Journal of Experimental Social Psychology 121 (2025) 104806

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Does protection come at a cost? A random stimuli approach to investigating the (side-)effects of misinformation inoculations☆

Teodora Spiridonova [a,*], Olga Stavrova [b], Ilja van Beest [a]

[a] Department of Social Psychology, School of Social and Behavioral Sciences, Tilburg University, Warandelaan 2, 5037 AB, Tilburg, Netherlands
[b] Department of Psychology (IPSY I), Lübeck University, Maria-Goeppert-Str. 9a, 23562 Lübeck, Germany

ARTICLE INFO

Keywords:
Misinformation
Inoculation
Fake news
Intervention
Cynicism

ABSTRACT

Inoculation – an intervention aimed at informing people of the threat of misinformation and the strategies used to spread it – is an increasingly popular approach for fighting fake news. While studies have shown inoculation to be effective in reducing the credibility of fake news, the evidence on whether it might also lead to undesirable side-effects, such as reduced credibility of true news, is mixed. Further, existing research has only rarely tested inoculation using real-life news, has not accounted for the potential issue of biased stimulus selection, and has not tested the assumed mechanism behind the inoculation's effectiveness: the higher presence of misinformation strategies in fake vs. true news. The present research was designed to fill these gaps. Using a random stimuli approach and a dataset of real-life true and fake news headlines, Study 1 showed that inoculation decreased perceived accuracy (but not trustworthiness) of fake news (without changing the perceived accuracy of true news), and did not render people more cynical. Additionally, Study 2 showed that fake news contained more misinformation strategies than true news, and Study 3 found that the inoculation worked better on headlines that used more (vs. fewer) misinformation strategies. In sum, our findings suggest that inoculation is unlikely to have side effects, yet its effectiveness might be more limited than previously assumed. We thus contribute to the broader literature on reducing misinformation, and research on the effectiveness of the inoculation approach in particular.

1. Introduction

The rise of misinformation has been described as one of the key challenges of the 21st century, carrying with it numerous consequences – from endangering public health (Larson et al., 2011; Poland & Spier, 2010; Van der Linden et al., 2021), to increasing political polarization and violence (Cook et al., 2017; Mozur, 2018), to eroding people's trust in media, government, and other important institutions (Iyengar & Massey, 2019). Over the past decade, researchers have proposed a possible solution to this "infodemic": inoculation interventions, which consist of exposing people to "weakened doses" of misinformation by making them familiar with the strategies used to spread misinformation (Compton, 2013; McGuire, 1964; Van der Linden et al., 2017).

While early research on inoculation showed promising results – demonstrating that it could reduce belief in misinformation across several domains, such as climate change (Van der Linden et al., 2017), vaccines (Jolley & Douglas, 2017), and GMOs (Wood, 2007), more recent studies have found mixed evidence for its effectiveness (Modirrousta-Galian & Higham, 2023). In addition, scholars have raised concerns that inoculation might have unintended side-effects, such as reducing the credibility not only of false but also of true information (e.g., Hoes et al., 2024). However, the evidence for such side-effects has been far from consistent, with some studies showing inoculation to be safe (Basol et al., 2021, Study 1; Harrop et al., 2023, Study 1; Lu et al., 2023), and other studies finding side-effects (Basol et al., 2021, Study 2; Guess et al., 2020; Hameleers, 2023; Harrop et al., 2023, Study 3; Hoes et al., 2024; Maertens et al., 2024; Modirrousta-Galian & Higham, 2023; Van der Meer et al., 2023).

Notably, few of these studies 1) used real-life news to test the effectiveness of inoculation, and 2) none accounted for the risk of biased stimulus selection when examining the impact of inoculation on the perception of specific news items. The goal of the present research was

☆ This paper has been recommended for acceptance by "Chadly Stern".

* Corresponding author.
E-mail addresses: T.Spiridonova@tilburguniversity.edu (T. Spiridonova), olga.stavrova@uni-luebeck.de (O. Stavrova), I.vanBeest@tilburguniversity.edu (I. van Beest).

https://doi.org/10.1016/j.jesp.2025.104806
Received 19 November 2024; Received in revised form 22 July 2025; Accepted 23 July 2025
Available online 11 August 2025

therefore to provide a more ecologically valid test of inoculation's intended effects and unintended side-effects, by sampling from a large-scale dataset of real-life news containing both "fake" and "true" headlines, and utilizing a random stimuli approach to account for methodological bias in stimulus selection (Judd et al., 2012). In addition, we went beyond assessing the effect of the inoculation intervention on the perceived credibility of specific news headlines, and additionally tested whether it would "spill over" and contaminate individuals' worldview by increasing their general cynicism (i.e., the belief that most people are driven by self-serving motives, Stavrova and Ehlebracht (2016)), and their cynicism towards the media (i.e., the belief that the media are driven by self-serving motives, Markov and Min (2022)) specifically.

Finally, inoculation interventions are based on the assumption that fake news contain specific misinformation "strategies" that true news do not and that drawing individuals' attention to these cues inoculates them against misinformation. Consequently, inoculation can only be effective if, within the real media environment, "fake news" use more misinformation strategies than "true news". Yet, to the best of our knowledge, this assumption has never been put to an empirical test. Here, using a larger dataset of real-life news headlines, we explore whether fake and true news indeed differ in the amount of misinformation strategies they include. Thus, the present research offers several contributions: It tests 1) the effectiveness and 2) the potential side-effects of the inoculation intervention using the state-of-the-art random stimuli approach (Judd et al., 2012), 3) It extends the scope of side-effects examined in previous research by testing a potential spill-over on media cynicism and general cynicism, and 4) It tests the very assumption that the success of the inoculation is based on.

1.1. Testing the (side-)effects of inoculation

Inoculation, also termed "prebunking", is a preventative approach to reducing the spread of, and people's belief in, misinformation. Influenced by early research on persuasion, inoculation theory is grounded in the idea that being exposed to weakened versions of certain persuasion tactics used to spread misinformation will lead to the development of "antibodies" that are activated when encountering real misinformation (Compton, 2013; McGuire, 1964). In other words, being aware of the threat of misinformation and the strategies that are used to spread it can help protect individuals from believing the kind of misinformation they encounter in their daily lives. Early research on inoculation posited that this approach was not only very effective, but also that it was "immune to side-effects". Indeed, many of the first studies that tested inoculation across different domains found that it both successfully reduced people's belief in misinformation, and increased discernment between true and fake news items – such that fake news were believed less, while true news were believed more, or to the same extent, compared to the control condition (Maertens et al., 2021; Roozenbeek et al., 2020; Roozenbeek et al., 2021; Roozenbeek & van der Linden, 2019). Further support for this notion came from a 2023 meta-analysis, which showed that on average, inoculation successfully increased discernment between misinformation and true information, and even increased the perceived credibility of true information (across 12 and 26 studies, respectively) (Lu et al., 2023).[1] Notably, however, both effects were subject to substantial heterogeneity across the studies, with standard mean differences ranging from very small to moderate (in the case of discernment), and even ranging from positive to negative (in the case of credibility assessments of true information).

In addition, more recent findings – which were not included in the aforementioned meta-analysis – have put into question both the effectiveness of inoculations in general, as well as the notion of "side-effect immunity". First, some scholars found evidence that inoculation lowered the credibility of not just false, but also true information (Guess et al., 2020; Harrop et al., 2023, Study 3; Maertens et al., 2024). Second, a re-analysis of inoculation studies published between 2019 and 2021 (Modirrousta-Galian & Higham, 2023) showed that only two out of five studies offered convincing evidence that discernment between true and false information was improved following the inoculation intervention, and argued that this result could be attributed to differences in ambiguity between the (true and fake) news items participants were presented with. Conversely, in three out of the five studies, the inoculation resulted in increased skepticism of all news, regardless of their veracity (Modirrousta-Galian & Higham, 2023). Similar findings were reported by Hoes et al. (2024), who compared the effectiveness of three misinformation interventions, including inoculation (here termed "media literacy intervention"), across different countries, and found that all three either increased skepticism of all news, or were not effective at all (Hoes et al., 2024). Finally, two studies that focused specifically on the threat component of inoculation (i.e., the initial warning that misinformation is something people should be wary of in daily life) found that it triggered a deception bias (i.e., a tendency to suspect all information of being deceptive), thus negatively affecting the credibility of real news (Hameleers, 2023; Van der Meer et al., 2023).

1.2. The problem of biased stimulus selection

Importantly, what a majority of previous studies did not take into account is how the ability to detect both inoculation effects and side-effects might be affected by the selection of (news) items used to test the success of the inoculation intervention. Specifically, many early inoculation studies tended to use fewer true news items than fake news items (e.g. 18 vs. 3), resulting in reduced reliability of the true (vs. fake) news items (see e. g. Maertens et al., 2020; Roozenbeek et al., 2021). Lower reliability implies a higher measurement error, which naturally limits a measure's ability to correlate with any other measures, including being affected by an experimental manipulation (Spearman, 1904). As a result, any difference in the effect of the inoculation on fake and true news might be the result of a methodological artefact (i.e., lower reliability of credibility judgments of true vs. fake news), rather than a true difference (see e. g. Meier & Perrig, 2000).

In addition, many inoculation studies use fabricated items for both true and fake news, with the main difference between the two being that the fake news items are intentionally designed to contain the misinformation strategies that the intervention targets, while the true news items are designed not to contain such strategies (e.g., Roozenbeek et al., 2021; Roozenbeek & van der Linden, 2019). As a consequence, it is plausible that the inoculation did not affect the credibility judgment of the true news items in these studies just because they were constructed without using the misinformation strategies targeted by the inoculation manipulation. Yet, whether true news are actually free of misinformation strategies "in real life" remains an open question. Overall, 10 of the 12 studies included in the aforementioned meta-analysis used artificial news items,[2] while four studies also used an unequal number of true vs. fake news items (Lu et al., 2023).

Finally, although some research in this field has tried to go beyond artificial news items (Basol et al., 2021; Roozenbeek et al., 2022), or tested inoculation on both artificial *and* real news items (Maertens et al., 2024), one additional issue remains. Namely, although all inoculation studies use a limited sample of news item stimuli, to the best of our knowledge, none of them account for the stimuli selection in their analyses. Typically, inoculation is tested on a number of true news and fake news items that are averaged to form mean credibility judgments of fake and true news, which are then compared to each other. This approach is

[1] Note that not all studies that measured the effect of inoculation on misinformation ($n = 31$) also measured its effect on true information ($n = 26$) and vice versa. Only the 12 studies mentioned here measured both at the same time.

[2] One study tested the inoculation using the Misinformation Susceptibility Test, which consists of both real and artificial news items.

problematic, as it essentially ignores the possibility of systematic variation between the stimuli. Such variation could result in a significant mean difference between the perceived credibility of fake and true news, which might not replicate in another study that uses different stimuli. Thus, the inconsistency of side-effects across the inoculation literature could be (partially) explained by such systematic stimulus variation. In addition, failing to account for stimulus variation can result in increased Type I error rates (Judd et al., 2012), such that inoculation appears more effective than it actually is.

In order to tackle these issues, and provide a thorough investigation of both the effectiveness and potential side-effects of inoculation, the present research thus followed a more ecologically valid and analytically robust approach. First, we tested the inoculation on real-life true and fake news items, randomly selected from a large-scale dataset of true and fake news headlines (McIntire, 2017). This allowed us to examine whether inoculation holds up in a more realistic context and whether its effect generalizes beyond artificially constructed stimuli. Second, we balanced the number of true and fake news items, to avoid differences in their reliabilities (Meier & Perrig, 2000). Third, we accounted for systematic stimulus variation by employing a mixed-effects model that treats both participants and stimuli as simultaneous random effects, thus reducing the risk of Type I errors (Judd et al., 2012). Using this improved methodology, we tested whether being exposed to an inoculation intervention would lower the credibility – as assessed in terms of perceived accuracy (H1a), and trustworthiness (H1b) – of both true and fake news headlines.

1.3. From reduced credibility to cynicism

While the possible side-effects of inoculation are gaining increased research attention (e.g., Hameleers, 2023; Van der Meer et al., 2023), the focus has largely remained on exploring the risk it might pose for the perceived credibility of true news. However, the threat component of inoculation interventions might not only undermine the perceived credibility of true (as well as fake) news, but also individuals' generalized beliefs about the media, in the form of media cynicism. Media cynicism is defined as "a generalized antagonism toward mainstream news media", characterized by negative expectations about media actors' journalistic integrity and the perception that media actors are driven by self-serving motives (such as increasing user engagement at any cost or promoting a certain political agenda) (Markov & Min, 2022). People who are cynical towards the media reject the notion that its goal is to serve the public, instead attributing purely malicious and manipulative intentions to media actors.

Because inoculation is predicated on warning individuals about the possibility that certain media will try to manipulate them, it stands to reason that this type of intervention might also increase individuals' cynicism towards the media. Indeed, previous research has shown that merely being exposed to discussions about fake news can lower people's trust in the media (Van Duyn & Collier, 2019). As within the inoculation intervention, individuals are not only confronted with the existence of misinformation, but are also explicitly warned about the different strategies used by media actors to spread it, stronger media cynicism might represent another – previously untested – side-effect of the inoculation treatment.

In addition to increasing people's media cynicism, there is reason to believe that inoculation might affect people's general cynicism as well. Cynicism is defined as "the belief that other people are driven primarily by self-interest" (Stavrova & Ehlebracht, 2016, 2019). Even though cynicism represents a moderately stable dimension of individual differences, it has been shown to increase after negative social experiences (Stavrova et al., 2020) and in situations that increase one's dependence on others (Stavrova & Ehlebracht, 2019). More importantly, cynicism has been shown to be particularly sensitive to the experience of threat (Stavrova et al., 2023), rendering increased cynicism another potential side-effect of the inoculation treatment. As such, our third goal was to investigate the potential consequences of inoculation for both general cynicism, and one of its domain-specific forms that is particularly relevant in the context of misinformation: media cynicism. Based on the notion that inoculation messages point to a clear vested interest that can be ascribed to both the producers of misinformation (i. e, the news media) and the consumers and spreaders of misinformation (i. e., people in general), we hypothesized that being exposed to an inoculation intervention would increase people's cynicism towards the news media (H2a), as well as their general cynicism (H2b).

1.4. Testing the underlying assumptions of inoculation theory

Beyond potentially inflating the effectiveness of inoculation interventions, the problem of biased stimulus selection we laid out previously also carries another, often overlooked, implication. Because inoculation research has largely opted to use artificial news items, the underlying assumption of inoculation paradigms – that fake news use "misinformation strategies" (e.g., Polarization, Emotional Language, Conspiracy Theories), while true news do not – has remained untested. Indeed, some scholars have raised the possibility that "some of the manipulative techniques the [inoculation] games seek to teach participants about are present in both fake news and true news", especially given a media climate that is focused on generating as many clicks as possible in the ever-shifting attention economy (Modirrousta-Galian & Higham, 2023, p. 66).

Despite these concerns, however, this possibility has not been formally tested in existing inoculation research. The second goal of this research was therefore to address this gap, by investigating whether fake news headlines indeed contain more misinformation strategies, and whether the inoculation would be more effective for headlines with a higher (vs. lower) misinformation strategy use. Specifically, we hypothesized that fake news headlines would be rated as using more misinformation strategies than true news headlines (H3) and that inoculation would be more effective for headlines that use misinformation strategies compared to headlines that do not (H4).

1.5. Overview of studies

We tested our hypotheses across three online studies. Study 1 was aimed at 1) providing a thorough investigation of the intended effects and unintended side-effects of inoculation (i.e., testing whether being inoculated would lower the perceived accuracy (H1a) and trustworthiness (H1b) of both true and fake news) using a random stimuli approach, 2) testing whether receiving an inoculation treatment would make people more cynical towards the news media (H2a), as well as people in general (H2b). Participants were first presented with either a text-based inoculation intervention, which described five misinformation strategies that they should look out for (Discrediting Opponents, Emotional Language, Polarization, Conspiracy Theories, and Clickbait), or a control text. They were then asked to rate the perceived accuracy and trustworthiness of 120 (60 true and 60 fake) news headlines randomly selected from a large-scale dataset of real-life headlines (McIntire, 2017), with each participant rating a random block of ten (five true and five fake) headlines. Subsequently, they completed measures of news media cynicism and general cynicism.

In Study 2, we instead had a new sample of participants rate the same headlines on the perceived presence of the five misinformation strategies targeted by the inoculation manipulation in Study 1 (Discrediting Opponents, Emotional Language, Polarization, Conspiracy Theories, and Clickbait). We then compared their ratings between the true and fake headlines, in order to test whether fake headlines contained more misinformation strategies than true headlines (H3). Finally, in Study 3, we combined the data from Study 1 and Study 2 to test whether the inoculation was more successful when applied to headlines that use more misinformation strategies, compared to those that use fewer misinformation strategies (H4). We report all manipulations, measures,

and exclusions in these studies.

The materials and data for all three studies can be accessed at: https://osf.io/mh3vj/?view_only=a7da587c3b39499ab58d84419f63dc88. All three studies, including hypotheses, procedures, and analysis plans, were pre-registered (see https://aspredicted.org/fsky-d3s4.pdf and https://aspredicted.org/ndb7-wcfs.pdf[3]). We did not deviate from these pre-registrations.

2. Study 1

2.1. Method

2.1.1. Inoculation treatment

Inoculation generally consists of two components: 1) a warning about encountering misinformation and 2) exposure to the strategies or arguments commonly used in misinformation campaigns. The inoculation literature contains different operationalizations of these components, with some taking the form of a text presenting in-depth counterarguments to common misinformation talking points in a particular domain (e.g., vaccine hesitancy, Vivion et al., 2022), and others using a gamified approach, where participants learn about general misinformation strategies, and can create examples of misinformation themselves (see Basol et al., 2020, Roozenbeek & van der Linden, 2019, 2020). In the present study, we employed a combination of these approaches, by using a text-based inoculation (see Cook et al., 2017; Van der Linden et al., 2017; Vivion et al., 2022), which warned participants about encountering misinformation and provided them with a list of strategies commonly used to spread misinformation that they should look out for. For the latter, we selected the most relevant strategies from the DEPICT framework (Basol et al., 2020; Maertens et al., 2020; Roozenbeek & van der Linden, 2019) – Discrediting Opponents, Emotional Language, Polarization, and Conspiracy Theories,[4] and included a fifth strategy that we considered particularly relevant in the context of news headlines – Clickbait. We selected this particular combination because it 1) includes the basic tenets of inoculation theory (a warning about misinformation and examples of the forms it can take), 2) is not limited to a particular domain, but applies to all types of news, 3) uses an established framework of misinformation strategies (DEPICT), 4) resembles news literacy interventions commonly employed in real-life contexts (where participants often lack the time to play a game or read longer texts with detailed arguments; see e.g., Facebook's "Tips to spot fake news", or "10 tips from WhatsApp to spot fake news"), and thus contributes to the external validity of our studies, and 5) is easily implemented, and therefore potentially easily scalable.

In Study 1, participants were randomly assigned to either an *inoculation* condition or a *control* condition. Participants in the inoculation condition received the inoculation text, while participants in the control condition read a text of similar length about the history of sushi (see Amazeen, 2021; Banas & Miller, 2013; Bessarabova & Banas, 2024). The complete inoculation and control texts can be found in the Supplementary Materials.

2.1.2. Headline selection

In order to test the inoculation on real-life news headlines, we sampled from a large-scale dataset ($N = 6{,}335$) of headlines, published during the 2016 U.S. election cycle, and classified as either "TRUE" or "FAKE" (McIntire, 2017). This dataset was originally created by merging two datasets comprised of true and fake news, respectively. The true news were webscraped from All Sides, a website hosting news articles from across the political spectrum, while the fake news were selected from an existing dataset of 13,000 articles, scraped from 244 websites and previously tagged as "bullshit" by the BS Detector Chrome Extension (Sieradski, 2016). More detailed information about the procedure can be found here. To achieve sufficient power, we aimed to select 120 (60 true and 60 fake) headlines from this dataset (Judd et al., 2017). To do this, we first shuffled the original order of headlines, separately for the true and fake news headlines. Next, we inspected the first 100 true and the first 100 fake headlines using seven exclusion criteria. These criteria were selected to ensure that headlines were clear, interpretable, and conveyed information that could be (correctly or not) evaluated in terms of its accuracy and trustworthiness, and that participants' familiarity with the events mentioned would not interfere with this evaluation. Specifically, headlines were excluded if they 1) were not phrased as clear statements (e.g., "Hope, Change, Lies and Greatness"); 2) were phrased as questions (e.g., "Statistical Propaganda: How many Syrians has US regime-change killed?"); 3) referred to an ongoing event (i.e., an event that was already concluded by the time of running the study, but phrased as ongoing in the headline), such as the 2016 Trump and Clinton campaigns (e.g., "Assange: Donald Trump Won't Be Allowed to Win – Clinton & ISIS Funded by Same Interests"); 4) were opinion pieces (e.g., "To finish MLK's work, face up to racism (Opinion)"); 5) did not constitute news (e.g., ads, comments, photos, videos); 6) were difficult to understand (i.e., the meaning was unclear or required very niche knowledge to grasp, such as "LUCIFER in the Temple of the Dog II"); 7) were deemed too outlandish (e.g., news about aliens).[5] We continued to sample in blocks of 100 using our exclusion criteria until we reached our desired sample size of $N = 60$ true headlines and $N = 60$ fake headlines. In total, 132 (out of 192 scanned) fake news and 126 (out of 186 scanned) true news headlines were excluded to reach this saturation point. A Chi-Square test comparing these ratios showed that they were not significantly different from one another ($\chi^2(1) = 0.04$, $p = .833$).

2.1.3. Sample size

To arrive at our desired sample size, we conducted a simulation-based power analysis using the R package simr (Green & MacLeod, 2016). We first determined that, to avoid attention attrition, each participant should only rate 10 of the 120 headline stimuli, and that each stimulus would therefore be presented 36 times (a number deemed sufficient to generate reliable mean inferences, see Rosenbusch et al., 2021). This resulted in a sample size of 432 participants. We simulated the dataset with these parameters. A total of 10,000 simulations of this data indicated that, based on an alpha level of .05, this sample size would allow us to detect a small effect of condition ($b = 0.10$), and a small interaction effect of condition × headline veracity ($b = 0.20$) with at least 80 % power. To account for participants failing the comprehension checks or completing the study in less than three minutes (i.e., our pre-registered exclusion criteria), we aimed to recruit a total of 450 participants. Note that our final sample size, $N = 388$ (see "Participants" section below) provided at least 80 % power to detect effects of $b = 0.19$

[3] We originally conceptualized Studies 2 and 3 as one study (this was later amended to improve the structure and readability of the manuscript). As a result, the hypotheses, procedures and analysis plans for both studies are contained within one pre-registration.

[4] The strategies Impersonation and Trolling were not considered relevant in this context, given that they are more common on social media than in the context of reading news headlines per se.

[5] Note that this procedure does not entirely eliminate the risk of stimulus selection bias, (i.e., it does not guarantee that the selected headlines are representative of the larger set of headlines they were selected from). Nevertheless, we believe that it helped to substantially reduce the stimuli selection bias present in previous inoculation studies by using a dataset comprised of real news headlines scraped from a variety of websites and including news from across the political spectrum (where previous inoculation studies frequently used datasets consisting entirely of artificial news items, i.e., news items that the researchers created specifically for the study, see e.g., Basol et al., 2020) and introducing (more) random sampling (where previous inoculation studies often used fixed datasets of 6–21 stimuli to test the inoculation, see e.g., Roozenbeek & van der Linden, 2019).

and above in our multilevel analyses, when accounting for an alpha level of .05.[6]

2.1.4. Procedure

Participants first read either the inoculation or the control text. Both texts were presented for at least 12 s, such that participants could only continue onto the next page after this set time had passed. This was followed by a comprehension check: In the inoculation condition, participants were asked "Which type of news should you be wary of?", with response options a) "News that focus on issues that pit two groups against each other" (correct), b) "News that make you doubt your existing beliefs and opinions." (incorrect), c) "News that use simple language." (incorrect). In the control condition, participants were given a similar comprehension question about the history of sushi text. Next, we presented participants with the 120 headlines, divided into twelve blocks. Each participant was randomly shown one of the twelve blocks, containing ten (five true and five fake) headlines, all presented in a random order. Participants then rated the perceived accuracy and trustworthiness of each headline. After completing the ratings, they filled out measures of media cynicism and general cynicism (presented in counterbalanced order). Finally, we asked participants to report their familiarity with the headlines, level of news literacy, and demographics. After completing the study, participants were debriefed about the goal of the study (including a detailed breakdown of which headlines were true and which were fake) and paid £0.90 for their participation.

2.1.5. Dependent measures

Perceived Headline Credibility. For each headline, participants were asked 1) "How accurate do you think the content of this headline is?" (*perceived headline accuracy*, 1 = not accurate at all; 7 = very accurate), 2) "How much do you trust the content of this headline?" (*perceived headline trustworthiness*, 1 = I do not trust it at all, 7 = I trust it completely).

Media Cynicism. We measured participants' cynicism towards the news media with the Media Cynicism scale (Markov & Min, 2022). This scale was developed specifically to measure media cynicism and consists of 10 items (e.g., "Journalists are prepared to lie to us whenever it suits their purposes.") rated on a 7-point scale (1 = strongly disagree, 7 = strongly agree). The scale showed excellent reliability in our sample (Cronbach's $\alpha = 0.89$; McDonald's $\omega = 0.92$).

General Cynicism. Participants' general cynicism was assessed with the Cynical Distrust scale (Greenglass & Julkunen, 1989), a common and well-validated measure of cynical beliefs (see e.g., Choy et al., 2021; Kaplan et al., 2004; Stavrova et al., 2020). The scale comprises eight items (e.g., "I think most people would lie to get ahead."), which were rated on a 7-point scale (1 = strongly disagree, 7 = strongly agree), and had very good reliability (Cronbach's $\alpha = 0.88$; McDonald's $\omega = 0.91$).

2.1.6. Control variables

Headline Familiarity. To account for the possibility that participants' familiarity with the headlines would impact our results, we included three headline familiarity items, namely: "How familiar were you with the names/people mentioned in the headlines you read earlier?", "How familiar were you with the locations mentioned in the headlines you read earlier?", and "How familiar were you with the topics of the headlines you read earlier?" (1 = very unfamiliar, 7 = very familiar).

News Literacy. We included a one-item measure of news interest ("How interested would you say you are in keeping up with the news?", 1 = not interested at all, 6 = very interested), and two items measuring news consumption: "How often do you check the news, on average?" (1 = Less than once a week, 5 = Several times a day) and "Which of the following do you use as a source of news (multiple answers possible)?", with the following options: TV, Radio, Print Newspapers, Online Newspapers, Social Media, Other [namely…].

Demographics. Finally, we asked participants to report their age, gender, education level, and political orientation.

2.1.7. Participants

Participants were recruited via Prolific in December 2022, resulting in an initial sample of $N = 465$ (inoculation condition: 232; control condition: 233). Of those participants, five were excluded because they completed the survey in less than three minutes (3 in the inoculation condition, 2 in the control condition), and 72 were excluded due to failing the comprehension check (47 in the inoculation condition (final $N = 182$), and 25 in the control condition (final $N = 206$)),[7] as per our pre-registered exclusion criteria. Thus, our final sample comprised a total of $N = 388$ participants (193 women, 188 men, 3 non-binary participants, and 4 participants who did not disclose their gender), with an average age of 28.81 ($SD = 10.13$). A majority of participants (63.5 %) had obtained at least a Bachelor's degree as their highest degree of education. In regards to their political orientation, participants were slightly left of center on average ($M = 3.49$, $SD = 1.36$). In terms of their news interest and consumption, participants were moderately interested in keeping up with the news ($M = 3.21$, $SD = 1.14$), and about half (53.1 %) checked the news at least once a day. They got their news primarily from social media (33.5 %) or online newspapers (29 %), followed by TV (22.7 %), radio (8.9 %), print newspapers (4.1 %), and other sources (1.8 %), such as YouTube, podcasts, etc. Finally, people's familiarity with the headlines was above or around the mid-point of the scale for the names and people mentioned ($M = 4.21$, $SD = 1.08$), the locations mentioned ($M = 3.71$, $SD = 1.24$), and the topics mentioned ($M = 3.45$, $SD = 1.11$) in the headlines.

2.2. Results

2.2.1. Correlations

An overview of correlations between the study variables can be found in the Supplementary Materials.

2.2.2. Perceived accuracy of headlines

Interaction. In order to test the effectiveness and potential side-effects of the inoculation on the perceived accuracy of headlines (H1a), we first compared participants' accuracy ratings for the true and fake news headlines across conditions using a multilevel regression,[8] with the following predictors: condition (1 = inoculation, 0 = control), headline veracity (1 = fake news, 0 = true news) and a condition x headline veracity interaction. We also included a random intercept of participants and headlines. Results showed a significant interaction effect between condition and headline veracity ($t(3401) = -2.40$, $b = -0.20$, $SE = 0.08$, CI[9] $[-0.37; -0.04]$, $p = .016$, $d = 0.08$). We probed this interaction using simple slopes, which showed that participants who were given the inoculation treatment could better differentiate the accuracy of fake and true news ($n = 182$, $t(169) = -5.84$, $b = -0.61$, $SE = 0.10$, CI $[-0.82; -0.41]$, $p < .001$) than participants in the control

[6] As calculated using a simulation-based sensitivity analysis conducted with the R package simr (Green & McLeod, 2016).

[7] Note that we compared the participants who were excluded based on the comprehension check item in each condition on our demographic and news literacy variables (age, education, political orientation, news consumption, and news interest), using independent samples *t*-tests, to ensure that this imbalance in exclusions would not result in a difference in competence between the two conditions. We found that the two groups did not differ significantly from one another on any of these variables (see Supplementary Materials for the full analyses).

[8] All multilevel analyses in these studies were conducted using the R package lme4 (Bates et al., 2015)

[9] Here and throughout the paper: 95 % CI.

condition (n = 206, $t(155) = -3.98$, $b = -0.41$, $SE = 0.10$, CI [−0.61; −0.21], $p < .001$) (see Fig. 1) – providing evidence that the inoculation worked as intended. Finally, additional simple slope analyses showed that the effect of the condition on perceived headline accuracy was significant for fake news ($t(646) = 3.72$, $b = -0.33$, $SE = 0.09$, CI [−0.50; −0.15], $p < .001$), but not true news ($t(646) = -1.40$, $b = -0.12$, $SE = 0.09$, CI [−0.05; 0.29], $p = .163$) – providing evidence against inoculation side-effects. The interaction effect remained significant when controlling for age, gender, familiarity with headlines (names, locations, and topics), news consumption, news interest, and political orientation (by including their interaction terms with condition into the model), as per our pre-registration (for the full results of these analyses, see Supplementary Materials).

2.2.3. Perceived trustworthiness of headlines

Interaction. Next, we tested the condition x headline veracity interaction on the perceived trustworthiness of headlines (H1b). There was no significant interaction effect between condition and headline veracity ($t(3400) = -1.11$, $b = -0.10$, $SE = 0.09$, CI [−0.27; 0.08], $p = .267$, $d = 0.04$), indicating that participants in the inoculation condition did not rate fake news as less trustworthy (compared to true news) than participants in the control condition (see Fig. 2). This effect remained consistent when controlling for age, gender, familiarity with headlines (names, locations, and topics), news consumption, news interest, and political orientation (by including their interaction terms with condition into the model). For the full results of these analyses, see Supplementary Materials.

2.2.4. Media cynicism

To test whether the inoculation treatment resulted in increased media cynicism, we ran a t-test comparing participants' media cynicism between the two conditions. We found no significant difference between conditions ($t(385) = -0.87$, $p = .386$, $d = 0.09$).[10] Participants exposed to the inoculation treatment (n = 182, $M = 4.68$, $SD = 0.96$) were not more cynical about the news media than participants in the control condition (n = 205, $M = 4.77$, $SD = 1.02$).

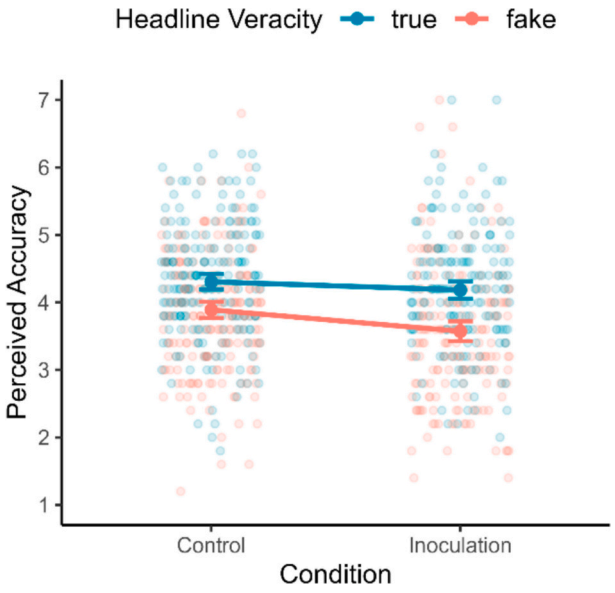

Fig. 1. Interaction effect between condition and headline veracity on perceived headline accuracy.
Note. Error bars indicate 95 % confidence intervals.

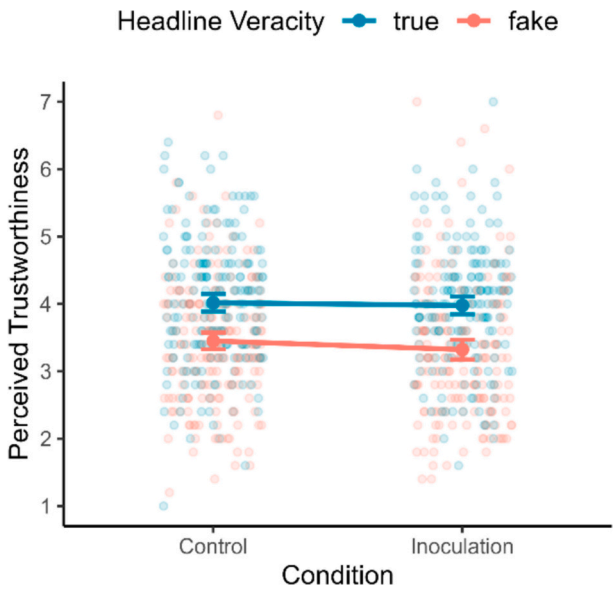

Fig. 2. Interaction effect between condition and headline veracity on perceived headline trustworthiness.
Note. Error bars indicate 95 % confidence intervals.

2.2.5. General cynicism

To test whether the inoculation treatment increased participants' general cynicism, we ran a t-test comparing participants' general cynicism between the two conditions. Again, there was no significant difference between conditions ($t(385) = -0.87$, $p = .383$, $d = 0.09$). Participants exposed to the inoculation treatment (n = 182, $M = 4.25$, $SD = 1.15$) were not more cynical than participants in the control condition (n = 205, $M = 4.35$, $SD = 1.17$).

2.2.6. Exploratory analyses

Main effects of condition and headline veracity. While our main confirmatory analyses were focused on the interaction effect between condition and headline veracity, we also explored the main effects of these variables on our dependent variables separately. Using multilevel regressions with condition as the predictor, and perceived accuracy and trustworthiness as the respective dependent variables, along with a random intercept of participants and headlines, we found that condition (1 = inoculation, 0 = control) negatively predicted perceived headline accuracy ($t(382) = -2.93$, $b = -0.22$, $SE = 0.08$, CI [−0.37; −0.07], $p = .004$, $d = 0.30$), but not perceived headline trustworthiness ($t(381) = -1.19$, $b = -0.09$, $SE = 0.08$, CI [−0.25; 0.06], $p = .234$, $d = 0.12$). Running the same analysis with headline veracity (1 = fake headline, 0 = true headline) as the sole predictor, we found that it negatively predicted both perceived headline accuracy ($t(112) = -5.33$, $b = -0.50$, $SE = 0.09$, CI [−0.69; −0.32], $p < .001$, $d = 1.00$) and perceived headline trustworthiness ($t(113) = -5.87$, $b = -0.60$, $SE = 0.08$, CI [−0.80; −0.40], $p < .001$, $d = 1.11$).

Including random slopes. To further test the robustness of our findings, we re-ran our main analyses and included 1) a random slope of condition across headlines, 2) a random slope of headline veracity across participants, and 3) both of these random slopes in the same model. This allowed us to account for additional heterogeneity in the effect of the condition across headlines, as well as in the rating of true vs. fake news across participants (see Judd et al., 2012). The interaction effect between the condition and headline veracity on perceived headline accuracy remained significant across all three models, while the interaction effects on perceived headline trustworthiness remained non-significant.

[10] Note that the sample size for these analyses was smaller than that of the total sample (N = 387), as one participant did not complete the cynicism measures.

The full results of these analyses can be found in the Supplementary Materials.

Bayesian analyses. Despite the fact that the effect of condition on the perceived accuracy of true news headlines did not reach significance, the estimated coefficient of this effect followed the same general pattern as the condition effect on fake news headlines ($b = -0.12$, $p = .163$). Thus, as an additional robustness check of our findings regarding the lack of inoculation side-effects, we conducted Bayesian analyses to examine the amount of evidence in favor of this null result. To do this, we first split our sample based on headline veracity and then examined the evidence for (the absence of) an effect of condition on perceived headline accuracy and trustworthiness for the true-news sample only. Bayesian analyses using the R packages brm (Bürkner, 2017) and bayestestR (Makowski et al., 2019) showed moderate evidence in favor of a null effect of condition on the perceived accuracy of true news headlines ($BF_{10} = 0.25$) and the perceived trustworthiness of true news headlines ($BF_{10} = 0.11$). Further, we used additional Bayesian analyses to examine the amount of evidence in favor of the inoculation's effect (or lack thereof) on the perceived accuracy and trustworthiness of the fake news headlines. The results of these analyses showed strong evidence in favor of the alternative hypothesis for the inoculation lowering the perceived accuracy of fake headlines ($BF_{10} = 24.65$) and moderate evidence in favor of the null hypothesis for the inoculation lowering the perceived trustworthiness of fake headlines ($BF_{10} = 0.29$).

Finally, we examined the evidence for (the absence) of an effect of condition on media cynicism and general cynicism using Bayesian analyses with the default priors in JASP (Heo et al., 2020). The results showed moderate evidence in favor of a null effect of condition on both media cynicism ($BF_{10} = 0.16$) and general cynicism ($BF_{10} = 0.16$).

2.3. Discussion

In sum, the results of Study 1 provided evidence against inoculation side-effects: First, our inoculation did not lower the perceived accuracy or trustworthiness of true headlines, relative to the control condition (i.e., there was no support for H1a or H1b). Second, being inoculated against misinformation did not increase individuals' cynicism towards the media, nor their general cynicism (i.e., there was no support for H2a or H2b). However, we also found mixed evidence for the effectiveness of the inoculation: It successfully increased participants' ability to distinguish between true and fake news when they were asked to evaluate their accuracy, but not when they were asked to evaluate their trustworthiness.

Notably, the ecologically valid approach we pursued meant that the headlines we showed participants were not constructed to either contain (in the case of fake headlines) or not contain (in the case of true headlines) the misinformation strategies pointed out in the inoculation. As such, the effectiveness of the inoculation could have been reduced due to the existence of fake news headlines that did not contain these strategies, as well as the existence of true news that did contain them. To follow up on this possibility, in Study 2, we aimed to test whether the fake news headlines were actually more likely to use misinformation strategies than the true news headlines (H3).

3. Study 2

3.1. Method

3.1.1. Misinformation strategy use

In order to determine whether fake news headlines were more likely to use misinformation strategies than true news headlines, we presented participants with the headlines from Study 1 ($N = 120$) and had them rate each one on the extent to which they included the five strategies participants had been warned about in the inoculation manipulation in Study 1 (Discrediting Opponents, Emotional Language, Polarization, Conspiracy Theories, and Clickbait). Specifically, for each of the ten headlines they read, we asked them to respond to the following questions: 1) "To what extent does this headline cast doubt on the reputation of certain groups or public figures?", 2) "To what extent does this headline appeal to emotions?", 3) "To what extent is this headline polarizing?", 4) "To what extent does this headline use conspiratorial language (i.e., tries to make you doubt mainstream narratives and provide alternative stories in which a small sinister group of people is responsible for doing harm to many)?", 5) "To what extent does this headline try to clickbait readers?" (1 = not at all, 7 = completely). In addition, we had participants rate the headlines on a sixth factor – Ambiguity. As previous research has claimed that the effects of inoculation could be explained by a difference in ambiguity between true and fake news, such that the fake news stimuli in certain studies were more ambiguous than the true news stimuli (Modirrousta-Galian & Higham, 2023), we included ambiguity as an exploratory variable in Study 2. Thus, for each headline, participants were also asked "To what extent is this headline ambiguous (i.e., can be interpreted in more than one way)?"

3.1.2. Headline familiarity, news literacy, and demographics

We used the same measures as in Study 1 (presented in randomized order following the misinformation strategy ratings), with the exception of the item asking about participants' preferred news sources.

3.1.3. Sample

Sample size. We based our desired sample size on Rosenbusch et al. (2021), who showed that 25 raters are sufficient to achieve good inter-rater reliabilities and thus generate reliable mean inferences. We therefore planned to have each headline rated on each dimension at least 25 times. With each participant rating 10 headlines out of the total sample of 120, this required a total sample of 300 participants. To account for participants failing an attention check or completing the study in less than two minutes (i.e., our pre-registered exclusion criteria), we aimed to recruit 330 participants. Note that our final sample size, $N = 350$ (see "Participants" section below) provided at least 80 % power to detect effects of $b = 0.39$ and above in our multilevel analyses, when accounting for an alpha level of .05.[11]

Participants. As in Study 1, we recruited participants via Prolific. Since some Study 1 participants indicated lesser familiarity with some of the headlines due to their focus on American politics, we restricted our Study 2 sample to US residents. Additionally, those who had already participated in Study 1 were not invited to participate in Study 2. Participants were recruited in December 2023 and the initial sample consisted of $N = 350$ individuals. Twenty-four participants were excluded due to failing an attention check, resulting in a final sample of $N = 326$ (179 women, 144 men, 1 non-binary, 1 bigender, 1 participant who did not disclose their gender). Participants had an average age of $M = 41.36$ ($SD = 13.75$) and the majority (59 %) had at least a Bachelor's degree. On average, they reported their political orientation as slightly left of center ($M = 3.40$, $SD = 1.63$). Overall, participants indicated a moderate interest in keeping up with the news ($M = 3.24$, $SD = 1.18$), with half of them (50.9 %) checking the news at least once a day. Finally, in regards to their familiarity with the headlines, participants indicated an above-average familiarity with the names/people mentioned ($M = 4.55$, $SD = 1.02$), the locations mentioned ($M = 4.04$, $SD = 1.22$), and the topics mentioned ($M = 3.38$, $SD = 1.24$) in the headlines.

3.2. Results

3.2.1. Correlations

Correlations between all misinformation strategy ratings can be found in Table 1. As correlations between participants' ratings of the

[11] As calculated using a simulation-based sensitivity analysis conducted with the R package simr (Green & McLeod, 2016).

Table 1
Correlations between misinformation strategy ratings.

N = 325	*1*	*2*	*3*	*4*	*5*	*6*
1 = Discrediting Opponents	M = 4.58 SD = 0.90					
2 = Emotional Language	.597***	M = 4.43 SD = 1.04				
3 = Polarization	.659***	.646***	M = 4.41 SD = 0.88			
4 = Conspiracy Theories	.566***	.486***	.572***	M = 3.87 SD = 1.12		
5 = Clickbait	.459***	.448***	.507***	.617***	M = 4.59 SD = 1.06	
6 = Ambiguity	.305***	.262***	.364***	.555***	.424***	M = 3.64 SD = 1.21

Note. ***$p < .001$, **$p < .01$, *$p < .05$.

core five misinformation strategies were relatively high ($r > .40$), and the reliability of a composite variable was shown to be very good (Cronbach's $\alpha = 0.86$; McDonald's $\omega = 0.89$), we combined them into the variable "misinformation strategy use" ($M = 4.37$, $SD = 0.80$) by computing a mean rating score across the five strategies, which we analyzed along with the separate scores for each rating.

3.2.2. Comparing misinformation strategy use

We ran a series of multilevel models[12] with headline veracity as a predictor of misinformation strategy use, including a random intercept of participants and headlines. We found a significant difference ($t(114) = 3.30$, $b = 0.46$, $SE = 0.14$, CI [0.19; 0.73], $p = .001$, $d = 0.62$) between fake and true news headlines, such that fake headlines were rated as using more misinformation strategies overall ($M = 4.58$, $SD = 1.39$) compared to true headlines ($M = 4.17$, $SD = 1.45$).

3.2.3. Exploratory analyses

Comparing the use of individual strategies. We also looked at the use of the five individual misinformation strategies separately. We found a significant difference between true and fake headlines in the "Polarization" strategy ($t(114) = 2.29$, $b = 0.38$, $SE = 0.17$, CI [0.06; 0.71], $p = .024$, $d = 0.43$), such that fake headlines were rated as more polarizing ($M = 4.57$, $SD = 1.79$) than true headlines ($M = 4.24$, $SD = 1.79$); this was also the case for the "Conspiracy Theories" strategy ($t(115) = 3.45$, $b = 0.58$, $SE = 0.17$, CI [0.25; 0.91], $p < .001$, $d = 0.64$), in that fake headlines were rated as using more conspiratorial language ($M = 4.12$, $SD = 1.98$) than true headlines ($M = 3.62$, $SD = 1.98$); and the "Clickbait" strategy ($t(109) = 5.12$, $b = 0.70$, $SE = 0.14$, CI [0.43; 0.97], $p < .001$, $d = 0.98$), such that fake headlines were rated as using more clickbait ($M = 4.92$, $SD = 1.72$) than true headlines ($M = 4.25$, $SD = 1.83$). While fake headlines were generally rated as discrediting public figures more ($M = 4.75$, $SD = 1.93$) than true headlines ($M = 4.41$, $SD = 1.97$), this difference between the true and fake headlines in the "Discrediting Opponents" strategy was not statistically significant ($t(116) = 1.98$, $b = 0.39$, $SE = 0.20$, CI [0.00; 0.78], $p = .051$, $d = 0.37$). Similarly, there was no significant difference in the use of the "Emotional Language" strategy between true and fake headlines ($t(111) = 1.75$, $b = 0.24$, $SE = 0.14$, CI [−0.03; 0.51], $p = .082$, $d = 0.33$), despite the overall pattern indicating that fake headlines appealed to emotions slightly more ($M = 4.53$, $SD = 1.76$) than true headlines ($M = 4.32$, $SD = 1.81$). An overview of these results can be found in Fig. 3.

Ambiguity. To explore whether real-life fake headlines were more ambiguous than real-life true headlines, we compared the "Ambiguity" ratings between them. We found no significant difference in the (perceived) ambiguity of fake and true headlines ($t(107) = 1.50$, $b = 0.15$, $SE = 0.10$, CI [−0.05; 0.34], $p = .138$, $d = 0.29$), that is, fake headlines were not rated as more ambiguous ($M = 3.70$, $SD = 1.84$) than true headlines ($M = 3.59$, $SD = 1.81$) (see Fig. 3). We followed this analysis up with an additional Bayesian analysis to test whether the lack of difference in ambiguity between true and fake news reflected a true null effect. We found moderate evidence in favor of the null hypothesis ($BF_{10} = 0.27$).

3.3. Discussion

The results of Study 2 showed that fake headlines indeed used more misinformation strategies than true headlines, thus providing support for H3. Our exploratory analyses additionally showed that this effect was largely driven by the Conspiracy Theories and Clickbait strategies, with Polarization showing only a small difference between true and fake headlines. Interestingly, while it followed the same overall pattern, the difference in the Discrediting Opponents and Emotional Language ratings between true and fake headlines was not statistically significant, indicating that content disparaging public figures, as well as emotional content, is likely to be present in both. In Study 3, we set out to test the hypothesized mechanism of inoculation by examining whether the inoculation would be more effective for headlines that, on average, use more misinformation strategies, compared to those that use fewer misinformation strategies.

4. Study 3: Joint analysis of Study 1 and Study 2 data

4.1. Method

For Study 3, we merged the datasets from Study 1 and Study 2, such

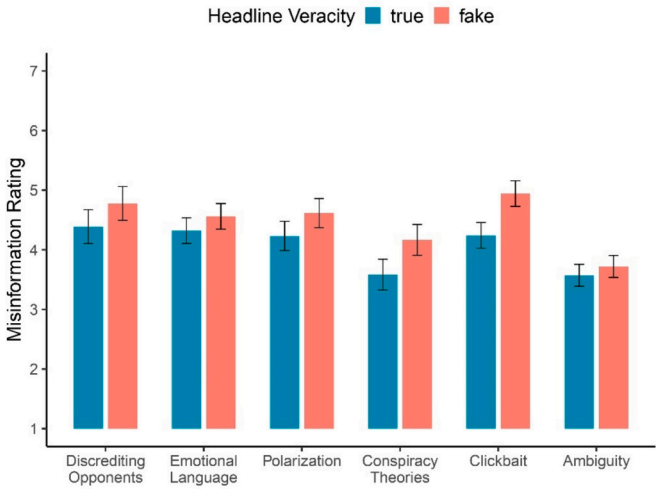

Fig. 3. Comparing the use of different misinformation strategies between true and fake news.
Note. Error bars indicate 95 % confidence intervals.

[12] Note that in our pre-registration, we indicated that we would use independent samples *t*-tests to test whether fake news headlines used more misinformation strategies than true news headlines. However, running multilevel regressions allowed us to account for systematic variation at the level of headlines and participants, which is why we opted for this more robust analysis instead.

that the information on each headline's perceived accuracy and trustworthiness following the inoculation (obtained from Study 1) was combined with the degree to which each headline was rated as high or low in misinformation strategy use (from Study 2). This allowed us to test whether the inoculation would work better on news headlines that used more misinformation strategies than on those that used fewer misinformation strategies (H4). Note that the sample size for Study 3 was therefore the same as the sample size for Study 1 ($N = 388$). A simulation-based sensitivity analysis conducted with the R package simr (Green & MacLeod, 2016) showed that this sample size provided at least 80 % power to detect effects of $b = 0.17$ and above in our multilevel analyses, when accounting for an alpha level of .05.

4.2. Results

4.2.1. Moderation by misinformation strategy use

First, we examined whether the use of misinformation strategies would affect the perceived accuracy of the news headlines across the Study 1 conditions. To do so, we ran a multilevel regression, with condition (1 = inoculation, 0 = control), misinformation strategy use (centered), and a condition x misinformation strategy use interaction as predictors, and perceived headline accuracy as the dependent variable. We also included a random intercept of participants and headlines. Results showed a significant interaction effect between condition and misinformation strategy use ($t(3601) = -2.52$, $b = -0.14$, CI [−0.24; −0.03], $SE = 0.05$, $p = .012$, $d = 0.08$). Using simple slopes to probe this interaction effect, we found that the inoculation reduced the perceived accuracy of headlines with an above average (+1 SD) misinformation strategy use ($t(653) = -3.81$, $b = -0.34$, $SE = 0.09$, CI [−0.51; −0.16], $p < .001$), but not that of headlines with a below average (−1 SD) misinformation strategy use ($t(649) = -1.31$, $b = -0.12$, $SE = 0.09$, CI [−0.29; 0.06], $p = .192$; see Fig. 4). Additional simple slope analyses revealed that misinformation strategy use had a stronger effect on the perceived accuracy of headlines in the inoculation condition ($t(174) = -5.55$, $b = -0.37$, $SE = 0.07$, CI [−0.50; −0.24], $p < .001$) than the control condition ($t(163) = -3.57$, $b = -0.23$, $SE = 0.07$, CI [−0.36; −0.11], $p < .001$).

Next, we ran the same analysis with perceived trustworthiness as the dependent variable. We found no significant interaction effect between condition and misinformation strategy use ($t(3605) = -1.22$, $b = -0.07$, $SE = 0.06$, CI [−0.18; 0.04], $p = .222$, $d = 0.04$), indicating that the inoculation did not affect the perceived trustworthiness of headlines with a higher use of misinformation strategies more than that of headlines with a lower use of misinformation strategies (see Fig. 5).

4.2.2. Exploratory analyses

Main effect of misinformation strategy use. Mirroring Study 1, we also explored the main effect of misinformation strategy use on the two credibility variables (perceived headline accuracy and trustworthiness), using multilevel regressions that included random intercepts of participants and stimuli. We found that misinformation strategy use was associated with lower perceived headline accuracy ($t(117) = -4.95$, $b = -0.30$, $SE = 0.06$, CI [−0.42, −0.18], $p < .001$, $d = 0.92$) and lower perceived headline trustworthiness ($t(116) = -5.52$, $b = -0.36$, $SE = 0.07$, CI [−0.49; −0.23], $p < .001$, $d = 1.02$).

Moderation by individual strategies. Furthermore, we examined the moderating effect of each misinformation strategy separately. Mirroring the results of Study 2 – which indicated that some strategies (i.e., Discrediting Opponents and Emotional Language) were not significantly different between true and fake news – we found no significant interactions between the inoculation condition and the use of the Discrediting Opponents ($t(3557) = -1.59$, $b = -0.06$, $SE = 0.04$, CI [−0.14; 0.01], $p = .112$, $d = 0.05$) or Emotional Language ($t(3715) = -1.77$, $b = -0.10$, $SE = 0.06$, CI [−0.21; 0.01], $p = .076$, $d = 0.06$) strategies on perceived headline accuracy. Conversely, the use of the Polarization ($t(3602) = -2.37$, $b = -0.11$, $SE = 0.05$, CI [−0.20; −0.02], $p = .018$, $d = 0.08$), Conspiracy Theories ($t(3539) = -2.09$, $b = -0.09$, $SE = 0.04$, CI [−0.18; −0.01], $p = .037$, $d = 0.07$), and Clickbait ($t(3698) = -3.38$, $b = -0.18$, $SE = 0.05$, CI [−0.28, −0.07], $p < .001$, $d = 0.11$) strategies significantly moderated the effect of the inoculation on perceived headline accuracy, such that the inoculation only significantly impacted the perceived accuracy of headlines that were rated as very polarizing (+1 SD: $t(655) = -3.72$, $b = -0.33$, $SE = 0.09$, CI [−0.50; −0.16], $p < .001$; -1SD: $t(651) = -1.37$, $b = -0.12$, $SE = 0.09$, CI [−0.29; 0.05], $p = .172$), conspiratorial (+1 SD: $t(652) = -3.60$, $b = -0.32$, $SE = 0.09$, CI [−0.49; −0.14], $p < .001$; -1SD: $t(648) = -1.54$, $b = -0.14$, $SE = 0.09$, CI [−0.31; 0.04], $p = .123$), or clickbaiting (+1 SD: $t(654) = -4.33$, $b = -0.39$, $SE = 0.09$, CI [−0.56; −0.21], $p < .001$; -1SD: $t(646) = -0.93$, $b = -0.08$, CI [−0.26; 0.09], $SE = 0.09$, $p = .353$). An overview of these results can be seen in Fig. 6.

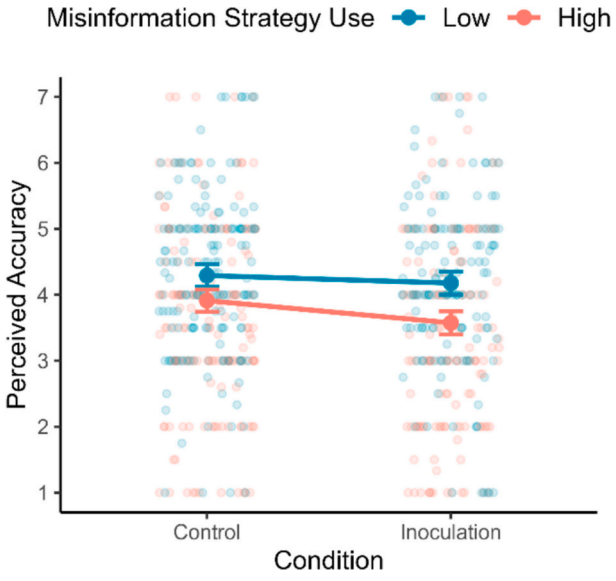

Fig. 4. Interaction effect between condition and misinformation strategy use on perceived headline accuracy.
Note: Low: -1SD; High: +1SD; error bars indicate 95 % confidence intervals.

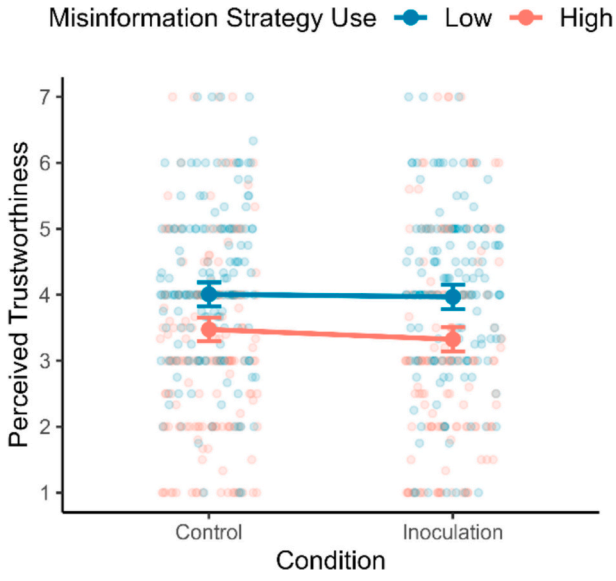

Fig. 5. Interaction effect between condition and misinformation strategy use on perceived headline trustworthiness.
Note: Low: -1SD; High: +1SD; error bars indicate 95 % confidence intervals.

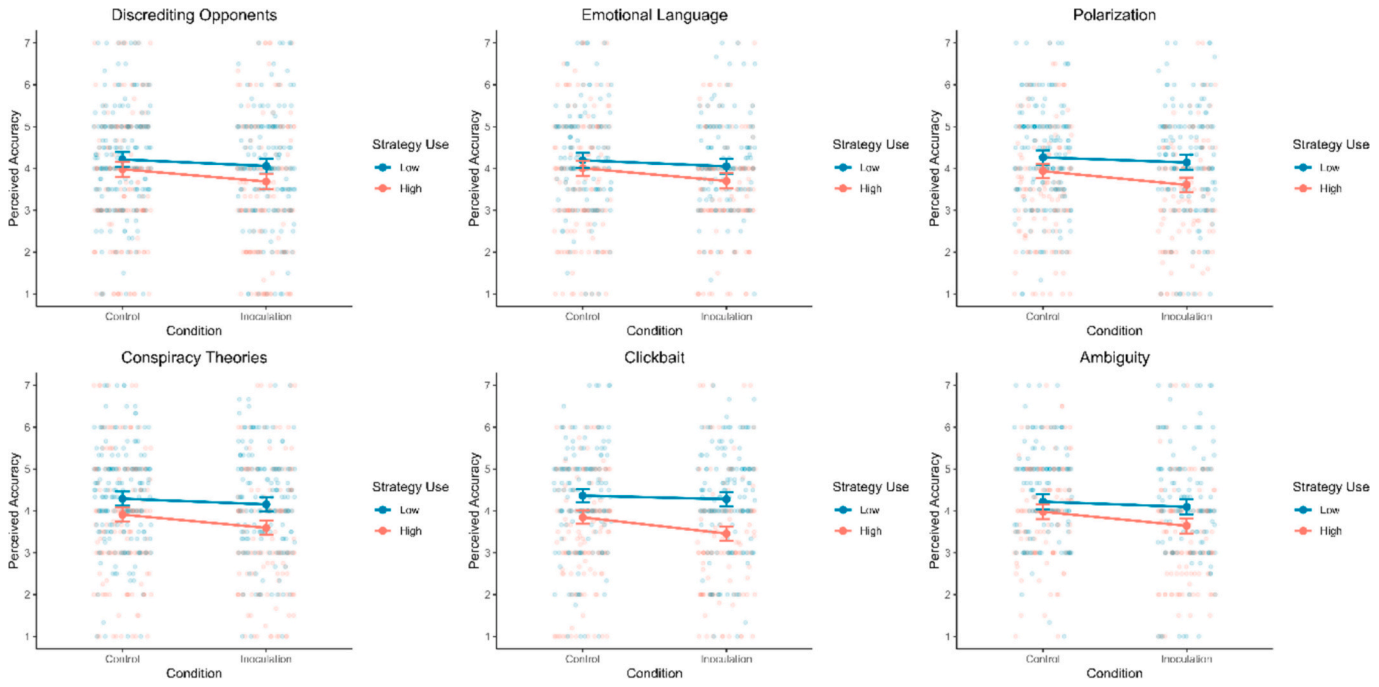

Fig. 6. Interaction effects between condition and the use of individual misinformation strategies on perceived headline accuracy.
Note: Low: -1SD; High: +1SD; error bars indicate 95 % confidence intervals.

There were no significant interaction effects between the inoculation and any of the misinformation strategies on the perceived trustworthiness of headlines (all *ps* > .05). We report the full statistics for these analyses in the Supplementary Materials.

Accounting for misinformation strategy use. To further explore the relevance of accounting for misinformation strategy use, we also examined whether controlling for the headlines' use of misinformation strategies would change the results of Study 1 (i.e., that participants given the inoculation treatment were more likely to rate fake headlines as less accurate than true headlines). The interaction between the condition and headline veracity became non-significant ($p = .08$) once we included the interaction between condition and misinformation strategy use in the model.

Next, we tested for a three-way interaction between condition, headline veracity and misinformation strategy use on the perceived accuracy of headlines. This interaction was not significant (but see Fig. 7 for a visual comparison of the effect of condition on misinformation strategy between true and fake news). The full results of both analyses are reported in the Supplementary Materials.

Including random slopes. To test the robustness of our findings, we again re-ran our main analyses including 1) a random slope of condition across headlines, 2) a random slope of headline veracity across participants, and 3) both of these random slopes in the same model. The interaction effect between the condition and misinformation strategy use on perceived headline accuracy remained significant across all three models, while the interaction effects on perceived headline trustworthiness remained non-significant. The full results of these analyses can be found in the Supplementary Materials.

4.3. Discussion

The findings of Study 3 demonstrated that misinformation strategy use indeed moderated the effect of the inoculation on the perceived accuracy of headlines. Specifically, the inoculation increased the

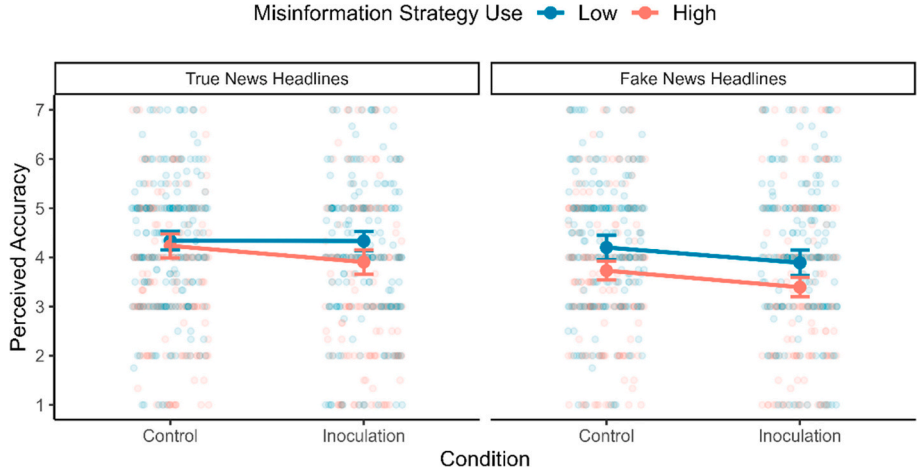

Fig. 7. Interaction effect between condition and misinformation strategy use on perceived headline accuracy for true vs. fake news.
Note: Low: -1SD; High: +1SD; error bars indicate 95 % confidence intervals.

perceived accuracy gap between headlines using more (vs. fewer) misinformation strategies, thus providing support for H4. Put differently, the inoculation worked better on headlines that used more (vs. fewer) misinformation strategies. In addition, we found that the interaction between the condition and headline veracity was no longer significant after controlling for the effect of misinformation strategy use, and that the condition x misinformation strategy use interaction effect was restricted to the misinformation strategies that significantly differed between true and fake news: Polarization, Conspiracy Theories, and Clickbait. This provides further evidence in support of the idea that the use of misinformation strategies in news headlines might be a mechanism through with the inoculation works, thereby validating assumptions about its underlying process.

5. General Discussion

5.1. Summary and implications

Over the past few years, inoculation has gained increasing popularity as a novel approach to fighting misinformation. Notably, the question of whether or not this approach is as effective – or as free from side-effects – as initially claimed, has been met with mixed results (Basol et al., 2021; Guess et al., 2020; Hameleers, 2023; Harrop et al., 2023; Hoes et al., 2024; Lu et al., 2023; Maertens et al., 2024; Modirrousta-Galian & Higham, 2023; Van der Meer et al., 2023). The aim of the present research was to provide a more robust and ecologically valid test of the effectiveness and unintended side-effects of inoculation by using a random stimuli approach (Judd et al., 2012) and sampling from a large-scale dataset of real-life headlines. As demonstrated by the findings of Study 1, we provide evidence against inoculation side-effects: Inoculating people did not reduce the extent to which they perceived true news headlines as accurate or trustworthy, did not increase their cynicism towards the media, and did not make them more cynical towards people in general. Thus, our findings align with previous research showing no negative repercussions of inoculation for future information processing (Basol et al., 2021, Study 1; Harrop et al., 2023, Study 1; Lu et al., 2023) or trust in institutions such as the media (Hoes et al., 2024; Leder et al., 2024).

Regarding the effectiveness of the inoculation in increasing discernment between true and fake news, our results were less conclusive – while inoculated participants were better able to distinguish between the accuracy of true and fake headlines than those who received a control treatment, this was not the case for the perceived trustworthiness of headlines. In other words, being inoculated (vs. receiving a control treatment) did not significantly reduce people's trust in fake news relative to true news. There could be several reasons for this discrepancy. First, it is possible that our operationalization of perceived headline credibility – as perceived accuracy on the one hand, and perceived trustworthiness on the other – captures two distinct constructs. In other words, it might be that an individual perceives a headline as accurate, but still does not trust it, or vice versa. Despite the high correlation between the two ($r = .80$), judgments of trustworthiness (vs. accuracy) might be shaped by a broader range of factors, such as pre-existing beliefs (e.g., Gervais et al., 2011; Ong et al., 2022). Indeed, previous research has shown that people can still believe information even when they know it is not fully accurate, as long as they consider its broader message to be true (Langdon et al., 2024). As such, whether something is perceived as accurate or trustworthy could be driven by different cognitive processes, and this distinction could have potentially undermined the effect of our manipulation.

Second, our studies showed a consistent main effect of headline veracity on perceived credibility, implying that people are generally quite good at distinguishing between true and fake news, even without being inoculated against misinformation (see also Basol et al., 2020; Roozenbeek & van der Linden, 2019). Indeed, when observing the effect of headline veracity (true vs. fake) in the control condition only, we found that this effect was larger for the perceived trustworthiness than for the perceived accuracy of headlines ($b = -0.56$ vs. $b = -0.41$). As such, the discrepancy in findings between the two types of headline ratings might be (at least partially) due to the fact that being inoculated does not substantially contribute to people's already good face-value judgment of true vs. fake headlines' trustworthiness.

Third, it is possible that the order of the two ratings influenced our findings, such that judging accuracy first affected how participants approached judgments of trustworthiness. For example, it could be that rating a headline as less accurate pushed participants to then lower their trust of it moreso than if they had been asked to judge the headline's trustworthiness alone. Conversely, it might be that the inoculation simply worked on whichever measure participants saw first, and that its effects wore off by the time participants responded to the second measure. Indeed, the overall pattern of results we found was not substantially different for trustworthiness vs. accuracy (i.e., both were negatively affected by the condition).

One final possibility is that, given our improved methodological and analytical approach, which is notably better suited for avoiding false positives (Judd et al., 2012), the strength of the overall inoculation effect was generally diminished, resulting in inconsistent findings. Previous research has generally found the effects of inoculation to be of small-to-medium size (d's between 0.18 and 0.22 for discernment between true and fake news; Lu et al., 2023), while the average across our studies was a very small effect of $d = 0.08$.

In addition to testing the inoculation itself, we were also interested in examining the underlying assumption of inoculation research – that fake news (headlines) use more misinformation strategies than true news (headlines). Indeed, the results of Study 2 show that, on average, fake news headlines are rated higher in misinformation strategy use. However, when comparing the presence of individual strategies (Discrediting Opponents, Emotional Language, Polarization, Conspiracy Theories, Clickbait) in the two types of headlines, the picture was less clear. Specifically, we found no significant difference between the Discrediting Opponents and Emotional Language ratings of true vs. fake news headlines.

This is particularly relevant, as Study 3 also showed that headlines rated as high in (average) misinformation strategy use were more affected by the inoculation (as one would expect given that the inoculation specifically addresses those strategies). Notably, this effect was driven by the three misinformation strategies whose use differed significantly between true and fake news – Polarization, Conspiracy Theories, and Clickbait – while we found no moderating effects of the Discrediting Opponents or Emotional Language strategies. Taken together, these findings indicate that the success of the inoculation depends (at least to some extent) on the use (and type) of the misinformation strategies that are included in it.

One possible implication of this finding is that inoculating against misinformation runs the risk of reducing the credibility of true headlines, if those headlines use specific misinformation strategies. Notably, with rising societal and political polarization on the one hand, and an increasing need for print media to keep itself relevant in the "attention economy" (e.g., through attention-grabbing headlines) on the other, it is likely that some of the commonly inoculated-against misinformation strategies will indeed be increasingly present in true news (headlines) as well (see also Feldman et al., 2015; Hart et al., 2020). While it can be argued that media using these strategies might not be of particularly high quality, and being overly cautious of it is therefore not necessarily detrimental, scholars have pointed out that even reputable news sources are prone to reporting that can be considered polarizing, emotional, and fearmongering (see Modirrousta-Galian & Higham, 2023). As such, to the extent that traditional media increasingly starts relying on such strategies to garner clicks, inoculation – or at least inoculation that focuses solely on warning individuals about the use of these particular strategies – might become a less effective tool in the future, particularly as it leads to increasing false positives (i.e., judging news as fake when it

is actually true).

Lastly, it is worth considering whether the negative effect of misinformation strategy use that we find in our analyses (i.e., the tendency to rate headlines using misinformation strategy as less accurate and trustworthy, regardless of whether one was previously inoculated) implies that using those strategies is not a "successful" approach for individuals looking to spread misinformation. Specifically, one could argue that as long as headlines containing such strategies do not appear credible at face value, this approach is actually prone to backfiring (i.e., fewer people falling for the misinformation). However, previous research has shown that merely encountering a certain piece of information repeatedly tends to increase its believability (Vellani et al., 2023). Therefore, as long as the use of misinformation strategies helps these headlines spread (e.g., by emotionally appealing to people or creating clickbait titles), being viewed as less accurate and trustworthy does not necessarily contradict their potential "success" in spreading misinformation.

5.2. Limitations and future directions

While the present research aimed to provide an improved methodological approach to testing inoculation, as well as its underlying assumptions, our study design is still subject to some limitations. First, we opted for a text-based inoculation, which focused on warning participants about four of the six misinformation strategies derived from the DEPICT framework (Basol et al., 2020; Maertens et al., 2020; Roozenbeek & van der Linden, 2019), along with a fifth strategy – Clickbait. This framing of the inoculation in line with common news literacy interventions (see e.g., Facebook's "Tips to spot fake news", or "10 tips from WhatsApp's to spot fake news") provided additional external validity, but also differs from other recent approaches, such as inoculation games where participants take the role of a fake news producer and actively spread misinformation in order to gain points (see e.g., "Bad News", Roozenbeek & van der Linden, 2019). Because these gamified inoculations use more examples of fake news, might be more engaging for participants, and are also administered for a longer period of time, it is possible that our findings were comparatively limited by the broader approach, shorter time frame, and more static nature of our text-based inoculation. As such, we call on future research to replicate these findings in a study design that combines our improved methodological approach with the dynamic context of a gamified inoculation.

Furthermore, we used comprehension checks (in multiple choice format) to ensure that participants fully read and understood the inoculation text we presented them with. However, the difficulty of these comprehension checks might have differed between conditions and resulted in an imbalance in the amount of exclusions in the inoculation (47 participants) vs. control condition (25 participants). Although both exclusion rates fall within the norm for comprehension check-based exclusions at 10–20 % (see e.g., Thomas & Clifford, 2017), we acknowledge that this imbalance might have partially affected our results. For example, it is possible that the increased difficulty of the comprehension check in the inoculation condition relative to the control condition resulted in a selection of participants who were more highly educated, or more news savvy. To test this possibility, we conducted exploratory analyses comparing the two samples of excluded participants on key demographic variables (i.e., education, news interest, and news consumption) and found that they did not significantly differ from one another. Nevertheless, there might have been other selection variables (e.g., general motivation to take part in the study or tiredness) that we were not able to compare. As such, we encourage future studies to ensure a similar difficulty level of the comprehension check questions across the two conditions.

Another trade-off arising from our study design concerns the choice to use a between-subjects approach to test the inoculation's effectiveness (i.e., having participants rate the true and fake news headlines only once). This was done to avoid contamination or order effects, given that we only tested the inoculation in a one-shot study. While previous research has used both between- and within-subjects designs, and our study is therefore not an outlier in this regard, we note that testing the inoculation's effect using a within-subjects design, in which the headlines are rated both before and after the inoculation, would have increased our power and likely produced larger effects than the ones we observe here.

In addition, our use of randomly selected real-life headlines ensured a certain level of ecological validity, but also limited our headline sample to a specific time and place (i.e., the 2016 US election cycle). Because we did not restrict our Study 1 sample to US participants, their (lack of) familiarity with the content of the headlines could have affected our findings. Note, however, that due to the centrality of the US in global discourse, and the accessibility of English-language news, global audiences are disproportionately likely to be exposed to US-centric news in their daily lives, especially during US election cycles (Wu, 2000). Indeed, we both measured and controlled for headline familiarity in Study 1, which did not significantly change the outcome of our analyses. Nevertheless, future research replicating our findings with US participants (or otherwise matching news and participant origin), as well as including more recent real-life headlines, would add to the robustness of our findings.

Similarly, the need to select real-life headlines that fit with the format of our questions about perceived accuracy and trustworthiness meant that a substantial amount of the headlines we initially sampled (around 65 %) had to be excluded, for example because they were not phrased as statements or were opinion pieces. This rather high number of exclusions exemplifies the "messiness" of real-life headlines (in contrast to the artificially created headlines used in previous inoculation research, see e.g., Basol et al., 2020), and further emphasizes the importance of aligning the methods we use to test inoculation with this level of real-life variability. Note that, despite this high number of exclusions, the exclusion rate was the same across headline type (true vs. fake).

Moreover, in choosing an equal number of true and fake news headlines, we prioritized a balanced design, which helps reduce potential issues arising from unequal cell sizes, such as unstable variance estimates, inflated standard errors, and reduced power to detect interactions (Landsheer & van den Wittenboer, 2015; Ramírez-Tapia et al., 2022) – but is not necessarily externally valid (i.e., representative of the amount of true and fake news that people actually read in their daily lives). Notably, previous research has shown that conservative-leaning news media often contains more fake news than liberal-leaning news media (e.g., Allcott & Gentzkow, 2017), implying that the ratio of fake-to-true news one consumes might differ by political orientation. As such, future research could more explicitly tailor the amount of true and fake headlines to be representative of the desired sample's actual news consumption, in order to further improve external validity.

Finally, our studies are somewhat limited by their short timeframe. This is particularly relevant in regards to the side-effects we examined. It might be that certain side-effects only develop after repeated exposure to inoculation over time, which our cross-sectional design could not effectively capture. Indeed, both media cynicism and general cynicism are often suggested to be responses to repeated instances of disappointment or betrayal (either by media actors or other people) (Markov & Min, 2022; Stavrova & Ehlebracht, 2016). With plans to extend the use of inoculation interventions, for example on the video-hosting platform YouTube (see also Roozenbeek et al., 2022), more longitudinal research is needed to examine the possibility that such side-effects develop over time, which thus cannot be ruled out entirely by cross-sectional research alone.

6. Conclusion

Using an improved analytical approach and real-life headlines, we

provide evidence for the robustness of inoculation interventions against side-effects, such as decreased credibility of true information and cynicism. However, we also find mixed support for the effectiveness of inoculation overall, as inoculated individuals perceive fake news as less accurate, but not less trustworthy than true news. In addition, we demonstrate that, while most misinformation strategies are more common in (real-life) fake than (real-life) true news, some strategies (e.g., emotional language) are present in both. Thus, while we do not deem inoculation to be harmful, we encourage the use of more rigorous analytical methods in testing its effectiveness, as well as a readiness to adapt its contents in the context of changing media landscapes in the future.

CRediT authorship contribution statement

Teodora Spiridonova: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Olga Stavrova:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Ilja van Beest:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2025.104806.

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Amazeen, M. A. (2021). Resisting covert persuasion in digital news: Comparing inoculation and reactance in the processing of native advertising disclosures and in article engagement intentions. *Journalism and Mass Communication Quarterly, 98*(4), 1129–1156. https://doi.org/10.1177/1077699020952131

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

Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. V. D. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society, 8*(1). https://doi.org/10.1177/20539517211013868, 20539517211013868.

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bessarabova, E., & Banas, J. A. (2024). Uncertainty and inoculation: Instilling resistance to anti-vaccination conspiracy propaganda. *Communication Quarterly*, 1–18. https://doi.org/10.1080/01463373.2024.2368124

Bürkner, P. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Choy, B. K., Eom, K., & Li, N. P. (2021). Too cynical to reconnect: Cynicism moderates the effect of social exclusion on prosociality through empathy. *Personality and Individual Differences, 178*. https://doi.org/10.1016/j.paid.2021.110871. article 110871.

Compton, J. (2013). Inoculation theory. In J. P. Dillard, & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (2nd ed., pp. 220–236). Sage Publications, Inc. https://doi.org/10.4135/9781452218410.n14.

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

Feldman, L., Hart, P. S., & Milosevic, T. (2015). Polarizing news? Representations of threat and efficacy in leading US newspapers' coverage of climate change. *Public Understanding of Science, 26*(4), 481–497. https://doi.org/10.1177/0963662515595348

Gervais, W. M., Shariff, A. F., & Norenzayan, A. (2011). Do you believe in atheists? Distrust is central to anti-atheist prejudice. *Journal of Personality and Social Psychology, 101*(6), 1189–1206. https://doi.org/10.1037/a0025882

Green, P., & MacLeod, C. J. (2016). Simr: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Greenglass, E. R., & Julkunen, J. (1989). Construct validity and sex differences in Cook-Medley hostility. *Personality and Individual Differences, 10*(2), 209–218. https://doi.org/10.1016/0191-8869(89)90206-7

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

Hameleers, M. (2023). The (un) intended consequences of emphasizing the threats of mis-and disinformation. *Media and Communication, 11*(2), 5–14. https://doi.org/10.17645/mac.v11i2.6301

Harrop, I., Roozenbeek, J., Madsen, J., & van der Linden, S. (2023). Inoculation can reduce the perceived reliability of polarizing social media content. *International Journal of Communication, 17*, 5291–5315.

Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and polarization in COVID-19 news coverage. *Science Communication, 42*(5), 679–697. https://doi.org/10.1177/1075547020950735

Heo, I., Veen, D., & Van de Schoot, R. (2020). Tutorial: JASP for Bayesian analyses with default priors. *Zenodo*. https://doi.org/10.5281/zenodo.4008339

Hoes, E., Aitken, B., Zhang, J., Gackowski, T., & Wojcieszak, M. (2024). Prominent misinformation interventions reduce misperceptions but increase scepticism. *Nature Human Behaviour, 8*(8), 1545–1553. https://doi.org/10.1038/s41562-024-01884-x

Iyengar, S., & Massey, D. S. (2019). Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences, 116*(16), 7656–7661. https://doi.org/10.1073/pnas.1805868115

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

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54–69. https://doi.org/10.1037/a0028347

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Kaplan, S. A., Bradley, J. C., & Ruscher, J. B. (2004). The inhibitory role of cynical disposition in the provision and receipt of social support: The case of the September 11th terrorist attacks. *Personality and Individual Differences, 37*(6), 1221–1232. https://doi.org/10.1016/j.paid.2003.12.006

Landsheer, J. A., & van den Wittenboer, G. (2015). Unbalanced 2 x 2 factorial designs and the interaction effect: a troublesome combination. *PLoS One, 10*(3). https://doi.org/10.1371/journal.pone.0121412. article e0121412.

Langdon, J. A., Helgason, B. A., Qiu, J., & Effron, D. A. (2024). "It's Not Literally True, But You Get the Gist:" How nuanced understandings of truth encourage people to condone and spread misinformation. *Current Opinion in Psychology, 57*. https://doi.org/10.1016/j.copsyc.2024.101788. article 101788.

Larson, H. J., Cooper, L. Z., Eskola, J., Katz, S. L., & Ratzan, S. (2011). Addressing the vaccine confidence gap. *The Lancet, 378*(9790), 526–535. https://doi.org/10.1016/S0140-6736(11)60678-8

Leder, J., Schellinger, L. V., Maertens, R., van der Linden, S., Chryst, B., & Roozenbeek, J. (2024). Feedback exercises boost discernment of misinformation for gamified inoculation interventions. *Journal of Experimental Psychology: General, 153*(8), 2068–2087. https://doi.org/10.1037/xge0001603

Lu, C., Hu, B., Li, Q., Bi, C., & Ju, X. D. (2023). Psychological inoculation for credibility assessment, sharing intention, and discernment of misinformation: Systematic review and meta-analysis. *Journal of Medical Internet Research, 25*. https://doi.org/10.2196/49255. article e49255.

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

Maertens, R., Götz, F. M., Golino, H. F., Roozenbeek, J., Schneider, C. R., Kyrychenko, Y., & van der Linden, S. (2024). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. *Behavior Research Methods, 56*(3), 1863–1899. https://doi.org/10.3758/s13428-023-02124-2

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. https://doi.org/10.1037/xap0000315

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software, 4*(40). https://doi.org/10.21105/joss.01541, 1541. 10.21105/joss.01541.

Markov, Č., & Min, Y. (2022). Understanding the public's animosity toward news media: Cynicism and distrust as related but distinct negative media perceptions. *Journalism*

and Mass Communication Quarterly, 99(4), 1099–1125. https://doi.org/10.1177/10776990211061764

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

McIntire, G. (2017). *fake_real_news_dataset* [Dataset]. https://github.com/GeorgeMcIntire/fake_real_news_dataset.

Meier, B., & Perrig, W. J. (2000). Low reliability of perceptual priming: Consequences for the interpretation of functional dissociations between explicit and implicit memory. *The Quarterly Journal of Experimental Psychology: Section A, 53*(1), 211–233. https://doi.org/10.1080/713755878

Modirrousta-Galian, A., & Higham, P. A. (2023). Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology: General, 152*(9), 2411–2437. https://doi.org/10.1037/xge0001395

Mozur, P. (2018, October 15). *A genocide incited on Facebook, with posts from Myanmar's military*. The New York Times. https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

Ong, H. H., Evans, A. M., Nelissen, R. M., & van Beest, I. (2022). Belief in karma is associated with perceived (but not actual) trustworthiness. *Judgment and Decision making, 17*(2), 362–377. https://doi.org/10.1017/S1930297500009141

Poland, G. A., & Spier, R. (2010). Fear, misinformation, and innumerates: how the Wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine, 28*(12), 2361–2362. https://doi.org/10.1016/j.vaccine.2010.02.052

Ramírez-Tapia, R., Ríos-Lira, A. J., Pantoja-Pacheco, Y. V., Vázquez-López, J. A., & Ruelas-Santoyo, E. A. (2022). Quantitative Analysis of the Balance Property in Factorial Experimental Designs 2^4 to 2^8. *Mathematics, 10*(20), 3812. https://doi.org/10.3390/math10203812

Roozenbeek, J., Maertens, R., McClanahan, W., & van der Linden, S. (2021). Disentangling item and testing effects in inoculation research on online misinformation: Solomon revisited. *Educational and Psychological Measurement, 81*(2), 340–362. https://doi.org/10.1177/0013164420940378

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications, 5*(1), 1–10. https://doi.org/10.1057/s41599-019-0279-9

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

Roozenbeek, J., van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances, 8*(34). https://doi.org/10.1126/sciadv.abo6254. article eabo6254.

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

Rosenbusch, H., Aghaei, M., Evans, A. M., & Zeelenberg, M. (2021). Psychological trait inferences from women's clothing: Human and machine prediction. *Journal of Computational Social Science, 4*(2), 479–501. https://doi.org/10.1007/s42001-020-00085-6

Sieradski, D. (2016). *BS Detector* [Chrome extension]. https://github.com/selfagency/bs-detector.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72–101. https://doi.org/10.2307/1412159

Stavrova, O., & Ehlebracht, D. (2016). Cynical beliefs about human nature and income: Longitudinal and cross-cultural analyses. *Journal of Personality and Social Psychology, 110*(1), 116–132. https://doi.org/10.1037/pspp0000050

Stavrova, O., & Ehlebracht, D. (2019). Broken Bodies, Broken Spirits: How Poor Health Contributes to a Cynical Worldview. *European Journal of Personality, 33*(1), 52–71. https://doi.org/10.1002/per.2183

Stavrova, O., Ehlebracht, D., & Ren, D. (2023). Cynical people desire power but rarely acquire it: Exploring the role of cynicism in leadership attainment. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12685. Advance online publication.

Stavrova, O., Ehlebracht, D., & Vohs, K. D. (2020). Victims, perpetrators, or both? The vicious cycle of disrespect and cynical beliefs about human nature. *Journal of Experimental Psychology: General, 149*(9), 1736–1754. https://doi.org/10.1037/xge0000738

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior, 77*, 184–197. https://doi.org/10.1016/j.chb.2017.08.038

Van der Linden, S., Dixon, G., Clarke, C., & Cook, J. (2021). Inoculating against COVID-19 vaccine misinformation. *EClinicalMedicine, 33*. https://doi.org/10.1016/j.eclinm.2021.100772

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 Meer, T. G., Hameleers, M., & Ohme, J. (2023). Can fighting misinformation have a negative spillover effect? How warnings for the threat of misinformation can decrease general news credibility. *Journalism Studies, 24*(6), 803–823. https://doi.org/10.1080/1461670X.2023.2187652

Van Duyn, E., & Collier, J. (2019). Priming and fake news: The effects of elite discourse on evaluations of news media. *Mass Communication and Society, 22*(1), 29–48. https://doi.org/10.1080/15205436.2018.1511807

Vellani, V., Zheng, S., Ercelik, D., & Sharot, T. (2023). The illusory truth effect leads to the spread of misinformation. *Cognition, 236*. https://doi.org/10.1016/j.cognition.2023.105421, 105421.

Vivion, M., Anassour Laouan Sidi, E., Betsch, C., Dionne, M., Dubé, E., Driedger, S. M., & Canadian Immunization Research Network (CIRN). (2022). Prebunking messaging to inoculate against COVID-19 vaccine misinformation: an effective strategy for public health. *Journal of Communication in Healthcare, 15*(3), 232–242. https://doi.org/10.1080/17538068.2022.2044606

Wood, M. L. M. (2007). Rethinking the inoculation analogy: Effects on subjects with differing preexisting attitudes. *Human Communication Research, 33*(3), 357–378. https://doi.org/10.1111/j.1468-2958.2007.00303.x

Wu, H. D. (2000). Systemic determinants of international news coverage: A comparison of 38 countries. *Journal of Communication, 50*(2), 110–130. https://doi.org/10.1111/j.1460-2466.2000.tb02844.x