

The findings of Experiment 2 replicated those of Experiment 1, with all interventions reducing belief in disinformation, source credibility, and feelings compared to control. Contrary to our hypothesis, the combined condition did not reduce belief in disinformation more than the other interventions alone. We instead found no significant differences between the intervention conditions. This finding aligns with Ecker et al. (2011), who found that once misinformation is

Figure 3

Experiment 2 (A) Disinformation Belief, (B) the Association Between CAMBI-NH and Postintervention Disinformation Belief (ρ = .36), (C) Source Credibility, and (D) Feelings Toward the Source



Note. Error bars are 95% confidence intervals. Black = control; red = correcting disinformation; blue = low expertise; purple = combined; dashed black = all interventions excluding control; CAMBI = Complementary and Alternative Medicine Beliefs Inventory; CAMBI-NH = CAM belief, specifically the Natural Treatment and Holistic Health subscales. See the online article for the color version of this figure.

believed, its endorsement cannot be reduced below a certain threshold, labeled the "level of irreducible persistence." However, the combined condition did reduce perceptions of source credibility and feelings toward the source more than highlighting low expertise alone. The current experiment also confirmed that people who endorsed CAM had higher belief in disinformation and were less likely to reduce their perceived credibility and feelings toward a source after being presented with a discreditation intervention. Finally, we found that higher e-health literacy was associated with *increased* disinformation beliefs, in contrast with previous findings (Scherer et al., 2021) and our hypotheses. This could reflect that participants with high self-reported health literacy are overconfident in their ability to evaluate information (see Kim et al., 2020).

Although the findings of Experiments 1 and 2 appear to be robust, there were several limitations in our study design. For instance, we did not examine whether the interventions lasted over time, and we did not account for pretest belief in the disinformation. Experiment 3 assessed belief in disinformation both prior to and after the interventions and included a 1-week delayed posttest. We also applied our paradigm to the topic of COVID-19 vaccination for several reasons: (a) to boost generalizability, (b) because the public had been saturated with corrective information about COVID-19 by the time of the experiments' launch, and (c) because we could focus on an unvaccinated population. Vaccination status could be a behavioral indicator that participants were more likely to believe in the disinformation (Loomba et al., 2021). Given that our previous interventions were extremely similar in their ability to reduce belief in disinformation, we also included a new intervention that highlighted the source's conflict of interest.

Experiment 3

Method

Experiment 3 was conducted in October 2022. We hypothesized that (a) all interventions would reduce belief in disinformation and source credibility more than the control condition; (b) these differences would remain after a 1-week period; (c) the conflict of interest intervention would reduce belief in new disinformation *more* than the other interventions, but there would be no difference between the low-expertise and correcting disinformation conditions (as per Experiments 1 and 2); (d) the low-expertise condition would reduce credibility ratings more than the correcting disinformation condition; and (e) participants' posttest belief in disinformation would be positively correlated with CAM belief and vaccine hesitancy. Focusing on unvaccinated and vaccinated groups independently was preregistered as an exploratory analysis. Please see Supplemental Material A, Table S1 for whether preregistered hypotheses were supported or rejected.

Design

This experiment was a parallel-arm prospective longitudinal intervention design. We assessed belief at three time points (pretest, immediate posttest, and 1-week delayed posttest), and the intervention arms included correcting disinformation, low expertise, conflict of interest, and a control condition.

Participants

Participants were 793 from Prolific over the age of 18. In order to have sufficient individuals who were vaccine hesitant, we recruited 50% of participants who reported to Prolific that they felt negatively about the COVID-19 vaccine and 50% of participants from the general Prolific population. Our exclusion criteria included participants who completed the task too quickly ($N = 0$), indicated that their data should be excluded due to lack of effort ($N = 2$), or did not complete both parts of the study ($N = 71$). Our final sample therefore included 720 people. Age ranged between 18 and 91 years (M = 39.09, SD = 12.8), with 353 males, 349 females, and 18 individuals who were nonbinary, preferred to self-describe, or who chose not to disclose their gender. There were 355 people who were vaccinated and 365 people who were unvaccinated. At the time of launch, 79.9% of the total U.S. population had received at least one vaccine (Centers for Disease Control and Prevention, 2024), and vaccines were readily available. Participants were paid $1.60 for an 8-min task in Part 1 and $0.80 for a 2-min task in Part 2.

Procedure and Stimuli

The procedure was similar to Experiment 2 with several changes. First, the vignettes were modified to be about a self-proclaimed medical researcher and vaccine expert, Dr. George Davis (see Supplemental Material C for details). In the pretest phase, participants were asked to rate their belief in four false claims regarding COVID-19 vaccines on a 7-point Likert scale. Importantly, these were *not yet attributed* to Dr. Davis but allowed us to get pretest evaluations of the claims. Participants were then randomly assigned to one of four conditions: correcting disinformation, low expertise, control, or new to this experiment, the conflict of interest condition. This condition described how Dr. Davis had been receiving money from lobbying groups and owned stocks in a company selling alternative medicines. Given that the feelings thermometer findings were identical to source credibility ratings in Experiments 1 and 2, they were omitted from this experiment. Additionally, participants completed the Vaccine Hesitancy Scale (a modified version of the 12-item Adult Vaccine Hesitancy Scale; Akel et al., 2021; Thaker, 2021) and were asked about their current vaccination status. Finally, participants were reinvited to take part in the delayed posttest 1 week later, which was identical to the immediate posttest.

Sample Size Justification

The effect size for disinformation belief was large in Experiment 1 (η_p^2 = .27) and Experiment 2 (η_p^2 = .23), and we expected Experiment 3 would require similar power. However, we boosted our sample to 800 participants to allow for attrition (≈15% expected) given that this was a two-wave study, and to alllow for subgroup analysis of participants who were unvaccinated.

Transparency and Openness

See Swire-Thompson and Kilgallen (2022) for preregistration (https://osf.io/7bjre), data, and analysis scripts.

Results

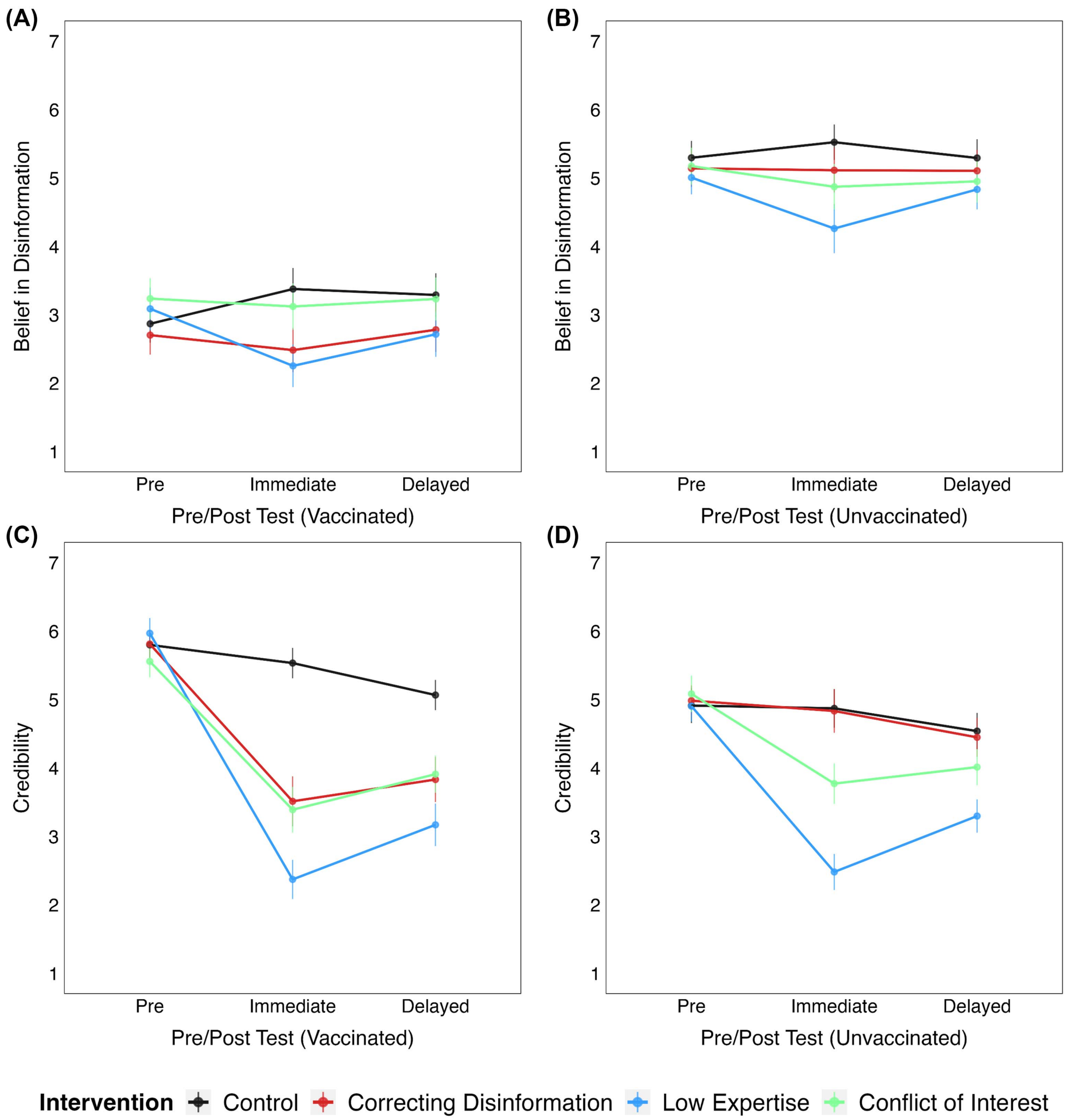
There were no significant differences between conditions for age (p = .908), gender (p = .174), education (p = .531), political partisanship (p = .966), ethnicity (p = .396), race (p = .638), and vaccination status (p = .971). For brevity, we report the following results with unvaccinated and vaccinated participants; see Supplemental Material H for omnibus analyses with all participants.

Belief in Disinformation

To examine how the interventions impacted disinformation belief in unvaccinated participants, we first conducted a 3 × 4 between-within ANOVA with between-subjects factors intervention (control vs. correcting disinformation vs. low expertise vs. conflict of interest) and within-subjects factors pre–post (pre vs. postimmediate vs. postdelayed). There was a significant main effect of intervention, $F(3, 361) = 4.39$; $p = .005$, $MSE = 4.84$, $\eta_p^2 = .04$, and a main effect of pre–post: $F(2, 722) = 6.24$; $p = .002$, $MSE = .66$, $\eta_p^2 = .02$, qualified by Intervention × Pre–Post interaction, $F(6, 722) = 6.99$; $p < .001$, $MSE = .66$, $\eta_p^2 = .06$. This showed that the impact of the interventions changed over time, as can be seen in Figure 4

Figure 4

Experiment 3 Belief in Disinformation for (A) Vaccinated Participants and (B) Unvaccinated Participants. Source Credibility for (C) Vaccinated Participants (D) Unvaccinated Participants.



Note. Error bars reflect 95% confidence intervals. Black = control; red = correcting disinformation; blue = low expertise; green = conflict of interest. See the online article for the color version of this figure.

(Panel B). Focusing on the immediate time point, we followed this up with a one-way ANCOVA, with pretest belief entered as a covariate. This revealed a main effect of intervention, $F(3, 360) = 13.18$, $p < .001$, $MSE = 1.28$, $\eta_p^2 = .10$, with the low-expertise and conflict of interest conditions both significantly lower than the control, $t(360) = 6.01$, $p < .001$, Cohen's $d = .89$; $t(360) = 3.23$, $p = .007$, Cohen's $d = .48$. Additionally, we found that the low-expertise condition reduced belief more than the conflict of interest condition, $t(360) = 2.79$, $p = .029$, Cohen's $d = .41$. However, the correcting disinformation condition showed no difference from control ($p = .364$). A one-way ANCOVA on delayed posttest belief showed that there were no significant differences between the conditions at the 1 week time point ($p = .319$), indicating that the differences in disinformation belief were confined to the immediate time point. We also conducted repeated measures ANOVAs (pre vs. postimmediate vs. postdelayed) for each intervention, confirming that there were no differences for the correcting disinformation condition ($p = .944$), but there were significant changes for the low expertise, $F(2, 186) = 14.28$, $p < .001$, $MSE = 1.00$, $\eta_p^2 = .13$, and conflict of interest conditions, $F(2, 180) = 3.99$, $p = .020$, $MSE = .56$, $\eta_p^2 = .04$.

The outcomes were similar for vaccinated participants. A 3 × 4 between–within ANOVA on disinformation belief with between-subjects factors intervention and within subjects factors pre–post revealed a significant Pre–Post × Intervention interaction,

$F(6, 702) = 9.80$; $p < .001$, $MSE = .69$, $\eta_p^2 = .08$. Again, this showed that the impact of the interventions changed over time. We followed this up with a one-way ANCOVA focusing on the immediate time point, with pretest belief entered as a covariate. A main effect showed that there were differences between the conditions, $F(3, 350) = 20.34$, $p < .001$, $MSE = 1.23$, $\eta_p^2 = .15$. All interventions were below control (correcting disinformation, $p < .001$, low expertise $p < .001$, and conflict of interest, $p = .008$), with low expertise having lower disinformation belief than correcting disinformation and conflict of interest ($p = .011$, $p < .001$, respectively). Focusing upon the delayed time point, a one-way ANCOVA found a significant main effect of intervention, $F(3, 350) = 6.05$, $p < .001$, $MSE = 1.31$, $\eta_p^2 = .05$. Belief remained significantly lower after 1 week than at pretest in the low-expertise condition, $t(350) = 4.24$, $p < .001$, Cohen's $d = .64$, while correcting disinformation ($p = .112$) and highlighting a conflict of interest ($p = .221$) were equivalent to control. It must be noted that pretest belief levels differed between groups, which highlights how important pre–post designs are, as a posttest-only design would have resulted in substantially different conclusions.

One caveat is that the control condition *increased* belief in the disinformation posttest, $t(174) = 2.42$, $p = .017$, 95% CI [.093, .923], likely because a seemingly reputable source was seen to be making these claims. To address this, we also conducted repeated measures ANOVAs (pre vs. postimmediate vs. postdelayed) for each intervention, confirming that there were no differences in the correcting disinformation condition ($p = .057$) nor conflict of interest condition ($p = .518$). However, there were significant changes in the low-expertise condition, $F(2, 170) = 18.64$; $p < .001$, $MSE = .81$, $\eta_p^2 = .18$. In sum, the only condition where belief in disinformation was reduced over a 1-week period was highlighting low expertise in the vaccinated group.

Source Credibility

To examine how the interventions impacted perceptions of source credibility in unvaccinated participants, we ran a 3 × 4 repeated measures ANOVA on credibility ratings with the within-subjects factor pre–post and the between-subjects factor intervention. We found a main effect of pre–post: $F(2, 722) = 123.97$, $p < .001$, $MSE = .87$, $\eta_p^2 = .26$, and a main effect of intervention: $F(3, 361) = 27.08$, $p < .001$, $MSE = 3.31$, $\eta_p^2 = .18$, qualified by an Intervention × Pre–Post interaction: $F(6, 722) = 33.25$, $p < .001$, $MSE = .87$, $\eta_p^2 = .22$. This suggests that the intervention's efficacy changed over time. Note that the assumption of sphericity was violated, but when correcting using the Greenhouse–Geisser method, all results replicated.

We followed this up with a one-way ANCOVA on the immediate posttest, with the pretest scores entered as a covariate. We found a significant main effect of intervention, $F(3, 360) = 70.65$, $p < .001$, $MSE = 1.64$, $\eta_p^2 = .37$. As can be seen in Figure 4 (Panel D), both highlighting low expertise and a conflict of interest significantly reduced perceived source credibility compared to control, $t(360) = 12.58$, $p < .001$, Cohen's $d = 1.86$; $t(360) = 6.19$, $p < .001$, Cohen's $d = .92$, respectively, while correcting disinformation did not ($p = .980$). Focusing upon the delayed time point, a one-way ANCOVA revealed a significant main effect of intervention, $F(3, 360) = 20.28$, $p < .001$, $MSE = 1.44$, $\eta_p^2 = .14$. Credibility ratings in the low-expertise and conflict of interest conditions were significantly lower than control, $t(360) = 6.98$, $p < .001$, Cohen's $d = 1.03$; $t(360) = 3.32$, $p = .005$, Cohen's $d = .50$, with the low-expertise condition reducing ratings even lower than conflict of interest, $t(360) = 3.64$, $p = .002$, Cohen's $d = .54$. Once again, correcting disinformation did not differ from control ($p = .909$).

Focusing on the vaccinated participants, we ran a 3 × 4 repeated measures ANOVA on credibility ratings with within-subjects factor pre–post and between-subjects factor intervention. We found an Intervention × Pre–Post interaction, $F(6, 702) = 40.80$, $p < .001$, $MSE = 1.04$, $\eta_p^2 = .26$, showing that the intervention's efficacy changed over time. We followed this up with a one-way ANCOVA on the immediate posttest, with the pretest scores as the covariate. We found a main effect of intervention, $F(3, 350) = 79.26$, $p < .001$, $MSE = 1.99$, $\eta_p^2 = .41$. Correcting disinformation, highlighting low expertise, and highlighting a conflict of interest all significantly reduced credibility ratings compared to control, $t(350) = 9.57$, $p < .001$, Cohen's $d = 1.43$; $t(350) = 15.12$, $p < .001$, Cohen's $d = 2.29$; $t(350) = 9.65$, $p < .001$, Cohen's $d = 1.44$, respectively. Focusing upon the delayed time point, a one-way ANCOVA also revealed a main effect of intervention, $F(3, 350) = 31.01$, $p < .001$, $MSE = 1.82$, $\eta_p^2 = .21$, with all interventions lower than control, correcting disinformation, $t(350) = 6.13$, $p < .001$, Cohen's $d = .92$; low expertise, $t(350) = 9.50$, $p < .001$, Cohen's $d = 1.44$; conflict of interest, $t(350) = 5.43$, $p < .001$, Cohen's $d = .81$. After 1 week, highlighting low expertise was still more effective than highlighting a conflict of interest and correcting disinformation, $t(350) = 4.12$, $p < .001$, Cohen's $d = .63$; $t(350) = 3.44$, $p < .001$, Cohen's $d = .52$.

Finally, we independently examined the dimensions of trust and expertise from Ohanian's (1990) Semantic scale (see Supplemental Material E for full analyses and details). For both vaccinated and unvaccinated participants, highlighting low expertise consistently reduced expertise more than the other interventions and trust more than (or equal to) the other interventions.

Individual Differences in Response to Intervention: CAM Beliefs and Vaccine Hesitancy

Although all CAMBI subscales correlated with belief in disinformation (see Supplemental Material F, Figure S4), we report CAMBI-NH results for consistency with the previous two experiments. CAMBI-NH scores were associated with belief in disinformation at pretest ($\rho = .40$, $p < .001$) and posttest ($\rho = .36$, $p < .001$). Given that we measured belief both pre- and postintervention, we were able to calculate disinformation belief residuals. We found that CAMBI-NH scores were correlated (albeit weakly) with disinformation belief residuals at immediate posttest ($\rho = .18$, $p < .001$) and after 1 week ($\rho = .12$, $p = .005$). To examine whether the CAMBI-NH scores had a different association with disinformation belief for each intervention, we conducted moderation analyses using the bootstrapping method. We found that the interventions similarly impacted disinformation belief regardless of CAMBI-NH score for both immediate ($p = .253$) and delayed posttests ($p = .844$). Regarding perceived credibility, CAMBI-NH scores were also associated with credibility residuals both immediately ($\rho = .17$, $p < .001$) and 1 week later ($\rho = .20$, $p < .001$). This suggested that people scoring high on the CAMBI-NH were less likely to reduce their credibility evaluations postintervention.

We finally examined whether vaccine hesitancy was associated with belief in disinformation and credibility ratings. We found that vaccine hesitancy was strongly correlated with belief in disinformation at pretest ($\rho = .82$, $p < .001$) and disinformation belief residuals at both immediate ($\rho = .32$, $p < .001$) and 1 week posttests ($\rho = .25$, $p < .001$). This illustrates that vaccine-hesitant participants were more likely to believe in disinformation initially and less likely to reduce belief in disinformation postintervention. Vaccine hesitancy was also correlated with credibility residuals immediately ($\rho = .29$, $p < .001$) and after 1 week ($\rho = .22$, $p < .001$), showing that vaccine-hesitant participants were less likely to reduce their source credibility ratings postintervention.

Discussion

Contrary to our hypotheses, only highlighting low expertise and a conflict of interest were effective at reducing belief in disinformation at the immediate posttest. Indeed, the only condition to last over the 1-week period was the low-expertise condition, exclusively in the vaccinated population. Correcting disinformation did not reduce disinformation belief in unvaccinated or vaccinated populations, likely because the public had been saturated with COVID-19 vaccine corrective information. Highlighting low expertise was the most effective intervention at reducing perceived credibility at both the immediate and delayed time points for both populations. For the vaccinated population, correcting disinformation was as effective as the conflict of interest condition for reducing credibility, but for the unvaccinated population it was equivalent to the control condition. Finally, participants who endorsed CAM and who were vaccine hesitant were less likely to reduce their belief in disinformation and perceived credibility postintervention.

General Discussion

The present study examined how people evaluated dubious health sources and their disinformation after various discreditation interventions. The first two experiments used cancer information, and the third used vaccine information. In Experiments 1 and 2, we found that all interventions (correcting disinformation, highlighting low expertise, and these combined) were equally effective at reducing belief in disinformation. Experiment 3 found that highlighting low expertise was more effective at reducing disinformation belief than highlighting a conflict of interest or correcting disinformation in both vaccinated and unvaccinated populations. Indeed, the only intervention that showed sustained belief reduction after 1 week was low expertise and only in the vaccinated population. Regarding perceived credibility and feelings toward the source, highlighting low expertise was consistently more effective than other interventions. The only condition that was more effective than highlighting low expertise alone was combining it with correcting disinformation in Experiment 2.

There are several potential reasons why highlighting low expertise was superior to the other interventions tested. First, highlighting low expertise not only reduced perceived expertise more than the other interventions, but it also reduced trust more than (or equal to) the other interventions. By revealing that the source did not have the expertise to provide accurate information, we also imply that they are untrustworthy, as they were lying or at least not portraying themselves accurately. In other words, our interventions were not clean dissociations of trust and expertise (Pornpitakpan, 2004). As such, we do not consider our findings in contrast to previous research that found trustworthiness to be more important than expertise when correcting misinformation (Ecker & Antonio, 2021; Guillory & Geraci, 2013). Another reason for the efficacy of the low-expertise condition could be that people are more concerned with domain expertise for health topics than for politics or unfolding news events like those used in prior studies (see Harris et al., 2016; Vraga & Bode, 2017).

It is also important to note the discrepancy in correcting disinformation's efficacy between the first two experiments and Experiment 3. In Experiments 1 and 2, correcting disinformation was as effective as the other interventions for reducing belief in disinformation. However, in Experiment 3, correcting disinformation was equivalent to the control for both vaccinated and unvaccinated populations. This is likely because the public was saturated with COVID-19 corrective information. As a result, people had likely heard the corrections previously, and if they maintained belief in the inaccuracy, it was not due to a lack of information. This suggests that, for topics where people are likely to already know the accurate information, it may be better to highlight other attributes of a dubious source instead of correcting their claims. Experiments 1 and 2 additionally found that correcting disinformation reduced feelings toward the source and source credibility ratings, in contrast to prior studies that used political stimuli (Aird et al., 2018; Nyhan et al., 2020; Swire-Thompson et al., 2020). This could be because politicians are expected to lie, whereas health professionals are not. To illustrate, 78% of U.S. adults thought nurses held high honesty and ethical standards, compared to only 8% of senators and 6% of members of Congress (Brenan & Jones, 2024). Experiment 3 found that the efficacy of correcting disinformation on reducing credibility depended on participants' vaccine status. If participants were vaccinated, correcting disinformation reduced credibility, but if they were not, it was equivalent to control. This could suggest that a violation of expected integrity is an important attribute for correcting disinformation to be effective, but future research should examine the mechanisms driving this difference (e.g., antiexpert attitudes, Motta, 2018, 2021).

We also examined several individual differences. We consistently found that people who endorsed CAM were not only more likely to believe disinformation initially, but less likely to reduce belief in disinformation, credibility ratings, and feelings postintervention. These findings align with previous research showing that people with positive attitudes toward CAM are more likely to believe in misinformation (Scherer et al., 2021). It also suggests that people who endorse CAM are not using traditional credibility cues, such as formal scientific training or track records of accuracy. For cancer care, this could result in adopting practices that delay evidence-based cancer treatment, refusing evidence-based cancer treatments, or using ineffective cancer treatments. We additionally examined self-reported closeness to cancer and vaccine hesitancy. In Experiment 1, we found weak evidence that people who were close to cancer were less likely to reduce their feelings toward the source postintervention, but this did not replicate in Experiment 2. By contrast, there was a strong association with vaccine hesitancy and initial belief in disinformation (aligning with Kricorian et al., 2022; Lee et al., 2022). Vaccine-hesitant participants were also less likely to reduce belief in disinformation and perceived credibility postintervention. It is logical that vaccine hesitancy plays a larger role than closeness to cancer, given that the former is more likely to be part of one's larger ideology and, thus, more influential for information processing, whereas the latter is an unlucky circumstance.

Though the findings of the present study are compelling, there are several limitations and opportunities for future research. First, in the present studies, we consider people who endorse complementary medicine and those who endorse alternative medicine as a homogenous group. Teasing these populations apart is an important next step, given that the decision to forego conventional cancer treatments for alternative treatments is uniquely different from seeking complementary treatments. Second, Briñol and Petty (2009) posited that if sources present weak arguments, less credible sources can produce more attitude change than credible sources. Future research could investigate the parameters of our findings by manipulating the strength of the argument made in the corrections. Third, it is unclear whether discreditation interventions are equally effective when the source is an individual or a media outlet (see Ecker et al., 2024). Finally, future research should consider specific interventions for people most susceptible to disinformation campaigns, such as those who believe in CAM or people who are vaccine hesitant.

Several practical recommendations can be made based upon the findings of the present study. First, practitioners should consider highlighting dubious health sources' low expertise. Bylund et al. (2023) found that clinicians currently address cancer misinformation using four strategies: working with the patients to understand the misinformation, correcting misinformation through education, advising patients about future online searches, and preserving the clinician–patient relationship. Highlighting low expertise has the advantage of reducing belief in disinformation as much as (or more than) other interventions, while optimally reducing perceived credibility and feelings toward the source. The combined low expertise plus correcting disinformation condition was also promising (Experiment 2), but the complete lack of efficacy when correcting disinformation in the unvaccinated population (Experiment 3) makes us cautious to recommend it for all audiences and information topics. Finally, practitioners should be aware that not all individuals will be equally responsive to interventions. For example, CAM belief and vaccine hesitancy present particular hurdles when discrediting disreputable health sources.

In sum, we found that highlighting low expertise was equivalent to or more effective than other interventions for reducing belief in disinformation and reducing the perceived credibility of dubious health sources. One unfortunate implication of the current findings is that people who *do* have expertise but spread disinformation nonetheless are the most damaging to the public (such as doctors; Larson, 2018). More nuanced interventions will need to be developed and tested to discredit these individuals. Nonetheless, for typical health disinformation sources who lack expertise, it appears that highlighting that they are unqualified, inexperienced, or unable to provide accurate information is a promising option for fact-checkers and practitioners.

Constraints on Generality

Experiments 1 and 2 were conducted in a U.S. representative sample to foster generalizability. Our sample was representative according to sex, age, and ethnicity, but did not take partisanship or other traits into consideration. Furthermore, our findings may not generalize as well to countries outside of the United States with different levels of trust for health professionals. In the United States, 56% of people trust doctors, yet around the world trust can be higher (such as Spain at 68%) or lower (such as South Korea at 38%; Global Trustworthiness Index, 2023). Although, it is important to note that, even in South Korea, doctors are the second most-trusted profession, behind only scientists. Experiment 3 did not recruit a representative sample but had a wide age range and equal gender representation (between 18 and 91 years with 353 males and 349 females). Our main constraint on generalizability is that our study was experimental, and our stimuli likely differed from real in-person or online disinformation sources (Simons et al., 2017). For example, the relationships that people have with real disinformation sources will often be longer than our 1 week period. Therefore, it would be ideal to replicate this study in a real-world setting.

References

Aird, M. J., Ecker, U. K. H., Swire, B., Berinsky, A. J., & Lewandowsky, S. (2018). Does truth matter to voters? The effects of correcting political misinformation in an Australian sample. *Royal Society Open Science*, *5*(12), Article 180593. https://doi.org/10.1098/rsos.180593

Akel, K. B., Masters, N. B., Shih, S.-F., Lu, Y., & Wagner, A. L. (2021). Modification of a vaccine hesitancy scale for use in adult vaccinations in the United States and China. *Human Vaccines & Immunotherapeutics*, *17*(8), 2639–2646. https://doi.org/10.1080/21645515.2021.1884476

Ambrosone, C. B., Zirpoli, G. R., Hutson, A. D., McCann, W. E., McCann, S. E., Barlow, W. E., Kelly, K. M., Cannioto, R., Sucheston-Campbell, L. E., Hershman, D. L., Unger, J. M., Moore, H. C. F., Stewart, J. A., Isaacs, C., Hobday, T. J., Salim, M., Hortobagyi, G. N., Gralow, J. R., Budd, G. T., & Albain, K. S. (2020). Dietary supplement use during chemotherapy and survival outcomes of patients with breast cancer enrolled in a cooperative group clinical trial (SWOG S0221). *Journal of Clinical Oncology*, *38*(8), 804–814. https://doi.org/10.1200/JCO.19.01203

American National Election Studies. (n.d.). *The ANES guide to public opinion and electoral behavior*. Retrieved January 5, 2023, from https://electionstudies.org/data-tools/anes-guide/

Barnes, R. M., Johnston, H. M., MacKenzie, N., Tobin, S. J., & Taglang, C. M. (2018). The effect of ad hominem attacks on the evaluation of claims promoted by scientists. *PLOS ONE*, *13*(1), Article e0192025. https://doi.org/10.1371/journal.pone.0192025

Bishop, F. L., Yardley, L., & Lewith, G. (2005). Developing a measure of treatment beliefs: The complementary and alternative medicine beliefs inventory. *Complementary Therapies in Medicine*, *13*(2), 144–149. https://doi.org/10.1016/j.ctim.2005.01.005

Brenan, M., & Jones, J. M. (2024). Ethics ratings of nearly all professions down in U.S. *Gallup News*. https://news.gallup.com/poll/608903/ethics-ratings-nearly-professions-down.aspx

Brewer, N. J., Turrise, S. L., Kim-Godwin, Y. S., & Pond, R. S., Jr. (2019). Nurses' knowledge and treatment beliefs: Use of complementary and alternative medicine for pain management. *Journal of Holistic Nursing*, *37*(3), 248–259. https://doi.org/10.1177/0898010118822212

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

Bylund, C. L., Mullis, M. D., Alpert, J., Markham, M. J., Onega, T., Fisher, C. L., & Johnson, S. B. (2023). Clinician communication with patients about cancer misinformation: A qualitative study. *JCO Oncology Practice*, *19*(3), e389–e396. https://doi.org/10.1200/OP.22.00526

Centers for Disease Control and Prevention. (2024). *COVID data tracker*. https://covid.cdc.gov/covid-data-tracker/

Chou, W. S., Oh, A., & Klein, W. M. P. (2018). Addressing health-related misinformation on social media. *JAMA*, *320*(23), 2417–2418. https://doi.org/10.1001/jama.2018.16865

Collins, P. J., Hahn, U., von Gerber, Y., & Olsson, E. J. (2018). The bi-directional relationship between source characteristics and message content. *Frontiers in Psychology*, *9*, Article 18. https://doi.org/10.3389/fpsyg.2018.00018

DeGutis, J., Mercado, R. J., Wilmer, J., & Rosenblatt, A. (2013). Individual differences in holistic processing predict the own-race advantage in recognition memory. *PLOS ONE*, *8*(4), Article e58253. https://doi.org/10.1371/journal.pone.0058253

Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology*, *36*(4), 424–435. https://doi.org/10.1037/0022-3514.36.4.424

Ecker, U. K. H., & Antonio, L. M. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition*, *49*(4), 631–644. https://doi.org/10.3758/s13421-020-01129-y

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

Ecker, U. K. H., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review*, *18*(3), 570–578. https://doi.org/10.3758/s13423-011-0065-1

Ecker, U. K. H., Prike, T., Paver, A., Scott, R., & Swire-Thompson, B. (2024). *Don't believe them! Reducing misinformation influence through source discreditation*. PsyArXiv. https://doi.org/10.31234/osf.io/5yj2v

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Flanagin, A. J., Winter, S., & Metzger, M. J. (2020). Making sense of credibility in complex information environments: The role of message sidedness, information source, and thinking styles in credibility evaluation online. *Information Communication and Society*, *23*(7), 1038–1056. https://doi.org/10.1080/1369118X.2018.1547411

Global Trustworthiness Index. (2023). *Ipsos Global Trustworthiness Index 2023: Who does the world trust?* Ipsos. https://www.ipsos.com/sites/default/files/ct/news/documents/2023-10/Ipsos-global-trustworthiness-index-2023.pdf

Green, R., Toribio-Flórez, D., & Douglas, K. M. (2023). Impressions of science and healthcare professionals who share anti-science conspiracy theories. *Routledge Open Research*, *2*, Article 37. https://doi.org/10.12688/routledgeopenres.17965.1

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition*, *2*(4), 201–209. https://doi.org/10.1016/j.jarmac.2013.10.001

Harris, A. J., Hahn, U., Madsen, J. K., & Hsu, A. S. (2016). The appeal to expert opinion: Quantitative support for a Bayesian network approach. *Cognitive Science*, *40*(6), 1496–1533. https://doi.org/10.1111/cogs.12276

Hess, A. C., Dodds, S., & Rahman, N. (2022). The development of reputational capital—How social media influencers differ from traditional celebrities. *Journal of Consumer Behaviour*, *21*(5), 1236–1252. https://doi.org/10.1002/cb.2074

Hughes, M. G., Griffith, J. A., Zeni, T. A., Arsenault, M. L., Cooper, O. D., Johnson, G., Hardy, J. H., Connelly, S., & Mumford, M. D. (2014). Discrediting in a message board forum: The effects of social support and attacks on expertise and trustworthiness. *Journal of Computer-Mediated Communication*, *19*(3), 325–341. https://doi.org/10.1111/jcc4.12077

Johnson, S. B., Park, H. S., Gross, C. P., & Yu, J. B. (2018a). Complementary medicine, refusal of conventional cancer therapy, and survival among patients with curable cancers. *JAMA Oncology*, *4*(10), 1375–1381. https://doi.org/10.1001/jamaoncol.2018.2487

Johnson, S. B., Park, H. S., Gross, C. P., & Yu, J. B. (2018b). Use of alternative medicine for cancer and its impact on survival. *Journal of the National Cancer Institute*, *110*(1), 121–124. https://doi.org/10.1093/jnci/djx145

Kelman, H. C., & Hovland, C. I. (1953). Reinstatement of the communicator in delayed measurement of opinion change. *Journal of Abnormal Psychology*, *48*(3), 327–335. https://doi.org/10.1037/h0061861

Kim, S., Park, K., & Jo, H. S. (2020). Gap between perceived eHealth literacy and ability to use online cancer-related information. *Journal of Korean Medical Science*, *35*(24), Article e187. https://doi.org/10.3346/jkms.2020.35.e187

Kricorian, K., Civen, R., & Equils, O. (2022). COVID-19 vaccine hesitancy: Misinformation and perceptions of vaccine safety. *Human Vaccines & Immunotherapeutics*, *18*(1), Article 1950504. https://doi.org/10.1080/21645515.2021.1950504

Larson, H. J. (2018). The biggest pandemic risk? Viral misinformation. *Nature*, *562*(7727), Article 309. https://doi.org/10.1038/d41586-018-07034-4

Lee, J. Y., & Sundar, S. S. (2013). To tweet or to retweet? *That is the question for health professionals on Twitter. Health Communication*, *28*(5), 509–524. https://doi.org/10.1080/10410236.2012.700391

Lee, S. K., Sun, J., Jang, S., & Connelly, S. (2022). Misinformation of COVID-19 vaccines and vaccine hesitancy. *Scientific Reports*, *12*(1), Article 13681. https://doi.org/10.1038/s41598-022-17430-6

Lindeman, M. (2011). Biases in intuitive reasoning and belief in complementary and alternative medicine. *Psychology & Health*, *26*(3), 371–382. https://doi.org/10.1080/08870440903440707

Lobera, J., & Rogero-García, J. (2021). Scientific appearance and homeopathy. Determinants of trust in complementary and alternative medicine. *Health Communication*, *36*(10), 1278–1285. https://doi.org/10.1080/10410236.2020.1750764

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the U.K. and USA. *Nature Human Behaviour*, *5*(3), 337–348. https://doi.org/10.1038/s41562-021-01056-1

Motta, M. (2018). The enduring effect of scientific interest on trust in climate scientists in the United States. *Nature Climate Change*, *8*(6), 485–488. https://doi.org/10.1038/s41558-018-0126-9

Motta, M. (2021). Republicans, not democrats, are more likely to endorse anti-vaccine misinformation. *American Politics Research*, *49*(5), 428–438. https://doi.org/10.1177/1532673X211022639

Nadarevic, L., Reber, R., Helmecke, A. J., & Köse, D. (2020). Perceived truth of statements and simulated social media postings: An experimental investigation of source credibility, repeated exposure, and presentation format. *Cognitive Research: Principles and Implications*, *5*(1), Article 56. https://doi.org/10.1186/s41235-020-00251-4

Norman, C. D., & Skinner, H. A. (2006). eHEALS: The eHealth Literacy Scale. *Journal of Medical Internet Research*, *8*(4), Article e27. https://doi.org/10.2196/jmir.8.4.e27

Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2020). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability. *Political Behavior*, *42*(3), 939–960. https://doi.org/10.1007/s11109-019-09528-x

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39–52. https://doi.org/10.1080/00913367.1990.10673191

Peterson, J. S., Swire-Thompson, B., & Johnson, S. B. (2020). What is the alternative? Responding strategically to cancer misinformation. *Future Oncology*, *16*(25), 1883–1888. https://doi.org/10.2217/fon-2020-0440

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social

Psychology, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Scherer, L. D., McPhetres, J., Pennycook, G., Kempe, A., Allen, L. A., Knoepke, C. E., Tate, C. E., & Matlock, D. D. (2021). Who is susceptible to online health misinformation? A test of four psychosocial hypotheses. *Health Psychology*, *40*(4), 274–284. https://doi.org/10.1037/hea0000978

Shen, F., Sheer, V. C., & Li, R. (2015). Impact of narratives on persuasion in health communication: A meta-analysis. *Journal of Advertising*, *44*(2), 105–113. https://doi.org/10.1080/00913367.2015.1018467

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Susmann, M. W., & Wegener, D. T. (2023). The independent effects of source expertise and trustworthiness on retraction believability: The moderating role of vested interest. *Memory & Cognition*, *51*(4), 845–861. https://doi.org/10.3758/s13421-022-01374-3

Swire-Thompson, B., Bodenger, J., Johnson, S., & Rossen, I. (2021). *Discrediting disinformation sources*. https://osf.io/kd9sh/

Swire-Thompson, B., Ecker, U. K., Lewandowsky, S., & Berinsky, A. J. (2020). They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation. *Political Psychology*, *41*(1), 21–34. https://doi.org/10.1111/pops.12586

Swire-Thompson, B., & Johnson, S. (2024). Cancer: A model topic for misinformation researchers. *Current Opinion in Psychology*, *56*, Article 101775. https://doi.org/10.1016/j.copsyc.2023.101775

Swire-Thompson, B., Johnson, S., & Bodenger, J. (2022). *Discrediting disinformation sources: Experiment 2*. https://osf.io/tq7ma/

Swire-Thompson, B., & Kilgallen, K. (2022). *Discrediting vaccine disinformation sources*. https://osf.io/7bjre

Swire-Thompson, B., & Lazer, D. (2022). Reducing health misinformation in science: A call to arms. *The Annals of the American Academy of Political and Social Science*, *700*(1), 124–135. https://doi.org/10.1177/00027162221087686

Thaker, J. (2021). The persistence of vaccine hesitancy: COVID-19 vaccination intention in New Zealand. *Journal of Health Communication*, *26*(2), 104–111. https://doi.org/10.1080/10810730.2021.1899346

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

Vraga, E. K., & Bode, L. (2020). Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Communication*, *37*(1), 136–144. https://doi.org/10.1080/10584609.2020.1716500

Vraga, E. K., Bode, L., & Tully, M. (2022). The effects of a news literacy video and real-time corrections to video misinformation related to sunscreen and skin cancer. *Health Communication*, *37*(13), 1622–1630. https://doi.org/10.1080/10410236.2021.1910165

Wertgen, A. G., & Richter, T. (2020). Source credibility modulates the validation of implausible information. *Memory & Cognition*, *48*(8), 1359–1375. https://doi.org/10.3758/s13421-020-01067-9

Westbrook, V., Wegener, D. T., & Susmann, M. W. (2023). Mechanisms in continued influence: The impact of misinformation corrections on source perceptions. *Memory & Cognition*, *51*(6), 1317–1330. https://doi.org/10.3758/s13421-023-01402-w

Received July 26, 2023
Revision received May 14, 2024
Accepted May 30, 2024 ■