

Benefits and Pitfalls of Debunking Interventions to Counter mRNA Vaccination Misinformation During the COVID-19 Pandemic

Science Communication
2022, Vol. 44(5) 531–558
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10755470221129608
journals.sagepub.com/home/scx



Philipp Schmid^{1,2}  and Cornelia Betsch^{1,2} 

Abstract

Misinformation about mRNA vaccination is a barrier in the global fight against the COVID-19 pandemic. Thus, authorities often rely on text-based refutations as a countermeasure. In two experiments ($N = 2,444$), text-based refutations effectively reduced the belief in misinformation and immunized participants against the impact of a misleading social media post. However, a follow-up ($N = 817$) questions the longevity of these debunking and prebunking effects. Moreover, the studies reveal potential pitfalls by showing a row of unintended effects of the refutations (lacking effect on intentions, backfire-effects among religious groups, and biased judgments when omitting information about vaccine side effects).

Keywords

persuasion, genetics, spirituality, vaccine hesitancy, inoculation theory

¹University of Erfurt, Germany

²Bernhard-Nocht-Institute for Tropical Medicine, Hamburg, Germany

Corresponding Author:

Philipp Schmid, Media and Communication Science, University of Erfurt, Nordhäuser Straße 63, Erfurt, 99089, Germany.

Email: philipp.schmid@uni-erfurt.de

The development of effective vaccines was a milestone and one of the greatest challenges in the pandemic response to the coronavirus disease 2019 (COVID-19). Among the most promising and now widely approved vaccines are so-called messenger RNA (mRNA) vaccines. In contrast to conventional protein-based vaccines, mRNA vaccines contain genetic information for a single element of the virus rather than live attenuated or inactivated pathogens (Abbasi, 2020; Pardi et al., 2018). The approved COVID-19 mRNA vaccines have been shown to be effective in reducing the probability of a COVID-19 infection, reducing the transmission of the disease and, if infected, reducing the severity of the disease progression (Centers for Disease Control and Prevention [CDC], 2021). In addition, mRNA vaccines are considered beneficial over conventional vaccines because, among other things, these new types of vaccines allow rapid manufacturing, which is essential for a successful pandemic response (Jackson et al., 2020; Pardi et al., 2018). However, the mere availability of promising prevention measures does not guarantee public acceptance of such measures. In fact, in many countries it remains questionable whether the public acceptance of COVID-19 vaccines will be sufficient to achieve a coverage rate that will allow pandemic control (MacPherson, 2020). Thus, vaccine acceptance is widely discussed as another major challenge after the development of an effective COVID-19 vaccine itself (Habersaat et al., 2020; Lazarus et al., 2021).

One reason why vaccine acceptance is likely to remain too low despite overwhelming scientific evidence in favor of vaccination is the impact of misinformation on individuals' judgments and health decision-making (Betsch, Schmid, et al., 2018). According to general health behavior theories (Ajzen, 1991) and domain-specific models of vaccine hesitancy (Betsch, Schmid, et al., 2018), misinformation can influence specific attitudinal beliefs about vaccination and thereby reduce an individual's intention to get vaccinated. For example, if an individual encounters the widespread misinformation that mRNA vaccines alter the human genome (Löffler, 2021), then this can reduce the individual's confidence in the safety of mRNA vaccines and ultimately reduce the intention to get vaccinated against COVID-19. The theoretically negative impact of such misinformation has been demonstrated in empirical evaluations during the course of the pandemic (Loomba et al., 2021).

However, individuals who are facing misinformation are not necessarily helpless. For example, individuals can evaluate the credibility of prominent dubious claims by consulting external sources (Farrell et al., 2019; Walter et al., 2020). Some external sources such as fact-checkers and health authorities even support individuals' credibility evaluations by providing explicit refutations of prominent misinformation. For example, scientists have refuted the misinformation that mRNA vaccines alter the human genome because,

among other reasons, the mRNA does not enter the nucleus where DNA is located (Hotez et al., 2021). Thus, the World Health Organization (WHO) presents this refutation as part of their COVID-19 MythBusters initiative on their webpages for the general public (WHO, 2021). Like WHO, many other fact-checkers rely on text-based approaches for the correction of misinformation. The ways in which text-based refutations are usually distributed vary from social media posts (Kim & Walker, 2020) and information leaflets at medical practices (Betsch, Rossmann, et al., 2018) to serious online games (Roozenbeek & van der Linden, 2019). Although there is some debate surrounding the unintended effects of refutations, challenging misinformation rather than ignoring it has been demonstrated repeatedly to be the superior strategy in mitigating its potential damage (Chan et al., 2017; Roozenbeek & van der Linden, 2019; Schmid & Betsch, 2019; Walter et al., 2021). Thus, providing text-based refutations of COVID-19 misinformation during the pandemic should support the public in evaluating the credibility of dubious claims. In fact, a recent evaluation of the MythBusters initiative revealed promising results for their effectiveness in social media (Vraga & Bode, 2021).

However, empirical evidence on the effectiveness of text-based refutations of vaccination misinformation during a pandemic is scarce and the extent to which corrective effects endure over time remains underexplored. Moreover, the introduction of COVID-19 mRNA vaccines may provide specific challenges for the design of text-based refutations of vaccination misinformation that are rarely addressed in previous research on vaccination misinformation. First, a text-based refutation about new mRNA vaccines may challenge the aversion to new technologies and scientific innovations that is assumed to be particularly prevalent among religious individuals (Allum et al., 2014; Brossard et al., 2009; Nisbet, 2005). Second, a text-based refutation about new mRNA vaccines based on scientific evidence may challenge the belief that truth can only be found through personal experience—a belief that is assumed to be particularly prevalent among spiritual individuals (Rutjens & van der Lee, 2020). Thus, both religiosity and spirituality may be barriers for the effectiveness of text-based refutations. Thus, the current studies aim to provide empirical evidence on (a) the effectiveness of text-based refutations of misinformation regarding COVID-19 mRNA vaccines during the COVID-19 pandemic and (b) examine the role of religiosity and spirituality in the success of refuting vaccination misinformation.

Designing Text-Based Refutations

Designing effective refutations does not mean simply to label misinformation as false. In fact, research suggests that refutations are most promising when

they consist of specific components (Lewandowsky et al., 2012; van der Linden et al., 2017). One key component is to introduce a clear warning of the falsehood of the misinformation and to name the misinformation (Ecker et al., 2020; Lewandowsky et al., 2012, 2020). While repeating misinformation is inadvisable, mentioning the misinformation a single time can help belief-updating, that is the formation of a new attitude or knowledge base (Ecker et al., 2017). In addition, adding a warning that is directly connected to the misinformation makes it unlikely that the correction will be ignored or that the misinformation will be misinterpreted as a fact (Lewandowsky et al., 2012, 2020). Another key component is to provide a detailed causal explanation of why the misinformation is incorrect (Ecker et al., 2020; Lewandowsky et al., 2020). According to research on mental models, refutations that only state the falsehood of misinformation do not provide enough detail (Lewandowsky et al., 2012). As a consequence, individuals can experience difficulties in remembering which proposition was correct or incorrect (Chan et al., 2017). The two outlined components are key parts of recommended best-practice for effective refutations and should also be applicable to the context of misinformation around COVID-19 mRNA vaccines. Thus, the *refutation hypothesis* predicts that debunking texts that contain the relevant components will reduce individuals' credibility judgments of prominent misinformation about COVID-19 mRNA vaccines compared to a control group. Based on studies on the long-term effectiveness of refutations (Maertens et al., 2021), we also expect that the refutation effect will be detectable after a 2-month delay.

Refutations as Prevention Measures

Refutations generally work via two distinct processes of belief updating. On one hand, refutations may serve as an intervention, given that individuals already believe in the misinformation; in this way, they serve as *debunking* (Chan et al., 2017; Lewandowsky et al., 2012). On the other hand, they may serve as prevention when individuals are entirely unfamiliar with the misinformation or when they encounter persuasive misinformation after receiving the refutation; in this way, they may serve as *prebunking* (Chan et al., 2017; Lewandowsky & van der Linden, 2021). For example, individuals reading the refutation of the MythBuster initiative on the WHO webpages may benefit from it because the refutation either corrects their false belief that mRNA vaccines alter the human genome (i.e., debunking), or it warns unknowing people that this misinformation is spreading and protects against future encounters with it (i.e., prebunking). While introducing a clear warning of the falsehood of the misinformation and providing a detailed causal explanation

of why the misinformation is incorrect are components of effective debunking texts, these components may also serve as prebunking for reasons outlined in the following paragraph.

The design of effective prebunking interventions is usually based on two components that originate from inoculation theory (Lewandowsky & van der Linden, 2021; McGuire & Papageorgis, 1961). First, a weakened form of the misinformation is introduced, and individuals are warned that they may encounter this misinformation. This threat induction puts individuals in a state of arousal that makes them more sensitive to being a potential target for misinformation (Compton & Ivanov, 2012). Second, it provides a refutation of the misinformation, giving individuals strong counterarguments to use when they are exposed to misinformation at a later point in time. This process is called refutational preemption (Lewandowsky & van der Linden, 2021). Following this rationale, introducing a clear warning regarding the misinformation in a debunking text may serve as a threat induction, while providing a detailed causal explanation of why the misinformation is incorrect may serve as refutational preemption. Thus, the *prebunking hypothesis* predicts that texts that contain relevant components of effective debunking will mitigate the impact of subsequent encounters with prominent misinformation about COVID-19 mRNA vaccines on individuals' credibility judgments compared to a control group. According to previous research, misinformation on vaccination also decreases individuals' confidence in vaccination and individuals' intentions to get vaccinated (Loomba et al., 2021; Schmid & Betsch, 2019). Thus, we also expect that the intervention texts reduce the negative impact of misinformation on confidence and intention measures.

Religiosity, Spirituality, and Unintended Effects

Despite the overall recommendation to challenge rather than ignore misinformation some researchers have raised concerns about potential backfire-effects when refuting misinformation. Backfire effects describe the counterintuitive finding that some individuals report higher (instead of lower) beliefs in misinformation after having received the refutation of the very same misinformation (Lewandowsky et al., 2020; Swire-Thompson et al., 2020). However, several studies have found no evidence for the presence of backfire-effects (Ecker et al., 2017, 2019; Wood & Porter, 2019). Thus, a recent review has concluded that backfire effects are not a robust empirical phenomenon and that it is unlikely that corrections cause such unintended effects on the group level (Swire-Thompson et al., 2020). While backfire effects seem to be a rare phenomenon, corrections may still prove ineffective under certain conditions. For example, refuting misinformation seems to be

least effective when the misinformation aligns with the worldviews of the audience (van der Linden et al., 2017). This may also be an issue when refuting misinformation about mRNA vaccines (Dragojlovic & Einsiedel, 2013). In fact, increased religiosity has repeatedly been found to be associated with increased opposition to emerging technologies like stem-cell research, nanotechnology, and gene-technology (Allum et al., 2014; Brossard et al., 2009; Nisbet, 2005). The typical narratives are either that the new technology “interferes with God’s creation” or that the technology is deemed “unnatural” (Dragojlovic & Einsiedel, 2013). While many religious medical expert communities and leaders of world religions have explicitly expressed their support toward COVID-19 vaccination (British Islamic Medical Association, 2020; Card, 2020; Jewish Medical Doctors working in UK, 2020), some concerns with the emerging mRNA technologies may remain as objections to vaccination among religious individuals or single religious leaders. These objections, in turn, may foster motivated reasoning and lower the effectiveness of refutations of misinformation about mRNA vaccines. Thus, the *religiosity as moderator hypothesis* predicts that the effectiveness of debunking texts about mRNA vaccines decreases with increasing levels of individual religiosity.

In addition, some recent research suggests that spirituality plays a central role in explaining the differences in overall levels of support for vaccination (Rutjens et al., 2022). Spiritual individuals do not necessarily feel a sense of belonging to a particular traditional religion or church and tend to see their own inner selves as the gateway to truth (Rutjens & van der Lee, 2020). Therefore, spiritual individuals’ own experiences are often considered the primary method of discovery, and aggregate research data seem of little relevance to their own belief-updating (Rutjens et al., 2022). Measures of spirituality may thus cover a different spectrum of believers who oppose vaccination as part of a specific worldview rather than an objection to emerging human-made technologies. Thus, the *spirituality as moderator hypothesis* predicts that the effectiveness of debunking texts about mRNA vaccines decreases with increasing levels of spirituality.

Other unintended effects may occur when evaluating interventions to counter misinformation. For example, a debunking text may effectively reduce credibility judgments simply because it demands that participants distrust any subsequent information (Roozenbeek & van der Linden, 2019). Moreover, debunking texts may be designed in such a way that they talk people into vaccination, rather than informing them about the pros and cons of vaccination. Thus, the *unintended effects research question* explores the impact of debunking texts on credibility judgments of unrelated facts and misleading pro-vaccine statements.

Overview

Based on two online experiments ($N = 2,444$), this contribution provides insights into whether debunking texts can reduce individuals' credibility judgments of prominent misinformation about COVID-19 mRNA vaccines (Experiment 1) and whether debunking texts can also serve as prebunking against the impact of subsequent misinformation (Experiment 2). Experiment 1 includes a 2-month follow-up measure of credibility to assess the long-term effects of debunking. Moreover, Experiment 2 includes the confidence in the safety of mRNA vaccines and the intention to get vaccinated against COVID-19 with mRNA-vaccines as additional outcome measures. Furthermore, we analyzed the role of religiosity (Experiment 1) and spirituality (Experiment 2) as potential moderators of the effectiveness of debunking texts and explored potential unintended effects of debunking texts.

Experiment I

In this experiment, we tested the refutation hypotheses. We expected that debunking texts that contain relevant components of effective debunking reduce individuals' credibility judgments of prominent misinformation about COVID-19 mRNA vaccines compared to a control group. We expected that this effect would occur immediately upon receipt of the debunking and after a delay of 2 months. Furthermore, we tested the religiosity as moderator hypothesis. That is, we expected that the effectiveness of debunking texts about COVID-19 mRNA vaccines decreases with increasing levels of religiosity. Finally, we explored the unintended effects research question. That is, we explored whether debunking texts also affect individuals' credibility judgments of unrelated facts or misleading pro-vaccine statements about COVID-19 mRNA vaccines compared to a control group. Following best practices, we only label preregistered hypothesis as confirmatory hypothesis and all other analyses as exploratory research questions (Wagenmakers et al., 2012).

Methods

The hypotheses, measures, and analyses for this experiment were preregistered. The preregistration protocol, data, and script for all analyses can be accessed from the open-access repository OSF (Schmid & Betsch, 2022).

Design and Participants. Each participant was assigned randomly to one of two conditions, resulting from the 2 (debunking vs. control; between

subjects) \times 3 (measurement before vs. immediately after the debunking versus 2 months after the debunking; within subjects) mixed design. An a priori power analysis conducted with G*Power (Faul et al., 2009) revealed a minimum of $n = 835$ participants to detect an effect size of $d = 0.25$ ($\alpha = 0.05$) with a power of 95% at the 2-month follow-up questionnaire. We expected a high drop-out rate of around 33% because the experiment was embedded in a larger survey; thus, we aimed to recruit at least 1,250 individuals for the onset of the experiment. The expected effect size was informed by the lower end of the confidence interval of the effectiveness of corrections from a previous meta-analysis (Walter et al., 2021). The external panel provider respondi.de invited and incentivized the participants. The provider used a quota sampling procedure to match the distribution of gender \times age with that of the German population (Supplementary Table 1). A total of $N = 1,387$ individuals participated in the online experiment, and $n = 821$ individuals finished the 2-month follow-up. Seven individuals participated twice at the same measurement date. Data of the second participations were excluded from analyses and data at the 2-month follow-up was only analyzed if participants did not participate multiple times at the first measurement date. Thus, the following analyses use the final sample sizes of $n = 1,382$ (age Mean $M = 45.40$, Standard Deviation $SD = 15.69$; 49.7% female) and $n = 817$ (age $M = 48.02$, $SD = 14.78$; 50.1% female) for the follow-up. For the 2-month follow-up, we missed our planned sample size by $n = 18$ and reached a final power of 94.5% to detect a small effect size, which is still higher than for most psychological studies (Bakker et al., 2012). Dropping out was neither significantly correlated with subjective credibility judgments of misinformation at baseline (Pearson's $r = -.01$, $p = .834$), nor with differences in credibility judgments between baseline and second measurement ($r = -.01$, $p = .846$), nor with experimental condition ($r = -.00$, $p = .894$). Thus, regarding the primary variables under investigation, we did not find evidence for any systematic dropout. In the follow-up sample, some participants were already vaccinated ($n = 33$) as the German vaccination campaign started on December 26, 2020. We did not exclude vaccinated individuals from analyses, but provide confirmatory analyses without vaccinated individuals as a robustness check.

Procedure. Participants indicated their self-reported religiosity and the perceived credibility of four fictitious news headlines. Depending on conditions, they then either received the debunking text or a control text (Figure 1). The credibility measures were again assessed immediately after the manipulation and at a 2-months follow-up.

Components of texts	Debunking/Prebunking condition Experiment 1 and 2	Control condition Experiment 1	Control condition Experiment 2
General facts	Messenger RNA vaccines are hopefuls in the pandemic Different production methods are currently being tested in the development of a vaccine against the coronavirus. One possibility for vaccine production uses modern genetic engineering to produce so-called messenger RNA (mRNA) vaccines. (...)	Messenger RNA vaccines are hopefuls in the pandemic Different production methods are currently being tested in the development of a vaccine against the coronavirus. One possibility for vaccine production uses modern genetic engineering to produce so-called messenger RNA (mRNA) vaccines. (...)	What has gotten better with the updates? As with many other digital applications, the development of the federal government's Corona warning app did not end with its launch. (...)
Debunking Component 1 Warning points to misinformation	Some people believe in the myth that vaccination with an mRNA vaccine can change the human genome. This concern is unfounded.	—	—
Debunking Component 2 Explains why misinformation is wrong	The mRNA does not enter the human cell nucleus, where the human genetic material is in the form of DNA, during vaccination. Even if the mRNA were to enter the cell nucleus, a combination of mRNA from vaccines and human DNA is considered chemically impossible. Therefore, the material of a coronavirus vaccination cannot change the human genome.	—	—
Repetition of general facts	(...) The fact is that mRNA vaccines are hopefuls in the pandemic, because these vaccines have the advantage over conventional vaccines that many vaccine doses can be produced within a few weeks.	(...) The fact is that mRNA vaccines are hopefuls in the pandemic, because these vaccines have the advantage over conventional vaccines that many vaccine doses can be produced within a few weeks.	—

Figure 1. Excerpts of the experimental stimuli.

Note. Participants in Experiment 1 received facts about mRNA vaccines both in the control and the intervention conditions. In the debunking/prebunking conditions participants additionally received a warning regarding the misinformation and an explanation why the misinformation was wrong (Lewandowsky et al., 2012). Participants in the control condition of Experiment 2 received an unrelated text with no reference to vaccines. Full materials can be accessed in Supplementary Table 2.

Materials. The debunking text specifically addressed the myth that mRNA vaccines can alter human DNA. The text included (a) a warning of the falsehood of the misinformation and (b) provided a detailed explanation of why the misinformation was incorrect. Figure 1 provides an excerpt of the core components of the full debunking text. The content of the debunking text was designed in consultation with infectious disease specialists from the Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany. The full text is available in Supplementary Table 2.

The text in the control condition was identical regarding the information about the mRNA vaccines but missed the two central components of effective debunking: information that points to misinformation and an explanation of why the misinformation is wrong (Figure 1).

Measures. The relevant measures are listed below and ordered as they appeared in the questionnaire. Mean and standard deviations of primary measures stratified by condition are reported as Supplementary Table 3.

Religiosity. Self-reported religiosity was measured with a validated German version (Klumparendt & Drenckhan, 2014) of the Religious Commitment Scale (A. B. Cohen et al., 2006). The original scale measures religious commitment with six items (e.g., How religious are you?). The authors of

the German version omitted the sixth item (How spiritual are you?) because this item measures spirituality rather than religiosity and, in line with our own thinking, spirituality is considered as a separate trait. Thus, we used the mean of the 5-item solution as an indicator of individuals' religiosity. The commitment was measured on a verbally anchored 7-point scale ranging from 1 = not at all to 7 = very much. Reliability analysis in the final sample showed a sufficiently high reliability score for the religiosity scale, Cronbach's alpha = .96.

Credibility judgments of (mis)information. Individuals judged the perceived credibility of a single false antivaccine news headline (mRNA vaccine alters the human genome) on a 7-point scale ranging from 1 = *not at all credible* to 7 = *very credible*. This item served as the primary outcome variable. The approach of judging news headlines as a primary measure was adapted from previous prebunking studies (Roozenbeek & van der Linden, 2019). We also measured the credibility of two factually correct statements (The next federal election in Germany is in 2021; The Summer Olympics are to be held in Tokyo) to test whether the debunking intervention merely demanded that participants judge any headline as not at all credible. Moreover, we measured the credibility of a misleading pro-vaccine headline (mRNA vaccine does not cause any side effects) to test whether the debunking intervention simply persuaded participants to falsely believe that side effects of vaccination do not exist. All credibility judgments were measured before and immediately and 2 months after individuals received the debunking or control text.

Additional variables. Participants also answered other pandemic-related questions reported elsewhere, as the experiment was embedded in a larger COVID-19 survey. The full list of variables of the survey can be accessed here: <http://dx.doi.org/10.23668/psycharchives.4398>.

Analysis. We used linear regression models to test the hypotheses and explore the research question. A significance level of $\alpha = .05$ was employed for all analyses. The models were calculated with the PROCESS macro in *R* (Model 1: Hayes, 2013). Tables displaying the results of the models are presented in Supplementary Table 4 to 7. Predictor variables were mean-centered. Thus, the regression coefficients *B* in the Results section represent estimated unstandardized mean differences between the debunking group and the control group on outcome measures when controlling for all other predictors in the models. All models were controlled for baseline values of the outcome measures. In addition to the outlined regression models, we also preregistered separate linear models without religiosity as a predictor to test the

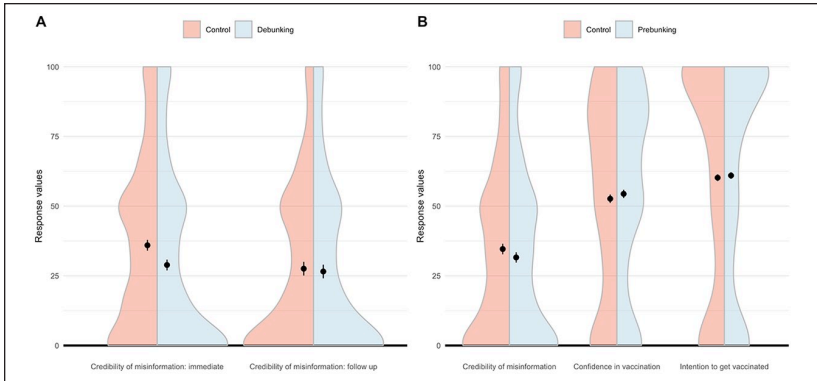


Figure 2. Effect of text-based debunking interventions on credibility judgments of misinformation in Experiment 1 (A) and effect of text-based prebunking on credibility judgments of misinformation, confidence in mRNA vaccines, and the intention to get vaccinated in Experiment 2 (B).

Note. Point estimates represent estimated means of primary outcomes, controlled for baseline values and model-specific moderators. Point estimates are reported in POMP values (percentage of maximum possible score). That is, scales range from 0 to 100, independent of the range of the original scale. Error bars are 95% confidence intervals.

refutation hypothesis in Experiment 1. However, these separate models were dropped due to scarcity in the Results section of this article because the full regression models include all effects of interest. We report the results of the separate linear models in Supplementary Table 8 for transparency. Results for the refutation hypotheses did not change when dropping religiosity as a predictor from the model. All continuous variables are reported as percentages of maximum possible scores of the original scales (POMP: P. Cohen et al., 1999). This linear transformation simplifies the interpretation of model parameters because all continuous scales range from 0 to 100. For example, a value of 50 on a POMP scale can be translated into 50% of the maximum possible score of the original scale independent of its original range.

Results

Test of Refutation Hypotheses. As shown in Figure 2A, participants who received the debunking text rated the false antivaccine news headline that mRNA vaccines alter the human genome as less credible compared to the control condition, $b = -7.06$, $[-9.81, -4.31]$, $t(1377) = -5.03$, $p < .001$. This is in line with the refutation hypothesis. However, we did not find a significant difference in credibility judgments between conditions at the

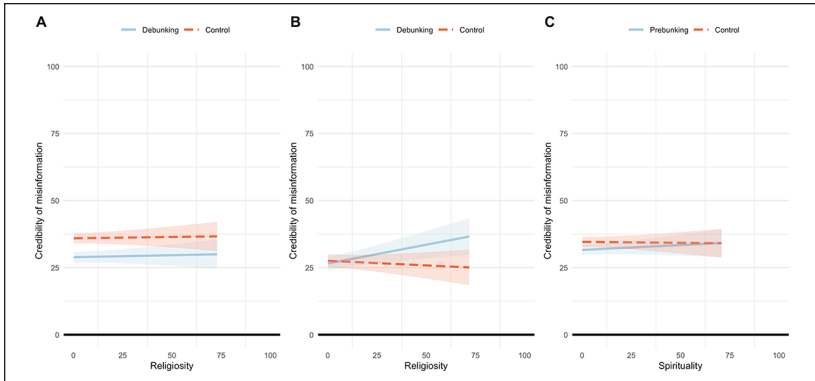


Figure 3. Credibility judgments of misinformation for intervention and control condition at different levels of religiosity for the immediate measurement (Experiment 1: A), the 2-month follow-up (Experiment 1: B), and at different levels of individuals' spirituality (Experiment 2: C).

Note. Lines represent estimated regression lines and shades are 95% confidence intervals. Estimates are reported in POMP values (percentage of maximum possible score). That is, scales range from 0 to 100, independent of the range of the original scale.

2-months follow-up, $b = -1.01$, $[-4.48, 2.47]$, $t(810) = -0.57$, $p = .569$. Explorative analyses revealed that this absence of a long-term effect did not result from a decay in the effectiveness of the debunking. In fact, as can be seen in Figure 2A, credibility judgments were slightly lower at the 2-months follow-up, compared to the immediate measurement, $t(406) = -0.87$, $p = .386$, Cohen's $d = .04$. This effect occurred in both conditions alike, that is, the control group also rated the misinformation as less credible when asked 2 months later, $t(409) = -5.74$, $p < .001$, $d = .28$, which neutralized the overall long-term advantage of the debunking intervention. Thus, the debunking text effectively decreased the believe in the false antivaccine news headline, but this effect was only apparent in the short term. General long-term effects may be masked by learning effects in the control group.

Test of Religiosity as Moderator Hypothesis. As a next step, we tested the *religiosity as moderator hypothesis*, that is, whether effects of the debunking text varied by individuals' religiosity (Figure 3A–B). As shown in Figure 3A, we did not find that the effectiveness of debunking on individuals' credibility judgments was moderated by religiosity—at least in the short term, $b = 0.01$, $[-0.09, 0.10]$, $t(1377) = 0.10$, $p = .918$. At the 2-month follow-up, however, we found a significant interaction effect between religiosity and debunking, $b = 0.17$, $[0.05, 0.30]$, $t(810) = 2.80$, $p = .005$. People with higher religiosity

demonstrated a backfire effect, such that after 2 months, highly religious people in the debunking condition judged misinformation as more credible than people in the control group (Figure 3B). Thus, while in the short run, debunking texts about mRNA vaccines seemed to work well, they may have triggered unintended effects in the long run when applied to highly religious groups.

The follow-up analyses were repeated without vaccinated participants (Supplementary Table 9). The patterns of results did not differ.

Exploration of Unintended Effects Research Question. To test whether the debunking intervention merely demanded participants to judge any headline as less credible, we asked individuals to also judge the credibility of factually correct statements before and after receiving the debunking. We did not find a significant decrease in credibility judgments for factually correct statements in the debunking condition, compared to the control condition (Supplementary Tables 5–6). In fact, participants in the debunking condition rated the factually correct statement about the Olympics as more credible compared to the control condition at the 2-month follow-up (Supplementary Table 6). Moreover, participants who received the debunking reported higher credibility judgments of the misleading news headline that the “mRNA vaccine does not cause any side effects” compared to the control group, $b = 3.77$, [1.28, 6.26], $t(1377) = 2.97$, $p = .003$. Thus, the debunking text may have created the false impression that because mRNA vaccines do not alter human DNA, they may have no side effects at all. However, we did not find the unintended effect of debunking at the 2-month follow-up, $b = -0.77$, [-4.49, 2.95], $t(810) = -0.41$, $p = .684$. An additional unsuccessful attempt to address the unintended effect with an additional fact box intervention at the 2-months follow-up is reported in Supplementary Results and Supplementary Table 10.

Additional Explorative Analyses. The religiosity as moderator hypothesis presumes that religiosity favors a worldview that is associated with lower support for vaccination. We indeed found that higher religiosity was associated with higher credibility judgments of misinformation at baseline ($r = .106$, $p = <.001$; Supplementary Figure 1A). This finding supports the notion that there is a positive association between religiosity and belief in misinformation about mRNA vaccination.

Experiment 2

In the second experiment, we tested the prebunking hypothesis. We expected that texts that contain relevant components of effective debunking will mitigate the impact of subsequent encounters with prominent misinformation about COVID-19 mRNA vaccines on individuals' credibility judgments

compared to a control group. Furthermore, we tested the spirituality as moderator hypothesis. That is, we expected that the effectiveness of the corrective text about COVID-19 mRNA vaccines decreases with increasing levels of spirituality. In line with Experiment 1, we explored the unintended effects research question. Moreover, Experiment 2 includes the confidence in the safety of mRNA vaccines and the intention to get vaccinated against COVID-19 with mRNA-vaccines as additional outcome measures. Again, preregistered hypothesis are labeled as confirmatory hypothesis and all other analyses as exploratory research questions.

Method

The hypotheses, measures, and analyses for this experiment were preregistered. The preregistration protocol, data, and script for all analyses can be accessed from the open-access repository OSF (Schmid & Betsch, 2022).

Design and Participants. Each participant was assigned randomly to one of two conditions, resulting from the 2 (prebunking vs. control; between subjects) \times 2 (measurement before prebunking versus immediately after the misinformation; within subjects) mixed design. Again, G*Power (Faul et al., 2009) was used to estimate the required sample size. Compared to Experiment 1, we aimed to detect smaller effect sizes of $d = .20$ with a high statistical power of 90% and thus aimed to recruit 1,054 unvaccinated individuals ($\alpha = 0.05$). Because we were interested in individuals' intentions to get vaccinated, we recruited only unvaccinated individuals for Experiment 2. A total of 1,056 individuals participated in the online experiment. Two individuals participated twice. Data of the second participations were excluded from analyses, resulting in the final sample size of $n = 1,054$ (age $M = 45.49$, $SD = 15.35$; 50.3% female).

Procedure. Participants first reported their spirituality and then judged the credibility of the same news headlines as participants in Experiment 1. As additional dependent variables, they then indicated their confidence in the mRNA vaccine and their intention to get vaccinated against COVID-19 as the baseline measure. Depending on conditions, they then either received the prebunking text or a control text (Figure 1), followed by the misinformation in the form of a misleading social media post. The same intervention text from Experiment 1 served as prebunking with the expectation that having read the prebunking can prevent the negative effect of the misinformation. The credibility, confidence, and intention measures were then repeated immediately after receiving the misinformation. Finally, participants received an

explicit debriefing that corrected the misinformation and provided official information about the safety of mRNA vaccines.

Materials. The prebunking text was identical to the debunking text of Experiment 1 (Figure 1), with one exception. As a precaution to avoid an increase in the false belief that mRNA vaccines do not cause any side effects, we added the disclaimer that “As with conventional vaccines, vaccine reactions and side effects can occur after mRNA vaccination. According to the Robert Koch-Institute, the most common reactions are pain at the injection site, fatigue and headaches.”

In this experiment, we also aimed to detect differences in individuals’ confidence in mRNA vaccines and their behavioral intention to get vaccinated. Following the rationale of general health behavior theories (Ajzen, 1991), the effect of the debunking text should be larger on credibility and confidence judgments and smaller on actual behavioral intentions. Thus, we replaced the conservative control condition from Experiment 1 with a less-challenging control condition to increase the power for detecting the weaker effects on the intention to get vaccinated (column 3 in Figure 1). The text of the new control condition was about technical information of updates for a Corona warning application for mobile phones and did not refer to mRNA vaccines whatsoever (Supplementary Table 2). As misinformation, all participants received an adaptation of a social media post that circulated in German social media groups at the time of the study and contained the claim that mRNA vaccines can alter the human genome. A translated version of the misinformation is provided as Supplementary Table 11.

Measures. The relevant measures are listed below and ordered as they appeared in the questionnaire. Mean and standard deviations of primary measures stratified by condition are reported in Supplementary Table 3.

Spirituality. Participants rated the following two items “To what extent do you consider yourself to be a spiritual person?” and “To what extent do others consider you to be a spiritual person?” on verbally anchored 7-point scales ranging from 1 = *not at all* to 7 = *to a strong extent*. The selected items were adapted from Rutjens and van der Lee (2020). Reliability analysis in the final sample showed a sufficiently high reliability score for the spirituality scale, Spearman Brown coefficient = .91.

Credibility of (mis)information. Individuals judged the perceived credibility of the same false and factually correct statements as in Experiment 1. The factually correct statement about Olympic games was dropped from Experiment

2 because, right before data collection, the media started to report about the possibility that the Olympic games may not take place at all.

Confidence in COVID-19 vaccination. Individuals judged their confidence in COVID-19 mRNA vaccines with a single item (I am completely confident that mRNA vaccines are safe) rated on a verbally anchored 7-point scale from 1 = *strongly disagree* to 7 = *strongly agree*. The selection of this item was informed by Betsch, Schmid, et al. (2018) and adapted to COVID-19.

Intention to get vaccinated. Individuals rated their intention to get vaccinated with an mRNA vaccine on a verbally anchored 7-point scale ranging from 1 = *I would definitely not get vaccinated* to 7 = *I would definitely get vaccinated*.

Additional variables. We also measured participants' judgments of other, vector-based COVID-19 vaccines. This allowed us to explore whether the misinformation had a specific impact on the attitudes and intentions toward mRNA vaccines or whether there were spillover effects to vector-based vaccines, too. These additional analyses are reported in Supplementary Methods and Results.

Analysis. We used the same preregistered linear regression models to test the predicted main effects and interaction effects as described in Experiment 1. The models only differed in that religiosity was replaced by spirituality as the moderator variable. Moreover, additional models including confidence in vaccination and intention to get vaccinated as outcome measures were analyzed. Tables displaying the results of all models are presented in Supplementary Tables 3, 4, 6, and 12.

Results

Test of Prebunking Hypothesis. As can be seen in Figure 2B, the prebunking text mitigated the impact of the subsequent antivaccination social media post. The prebunking text effectively served as an “immunization” against misinformation, as credibility judgments in the prebunking condition were lower than in the control condition, $b = -3.01$, $[-5.66, -0.36]$, $t(1,049) = -2.23$, $p = .026$. This benefit was, however, not observed for individuals' intention to get vaccinated, $b = 0.78$, $[-0.92, 2.48]$, $t(1,049) = 0.90$, $p = .368$. There was also no effect of the prebunking on confidence in mRNA vaccines, $b = 1.74$, $[-0.37, 3.85]$, $t(1049) = 1.62$, $p = .106$.

Spirituality as Moderator Hypothesis. We did not find that the effectiveness of prebunking on individuals' credibility judgments was moderated by spirituality, $b = 0.04$, $[-0.05, 0.14]$, $t(1,049) = 0.89$, $p = .375$ (Figure 3C). Likewise, we did not find any statistically significant interaction effects of prebunking and spirituality on individuals' intention or confidence in vaccination (Supplementary Table 12).

Exploration of Unintended Effects Research Question. Similar to Experiment 1, we did not find a significant decrease in credibility judgments for the factually correct statement in the prebunking condition as compared to the control condition (Supplementary Table 5). Unlike in Experiment 1, and potentially owing to the additional disclaimer that pointed to nonsevere side effects, we found no evidence that individuals who received the prebunking rated the misleading pro-vaccine statement as more credible (Supplementary Table 7).

Additional Explorative Analyses. The spirituality as moderator hypothesis presumes that spirituality favors worldviews that are associated with lower support for vaccination. We indeed found that higher spirituality was associated with higher credibility judgments of misinformation at the baseline measurement ($r = .212$, $p \leq .001$), lower levels of confidence in vaccination ($r = -.118$, $p \leq .001$) and a lower intention to get vaccinated ($r = -.118$, $p \leq .001$; Supplementary Figure 1B). While spirituality may not moderate the effectiveness of prebunking, these findings support the notion that spirituality, just like religiosity (Experiment 1), is positively associated with the belief in misinformation about mRNA vaccination.

In some cases, prebunking effects may not be detected because the misinformation itself has no effect. Comparing premeasures and postmeasures in the control condition, however, revealed that in this experiment, the social media message indeed had an effect: it increased the perceived credibility of the false headline, $t(515) = 2.07$, $p = .039$, $d = 0.09$, and decreased the intention to get vaccinated with an mRNA vaccine, $t(515) = -3.52$, $p < .001$, $d = 0.15$. No such effect was found on individuals' confidence in mRNA vaccines, $t(515) = -0.04$, $p = .968$, $d = < 0.01$. Thus, protective effects from the prebunking can be assumed, at least for judgments of the credibility of false headlines and the intention to get vaccinated with an mRNA vaccine.

Discussion

When new scientific innovations such as mRNA vaccines are introduced, they often provide a breeding ground for doubts, misunderstandings and misinformation. Results from the refutation hypothesis in Experiment 1 showed

that text-based debunking attempts can effectively reduce the perceived credibility of misinformation about mRNA vaccines—at least in the short run. We did not observe a benefit of the text-based debunking on credibility judgments after a 2-month delay. However, the results provide no evidence that this absence of a benefit was due to a decay of the effect in the intervention group. Rather, the result of lower credibility judgments of the control group at the 2-months follow-up may account for this. This change in the control group's perception may not be surprising, given the fact that several health authorities began to provide corrective information about the misinformation that mRNA alters human DNA in several media channels in Germany (e.g., Robert Koch-Institut [RKI], 2020)—corrective information that was available to both groups between the two measurements of Experiment 1. Thus, the results of Experiment 1 strengthen previous findings about the effectiveness of refutations (Chan et al., 2017; Vraga & Bode, 2021; Walter et al., 2020) by revealing their usefulness in the midst of a pandemic. In the light of the findings, we recommend to maintain and increase current active efforts by health authorities and fact-checkers to provide explicit corrections of misinformation (e.g., MythBusters; WHO, 2021), since they are an effective and economical measure of health communication during a pandemic.

The recommendation to use debunkings in times of a health crisis is also supported by theoretical considerations. According to research on the theory of motivated reasoning, people are motivated to be accurate, that is, they attend to relevant information more carefully when they have more incentive to invest effort in processing complex information (Bolsen & Druckman, 2015; Taber & Lodge, 2006). The possibility of preventing the risks of a potentially lethal COVID-19 infection may provide a good-enough incentive to invest effort into processing debunking texts about vaccination during a pandemic.

The results from the religiosity as moderator hypothesis suggest that accuracy may not be the only prevalent motivation when processing debunking about mRNA vaccination misinformation. In fact, we found evidence for a backfire effect of text-based debunking among religious individuals 2 months after debunking. According to research on the theory of motivated reasoning, people are motivated to arrive at a particular conclusion, that is, they attend more to information that is consistent with their prior beliefs when their worldviews are challenged by new information (Bolsen & Druckman, 2015, 2018). A debunking text about new mRNA vaccines may have challenged the aversion to new technologies and scientific innovations among religious individuals. As a response, they may have generated counterarguments against debunking, thus strengthening their initial worldviews.

But why does this effect occur only after 2 months and not immediately after reading the information? There is evidence that individuals who are motivated to arrive at a particular conclusion still do not openly lie to themselves (Taber & Lodge, 2006) but that a desired conclusion is drawn only if they can maintain the *illusion of objectivity* (Kunda, 1990). That is, they search for an apparently objective justification for their conclusion and do not realize that the process of building that justification is biased by focusing on specific memories that only confirm their conclusion (Kunda, 1990; Taber & Lodge, 2006). Maintaining this illusion may be more difficult immediately after being confronted with a disconfirming debunking text than after a delay of 2 months—when the debunking can be ignored and a memory search can be directed toward counterarguments that support the aversion against new technologies such as mRNA vaccination. In line with this, previous researchers have suggested that worldview backfire effects may, in fact, only occur after some delay (Wittenberg & Berinsky, 2020).

It is important to note that the sample in Experiment 1 was not very religious, and thus confirmatory replications, including highly religious groups, are needed to confirm the existence of a backfire effect among this target group. Moreover, replicating this finding is necessary before drawing conclusions for practitioners, since backfire effects after debunking are considered a rare phenomenon, and replications of previous backfire effects have failed (Haglin, 2017). Furthermore, using a single-item measure (as in this study) can further be unreliable (Swire-Thompson et al., 2020), which can also create false impressions of backfire effects (Swire-Thompson et al., 2022).

During the absence of further studies, giving religious leaders a voice when designing mRNA debunking texts may be useful (Valente & Pumpuang, 2007). Many leaders of world religions have explicitly expressed their support for vaccination against COVID-19 (British Islamic Medical Association, 2020; Card, 2020; Jewish Medical Doctors working in UK, 2020). Communicating this fact, along with debunking myths around mRNA vaccination, may reduce biased reasoning among religious groups.

Exploring the unintended effects research question also revealed potential pitfalls when designing refutations. Our findings from Experiment 1 suggest that debunking texts about COVID-19 vaccines can increase the false belief that COVID-19 vaccinations never cause any side effects. We assume this occurred because the debunking did not adequately address the potential risks of vaccination. Mentioning potential side effects in Experiment 2 prevented this unintended effect. Thus, to support informed decision-making and to build trust (Habersaat et al., 2020), we recommend reporting side effects in a transparent manner when designing debunking interventions about the risks of vaccination.

Results from the prebunking hypothesis in Experiment 2 revealed that materials designed for debunking can also serve as prebunking, and thereby protect individuals against subsequent misleading information, for example, from social media posts. These results are in line with previous findings about the effectiveness of corrections as prebunking interventions (Tay et al., 2022; Vraga et al., 2020), and further underline the usefulness of simple text-based debunkings during a pandemic. Moreover, the results are in line with research on the theory of motivated reasoning that suggests that prebunking can be particularly effective because it may address two motivated reasoning approaches (Bolsen & Druckman, 2015, 2018). First, individuals motivated to be accurate may use prebunking information as relevant information in their subsequent reasoning. Second, unknowing individuals that are not motivated to invest efforts in accuracy may use the prebunking to form a prior belief that is then defended against any subsequent (mis)information, regardless of its accuracy (Bolsen & Druckman, 2015).

Interestingly, we did not find evidence that prebunking protected against the impact of misleading social media posts on individuals' intention to get vaccinated. Maybe this was a cost of simply turning the debunking into a prebunking without further changes to the intervention. Sophisticated prebunking attempts explicitly warn individuals that they will be potential targets of disinformation (Lewandowsky & van der Linden, 2021). This warning is intended to trigger a feeling of threat and increase the receiver's motivation to resist the maleficent attempt (van der Linden et al., 2020). The prebunking text used in this study instead warned individuals about the misinformation but did not explicitly raise the expectation that one will be a potential target of that misinformation in the near future. Moreover, typically the misleading argumentation is explicitly explained during the prebunking, which provides an additional explanation of why misinformation is false and allows individuals to generalize their knowledge to other topics (Ecker et al., 2022). The prebunking text used in this study focused only on the content and did not uncover the fallacies present in the social media post. Finally, the debunking text started with a technical summary of general facts (Figure 1), which was probably not particularly engaging, and may have made the intervention less effective due to a lack of engagement and interest.

We found no evidence for the spirituality as moderator hypothesis in Experiment 2. In light of the findings from Experiment 1, we assume that the moderating effects may only show after a delay. Future studies should investigate that assumption.

The studies have several limitations. Experiment 1 and Experiment 2 were conducted at different times during the pandemic. No mRNA vaccine had been approved yet by the European Medicines Agency at the start of

Experiment 1, while vaccines were already being administered during the 2-months follow-up, and around 31 million Germans were already vaccinated at least once when Experiment 2 started (Supplementary Figure 2). At the time of Experiment 2, approximately 69.4 million of the total population in Germany (83.2 million) were over 18 years of age and thus eligible to be vaccinated. In relation to this group, 31 million vaccinated individuals correspond to a vaccination rate of about 44.7%. Thus, comparisons between the experiments need to be treated with caution because the results may have been influenced by time and context. However, assessing the effectiveness of debunking texts at different stages of the pandemic is also a unique strength of the current studies because it reveals the usefulness of simple, often-used and recommended text-based approaches at different stages of a pandemic. Another limitation is the rather highly educated sample pool. Debunking texts need to be understood in order to be effective. Thus, disseminating the chosen interventions in lower-educated or less-literate populations could lead to a decrease in the observed effectiveness. Other formats such as videos may be more suitable and should be tested in similar manners as prebunking and debunking interventions. Finally, both samples reported low levels of religiosity and spirituality. Thus, assumptions about how debunking affects the credibility judgments in highly religious or spiritual groups are limited and effects may be underestimated.

Conclusion

Text-based debunking containing detailed refutations can be an effective approach to counter misinformation about mRNA vaccination during a pandemic. In fact, detailed refutations can also serve as prebunking, that is, they can immunize individuals against subsequent misleading information, for example, as seen in social media posts. However, the conditions under which debunking and prebunking effects persist over time need to be further examined. Protective effects may be limited to beliefs rather than behavioral intentions. Despite the limitations of the present studies, we recommend actively providing text-based corrections of misinformation to the public because they can be an effective and economic measure of science communication during a pandemic. However, designers of debunking interventions should be aware of how such texts should be constructed (Ecker et al., 2020; Lewandowsky et al., 2012, 2020) and that the effect may be limited or backfire. As demonstrated, debunking that introduces a new technology, such as mRNA, may backfire among religious groups. To reduce the risk of backfire, we speculate that it can be helpful to highlight that leaders of world religions have explicitly expressed their support for the COVID-19 vaccination. Moreover, debunking texts that

omit information about the side effects of vaccination may introduce new pro-vaccination myths that interfere with informed decision-making. Therefore, the risks of vaccination should be transparently reported to avoid overgeneralization and foster informed decision-making.

Acknowledgment

The authors thank Dr. Parichehr Shamsrizi and Prof. Dr. Marylyn Addo for their help in designing the debunking texts.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 964728 (JITSUVAX) during completion of this paper. The study was conducted as part of Germany's COVID-19 Snapshot Monitoring (COSMO), a joint project of the University of Erfurt (Cornelia Betsch [PI], Lars Korn, Philipp Sprengholz, Philipp Schmid, Lisa Felgendreff, Sarah Eitze), the Robert Koch Institute (RKI; Lothar H. Wieler, Patrick Schmich), the Federal Center for Health Education (BZgA; Heidrun Thaiss, Freia De Bock), the Leibniz Institute of Psychology (ZPID; Michael Bosnjak), the Science Media Center (SMC; Volker Stollorz), the Bernhard Nocht Institute for Tropical Medicine (BNITM; Michael Ramharter), and the Yale Institute for Global Health (Saad Omer).

ORCID iDs

Philipp Schmid  <https://orcid.org/0000-0003-2966-0806>

Cornelia Betsch  <https://orcid.org/0000-0002-2856-7303>

Supplemental Material

Supplemental material for this article is available online at <http://journals.sagepub.com/doi/suppl/10.1177/1075547020951794>.

References

- Abbasi, J. (2020). COVID-19 and mRNA vaccines—First large test for a new approach. *JAMA—Journal of the American Medical Association*, 324(12), 1125–1127. <https://doi.org/10.1001/jama.2020.16866>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)

- Allum, N., Sibley, E., Sturgis, P., & Stoneman, P. (2014). Religious beliefs, knowledge about science and attitudes towards medical genetics. *Public Understanding of Science*, 23(7), 833–849. <https://doi.org/10.1177/0963662513492485>
- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7(6), 543–554. <https://doi.org/10.1177/1745691612459060>
- Betsch, C., Rossmann, C., Pletz, M. W., Vollmar, H. C., Freytag, A., Wichmann, O., Hanke, R., Hanke, W., Heinemeier, D., Schmid, P., Eitze, S., Weber, W., Reinhardt, A., K upke, N. K., Forstner, C., Fleischmann-Struzek, C., Mikolajetz, A., R omhild, J., Neufeind, J., . . . Reinhart, K. (2018). Increasing influenza and pneumococcal vaccine uptake in the elderly: Study protocol for the multi-methods prospective intervention study Vaccination60+. *BMC Public Health*, 18(1), 885. <https://doi.org/10.1186/s12889-018-5787-9>
- Betsch, C., Schmid, P., Heinemeier, D., Korn, L., Holtmann, C., & B ohm, R. (2018). Beyond confidence: Development of a measure assessing the 5C psychological antecedents of vaccination. *PLOS ONE*, 13(12), Article e0208601. <https://doi.org/10.1371/journal.pone.0208601>
- Bolsen, T., & Druckman, J. N. (2015). Counteracting the politicization of science. *Journal of Communication*, 65(5), 745–769. <https://doi.org/10.1111/jcom.12171>
- Bolsen, T., & Druckman, J. N. (2018). Do partisanship and politicization undermine the impact of a scientific consensus message about climate change? *Group Processes and Intergroup Relations*, 21(3), 389–402. <https://doi.org/10.1177/1368430217737855>
- British Islamic Medical Association. (2020). *Position statement on the Pfizer/BioNTech Covid-19 vaccine*. <https://britishima.org/pfizer-biontech-covid19-vaccine/>
- Brossard, D., Scheufele, D. A., Kim, E., & Lewenstein, B. V. (2009). Religiosity as a perceptual filter: Examining processes of opinion formation about nanotechnology. *Public Understanding of Science*, 18(5), 546–558. <https://doi.org/10.1177/0963662507087304>
- Card, L. F. (2020). *Note of the congregation for the doctrine of the faith on the morality of using some anti-Covid-19 vaccines*. https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_con_cfaith_doc_20201221_nota-vaccini-anticovid_en.html
- Centers for Disease Control and Prevention. (2021). *Science Brief: COVID-19 Vaccines and Vaccination*. <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html>
- Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarrac ın, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546. <https://doi.org/10.1177/0956797617714579>
- Cohen, A. B., Malka, A., Rozin, P., & Chermas, L. (2006). Religion and unforgivable offenses. *Journal of Personality*, 74(1), 85–118. <https://doi.org/10.1111/j.1467-6494.2005.00370.x>

- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research, 34*(3), 315–346. https://doi.org/10.1207/S15327906MBR3403_2
- Compton, J., & Ivanov, B. (2012). Untangling threat during inoculation-conferred resistance to influence. *Communication Reports, 25*(1), 1–13. <https://doi.org/10.1080/08934215.2012.661018>
- Dragojlovic, N., & Einsiedel, E. (2013). Playing God or just unnatural? Religious beliefs and approval of synthetic biology. *Public Understanding of Science, 22*(7), 869–885. <https://doi.org/10.1177/0963662512445011>
- Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition, 6*(2), 185–192. <https://doi.org/10.1016/j.jarmac.2017.01.014>
- Ecker, U. K. H., Lewandowsky, S., 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., Jayawardana, K., & Mladenovic, A. (2019). Refutations of equivocal claims: No evidence for an ironic effect of counterargument number. *Journal of Applied Research in Memory and Cognition, 8*(1), 98–107. <https://doi.org/10.1016/j.jarmac.2018.07.005>
- Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology, 111*(1), 36–54. <https://doi.org/10.1111/bjop.12383>
- Farrell, J., McConnell, K., & Brulle, R. (2019). Evidence-based strategies to combat scientific misinformation. *Nature Climate Change, 9*(3), 191–195. <https://doi.org/10.1038/s41558-018-0368-6>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Habersaat, K. B., Betsch, C., Danchin, M., Sunstein, C. R., Böhm, R., Falk, A., Brewer, N. T., Omer, S. B., Scherzer, M., Sah, S., Fischer, E. F., Scheel, A. E., Fancourt, D., Kitayama, S., Dubé, E., Leask, J., Dutta, M., MacDonald, N. E., Temkina, A., . . . Butler, R. (2020). Ten considerations for effectively managing the COVID-19 transition. *Nature Human Behaviour, 4*(7), 677–687. <https://doi.org/10.1038/s41562-020-0906-x>
- Haglin, K. (2017). The limitations of the backfire effect. *Research and Politics, 4*(3), 2053168017716547. <https://doi.org/10.1177/2053168017716547>
- Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis* (1st ed.). Guilford Press. <https://doi.org/978-1-60918-230-4>
- Hotez, P., Batista, C., Ergonul, O., Figueroa, J. P., Gilbert, S., Gursel, M., Hassanain, M., Kang, G., Kim, J. H., Lall, B., Larson, H., Naniche, D., Sheahan, T., Shoham, S., Wilder-Smith, A., Strub-Wourgaft, N., Yadav, P., & Bottazzi, M. E. (2021). Correcting COVID-19 vaccine misinformation: Lancet

- Commission on COVID-19 Vaccines and Therapeutics Task Force Members*. *EClinicalMedicine*, 33, 100780. <https://doi.org/10.1016/j.eclinm.2021.100780>
- Jackson, N. A. C., Kester, K. E., Casimiro, D., Gurunathan, S., & DeRosa, F. (2020). The promise of mRNA vaccines: A biotech and industrial perspective. *NPJ Vaccines*, 5(1), 11. <https://doi.org/10.1038/s41541-020-0159-8>
- Jewish Medical Doctors working in UK. (2020). *Letter to Jewish press RE Pfizer-BioNTech mRNA Covid vaccine*. <https://docs.google.com/document/d/1ldvpDSrzJommeevkvj6ynFaEG0FkIdlSGfRIFz9Smm8/edit>
- Kim, H., & Walker, D. (2020). Leveraging volunteer fact checking to identify misinformation about COVID-19 in social media. *Harvard Kennedy School Misinformation Review*, 1(3). <https://doi.org/10.37016/mr-2020-021>
- Klumparendt, A., & Drenckhan, I. (2014). Deutsche Version der Religious Commitment Scale. *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen*. ZIS, 8542–8550. <https://zis.gesis.org/skala/Klumparendt-Drenckhan-Deutsche-Version-der-Religious-Commitment-Scale>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., Kimball, S., & El-Mohandes, A. (2021). A global survey of potential acceptance of a COVID-19 vaccine. *Nature Medicine*, 27(2), 225–228. <https://doi.org/10.1038/s41591-020-1124-9>
- Lewandowsky, S., Cook, J., Ecker, U., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., . . . Zaragoza, M. S. (2020). *Debunking handbook 2020*. <https://doi.org/10.17910/b7.1182>
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, Supplement*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 32(2), 348–384. <https://doi.org/10.1080/10463283.2021.1876983>
- Löffler, P. (2021). Review: Vaccine myth-buster—Cleaning up with prejudices and dangerous misinformation. *Frontiers in Immunology*, 12, 663280. <https://doi.org/10.3389/fimmu.2021.663280>
- 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 UK and USA. *Nature Human Behaviour*, 5(3), 337–348. <https://doi.org/10.1038/s41562-021-01056-1>
- MacPherson, Y. (2020). What is the world doing about COVID-19 vaccine acceptance? *Journal of Health Communication*, 25(10), 757–760. <https://doi.org/10.1080/10810730.2020.1868628>

- Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1–16. <https://doi.org/10.1037/xap0000315>
- McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *The Journal of Abnormal and Social Psychology*, 62(2), 327–337. <https://doi.org/10.1037/h0042026>
- Nisbet, M. C. (2005). The competition for worldviews: Values, information, and public support for stem cell research. *International Journal of Public Opinion Research*, 17(1), 90–112. <https://doi.org/10.1093/ijpor/edh058>
- Pardi, N., Hogan, M. J., Porter, F. W., & Weissman, D. (2018). mRNA vaccines—A new era in vaccinology. *Nature Reviews Drug Discovery*, 17(4), 261–279. <https://doi.org/10.1038/nrd.2017.243>
- Robert Koch-Institut. (2020). *Wirkweise und potentielle Risiken der mRNA-Impfstoffe gegen COVID19*. <https://www.youtube.com/watch?v=0LnkoEOHSiM>
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1), 65. <https://doi.org/10.1057/s41599-019-0279-9>
- Rutjens, B. T., Sengupta, N., der Lee, R. van, van Koningsbruggen, G. M., Martens, J. P., Rabelo, A., & Sutton, R. M. (2022). Science skepticism across 24 countries. *Social Psychological and Personality Science*, 13(1), 102–117. <https://doi.org/10.1177/19485506211001329>
- Rutjens, B. T., & van der Lee, R. (2020). Spiritual skepticism? Heterogeneous science skepticism in the Netherlands. *Public Understanding of Science*, 29(3), 335–352. <https://doi.org/10.1177/0963662520908534>
- Schmid, P., & Betsch, C. (2019). Effective strategies for rebutting science denialism in public discussions. *Nature Human Behaviour*, 3(9), 931–939. <https://doi.org/10.1038/s41562-019-0632-4>
- Schmid, P., & Betsch, C. (2022). *Data and material: Benefits and pitfalls of debunking interventions to counter mRNA vaccination misinformation during the COVID-19 pandemic*. Open Science Framework. <https://doi.org/10.17605/OSF.IO/DK52C>
- Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition*, 9(3), 286–299. <https://doi.org/10.1016/j.jarmac.2020.06.006>
- Swire-Thompson, B., Miklaucic, N., Wihbey, J. P., Lazer, D., & DeGutis, J. (2022). The backfire effect after correcting misinformation is strongly associated with reliability. *Journal of Experimental Psychology: General*, 151(7), 1655–1665. <https://doi.org/10.1037/xge0001131>
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>
- Tay, L. Q., Hurlstone, M. J., Kurz, T., & Ecker, U. K. H. (2022). A comparison of prebunking and debunking interventions for implied versus explicit

- misinformation. *British Journal of Psychology*, *113*(3), 591–607. <https://doi.org/10.1111/bjop.12551>
- Valente, T. W., & Pumpuang, P. (2007). Identifying opinion leaders to promote behavior change. *Health Education and Behavior*, *34*(6), 881–896. <https://doi.org/10.1177/1090198106297855>
- van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008. <https://doi.org/10.1002/gch2.201600008>
- van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. *Frontiers in Psychology*, *11*, 566790. <https://doi.org/10.3389/fpsyg.2020.566790>
- Vraga, E. K., & Bode, L. (2021). Addressing COVID-19 misinformation on social media preemptively and responsively. *Emerging Infectious Diseases*, *27*(2), 396–403. <https://doi.org/10.3201/EID2702.203139>
- Vraga, E. K., Kim, S. C., Cook, J., & Bode, L. (2020). Testing the effectiveness of correction placement and type on Instagram. *International Journal of Press/Politics*, *25*(4), 632–652. <https://doi.org/10.1177/1940161220919082>
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. <https://doi.org/10.1177/1745691612463078>
- Walter, N., Brooks, J. J., Saucier, C. J., & Suresh, S. (2021). Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. *Health Communication*, *36*(13), 1776–1784. <https://doi.org/10.1080/10410236.2020.1794553>
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350–375. <https://doi.org/10.1080/10584609.2019.1668894>
- Wittenberg, C., & Berinsky, A. J. (2020). Misinformation and its correction. In J. A. Tucker & N. Persily (Eds.), *Social media and democracy: The state of the field, prospects for reform* (pp. 163–198). Cambridge University Press.
- Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135–163. <https://doi.org/10.1007/s11109-018-9443-y>
- World Health Organization. (2021). *Coronavirus disease (COVID-19) advice for the public: Mythbusters*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>

Author Biographies

Philipp Schmid is psychologist and postdoctoral researcher at the University of Erfurt, Germany. He studies the psychology of science denialism and health misinformation and aims to support people's informed decision-making in health, for example, vaccination. He applies a persuasion psychology perspective to understand the impact of misinformation in health communication and to develop and evaluate promising interventions.

Cornelia Betsch is psychologist and professor of health communication at the University of Erfurt and leads the Health Communication working group at the Bernhard-Nocht-Institute for Tropical Medicine in Hamburg. She studies the psychological mechanisms of health behaviors and aims at increasing planetary health by evidence-based health communication. She applies a judgment and decision-making and strategic interaction perspective to understanding health behavior—especially with regard to fighting emerging infectious diseases, vaccination decision-making, prudent use of antibiotics, and climate friendly behavior.