International Journal of Communication 17(2023), 5291–5315 1932–8036/20230005

Inoculation Can Reduce the Perceived Reliability of Polarizing Social Media Content

ISOBEL HARROP*
London School of Economics and Political Science, UK

JON ROOZENBEEK*[1]
University of Cambridge, UK

JENS KOED MADSEN
London School of Economics and Political Science, UK

SANDER VAN DER LINDEN
University of Cambridge, UK

Little research is available on psychological interventions that counter susceptibility to polarizing online content. We conducted 3 studies (n_1 = 472, n_2 = 193, n_3 = 772) to evaluate whether psychological resistance against polarizing social media content can be conferred, using the *Bad News* game, a "technique-based inoculation" intervention that simulates a social media feed. We investigate (1) whether technique-based inoculation can reduce susceptibility to content designed to fuel intergroup polarization; (2) whether technique-based inoculation can offer cross-protection against misinformation techniques that people were not inoculated against; and (3) whether political ideology plays a role in how people engage with anti-misinformation interventions. In Studies 1 and 3 (but not Study 2), we found that technique-based inoculation significantly reduces the perceived reliability of polarizing content and offers partial cross-protection against untreated misinformation techniques. We found no effect for attitudinal certainty and news-sharing intentions. Finally, we report preliminary evidence that people may choose to engage with politically congruent news topics within the intervention.

Isobel Harrop: izzy.harrop@gmail.com
Jon Roozenbeek (corresponding author): jjr51@cam.ac.uk
Jens Koed Madsen: j.madsen2@lse.ac.uk
Sander van der Linden: sander.vanderlinden@psychol.cam.ac.uk
Date submitted: 2021-11-17

[1] Isobel Harrop and Jon Roozenbeek contributed equally to this work. We have no conflict of interest to disclose. We are grateful for funding from the British Academy (#PF21\210010), IRIS Coalition (UK Government, #SCH-00001-3391), and JITSUVAX (EU Horizon 2020, #964728).

Keywords: misinformation, polarization, inoculation theory, gamification, social media

The proliferation of misinformation on social networks is a significant societal problem that has proved difficult to mitigate at scale (Van Bavel et al., 2020). For example, belief in misinformation has been linked to reduced compliance with public health guidelines and lower intentions to get vaccinated against COVID-19 (Roozenbeek et al., 2020).

Recently, researchers have begun to explore the spread of information on social media that is not necessarily *false* but harmful in other ways (Simchon, Brady, & Van Bavel, 2021). For example, disinformation campaigns are often not aimed at spreading false information per se but rather seek to sow distrust and increase intergroup divisions (Keller, Schoch, Stier, & Yang, 2020). To do so, disinformation producers identify contentious issues such as abortion, racial strife, or government overreach and paint an exaggerated or emotionally charged picture of a particular topic. The goal of spreading such content is to "troll and divide" (Simchon et al., 2021, p. 1), and this method appears to be successful. Rathje, Van Bavel, and van der Linden (2021) showed that out-group-focused language predicts online engagement: Negative out-group language evokes anger and is shared and retweeted significantly more than (positive) content about in-groups. Given the prevalence of online echo chambers, exposure to polarizing social media content may be associated with increased affective as well as political polarization (see Kubin & von Sikorski, 2021, for a review).

Despite several studies investigating the possibility of reducing susceptibility to polarizing online content (Allcott, Braghieri, Eichmeyer & Gentzkow, 2020; Bail et al., 2018), little research has explored the feasibility of reducing polarization on social media through psychological interventions (Kubin & von Sikorski, 2021).

Inoculation Theory

Inoculation theory (McGuire, 1964) is a framework for designing interventions aimed at reducing susceptibility to persuasion and manipulative online content. This approach is inspired by a biomedical analogy in which a weakened dose of a viral pathogen triggers the production of antibodies to help fight off future infection. Inoculation theory posits that the same can be achieved with information: Preemptively exposing people to weakened doses of misinformation can confer psychological resistance against future unwanted persuasion. Inoculation works through two mechanisms: (1) a forewarning of a threat against a person's beliefs to motivate a resistance response and (2) a preemptive refutation of the persuasive argument, which scaffolds people's counterarguments against misinformation (Compton, Van der Linden, Cook, & Basol, 2021). Prior research has investigated the effectiveness of inoculation interventions within the context of misinformation about vaccines (Jolley & Douglas, 2017), climate change (Cook, Lewandowsky, & Ecker, 2017), and immigration (Roozenbeek & van der Linden, 2018). More recently, researchers have explored *technique-based inoculations*, which seek to confer psychological resistance against common techniques or tropes that underlie misinformation, rather than individual examples (see Traberg, Roozenbeek, & van der Linden, 2022, for an overview).

An open question within inoculation research is the extent to which inoculating individuals against a *specific* misinformation technique (such as the use of polarizing language) also confers psychological resistance against *other* misinformation techniques that people were not inoculated against—so-called cross-protection (Parker, Rains, & Ivanov, 2016). Cross-protection is theorized to work through reflective cognition about the topic of the inoculation, which makes people more aware of, and therefore less susceptible to, related persuasion attempts. If cross-protection can be shown to occur within the context of online misinformation, then inoculation interventions could confer resistance against multiple techniques used to mislead people (see also Roozenbeek, Traberg, & van der Linden, 2022).

Active Inoculation: The *Bad News* Game

As opposed to passive inoculation, whereby people are provided with a refutation (or "prebunk") by the experimenter, active inoculation occurs when people generate their own counterarguments against the misinformation (Compton et al., 2021). One example of an active inoculation intervention is the online game *Bad News* (https://www.getbadnews.com/). Players take on the role of a misinformation producer and learn about six common misinformation techniques: impersonating fake accounts; using emotionally manipulative language; spreading conspiracy theories; discrediting opponents through ad hominem attacks; evoking outrage through trolling; and fueling intergroup polarization (van der Linden & Roozenbeek, 2020). Their goal is to start a "fake news" website and gain followers while building credibility. If players' credibility meter drops to 0, they lose the game. Players win by playing through each level, using each manipulation technique, and eventually becoming a "fake news tycoon." Figure 1 shows screenshots of the game environment.

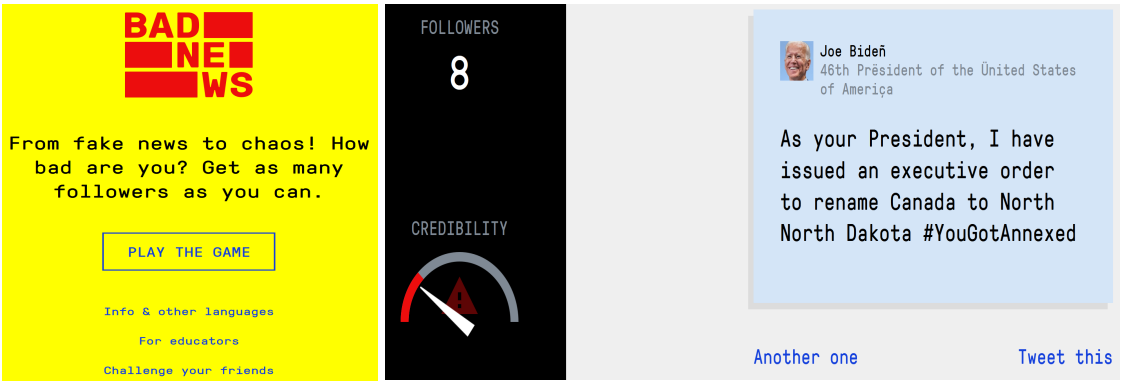


Figure 1. Screenshots of the Bad News (https://www.getbadnews.com/) landing page (left) and game environment (right).

Previous research into the effectiveness of *Bad News* has shown that playing the game significantly reduces the perceived reliability of social media content that makes use of the misinformation techniques learned in the game (Roozenbeek & van der Linden, 2019); that players become significantly more confident in their ability to assess misinformation (Basol, Roozenbeek, & van der Linden, 2020); and that these effects remain significant for several weeks (Maertens, Roozenbeek, Basol, & van der Linden, 2021). However, several questions about the effectiveness of the game remain unanswered (Traberg et al., 2022). Basol et

al. (2020) highlight that *Bad News* conferred the smallest effect for the perceived reliability of content that uses polarizing language. The authors suggest that prior attitudes related to participants' political ideology (Roozenbeek et al., 2022) may be responsible for this reduced effect, given that *Bad News* allows players to make choices within the game that are congenial to their political beliefs. Players may self-select into being exposed only to politically congruent polarizing content and therefore not learn to spot the "polarization" technique in content that is incongruent with their beliefs.

Within the game, players are given the option to spread misinformation about either the government or large corporations, or about either rising crime rates or police brutality. In the United States, Republicans/conservatives are more concerned than Democrats/liberals about rising crime rates (Pew Research Center, 2020) and are more distrustful of the government (at least when this study was conducted; Pew Research Center, 2019, 2021). Conversely, U.S. Democrats/liberals are more likely than Republicans/conservatives to view police brutality and the influence of large corporations on politics and society as important societal issues (Pew Research Center, 2019). Thus, if political attitudes play a role in how people engage with the *Bad News* intervention, we would expect Republican/conservative players to be more likely to choose to spread misinformation about government overreach and rising crime rates. Conversely, we would expect Democrat/liberal players to be more likely to spread misinformation about police brutality and large corporations. *Bad News* allows for the recording of participants' in-game choices, providing an opportunity to test whether this partisan self-selection plays a role in the effectiveness of the game as an inoculation intervention. Investigating whether this is the case is important to the field of misinformation research because previous research has shown that there may be discrepancies between the political left and right in terms of the effectiveness of some psychological interventions (Rathje, Roozenbeek, Traberg, Van Bavel, & van der Linden, 2022).

Finally, *Bad News* and other inoculation interventions have primarily tested whether people's ability to spot misinformation improves post-gameplay. However, the influence of inoculations on people's attitudes has been investigated far less. We do not know whether technique-based inoculation influences people's attitudes toward out-groups such as members of an opposing political party. It is important to investigate whether misinformation interventions have inadvertent side effects, like increasing affective polarization (Druckman & Levendusky, 2019).

In this study, we address three questions. First, can inoculation improve people's ability to recognize polarizing social media content? Second, can inoculating people against one misinformation technique confer cross-protection (i.e., psychological resistance against misinformation that makes use of other techniques that people were not inoculated against)? And third, what role do political attitudes play in how people engage with misinformation interventions?

The Present Research

To address these questions, we ran three separate studies using the *Bad News* game, conducted in September and October 2021. To test whether cross-protection against untreated misinformation techniques was achieved, we created a shortened version of the game, which only featured the polarization scenario (https://www.getbadnews.com/polarization). This version takes about 3–5 minutes to complete.

By playing this shortened version, people are only inoculated against the polarization technique and not the other techniques in the full *Bad News* game, which takes about 15 minutes.

We list our specific hypotheses separately in each study. Our OSF page contains all information required to replicate our findings and methods, including our data sets, Qualtrics surveys, preregistrations, supplementary tables, and our analysis and visualization scripts: https://osf.io/v9y6t/. All supplementary figures and tables are labeled with S. More information about the samples from Studies 1–3 can be found in Table S1 and Supplement S1. All studies were approved by the Department of Psychological and Behavioural Science at the London School of Economics (Reference #25363).

Study 1

Following Roozenbeek and Van der Linden (2019), we implemented a voluntary pre–post survey within the *Bad News* game. Because the game is available online, we relied on a convenience sample for participant recruitment (players from all over the world navigate to the *Bad News* website, so we did not limit our sample to a particular country or region). We tested the following (non-preregistered) hypotheses:

H1: Participants who play the Bad News game rate "polarizing" social media content as significantly less reliable post-gameplay (posttest) than at the start of the game (pretest).

H2: Participants who play Bad News rate social media content making use of the "conspiracy" technique as significantly less reliable in the posttest as compared with the pretest.

H3: Participants are more likely to choose to spread misinformation about an ideologically incongruent news topic than an ideologically congruent one within Bad News (i.e., left-wing participants are more likely to choose large corporations or police brutality, and right-wing participants are more likely to choose the government and rising crime rates).

H0: Participants who play Bad News do not rate non-misinformation ("real news") as significantly less reliable in the posttest as compared with the pretest.

Method, Sample, and Procedure

After a brief introduction to familiarize players with the game mechanics, participants (n = 472; 44.9% male, 66.1% between 18 and 29 years of age; slightly left-leaning politically, M = 3.70, SD = 1.65 on a 7-point scale; see Supplement S1 and Table S1 for the full sample composition) were asked if they wanted to participate in a scientific study. If they agreed, they were asked to provide informed consent to have their responses recorded. All responses were recorded anonymously, and no personally identifying data were stored.

Participants then performed an item-rating task as the pretest, consisting of nine social media posts. Four of these posts made use of polarizing language and were used in previous research on *Bad News* (Basol et al., 2020; Maertens et al., 2021). We also included two social media posts that made use of the

"conspiracy" technique, also used in previous *Bad News* research. Finally, we included three "real news" posts that did not contain any misinformation. Participants were asked to rate the reliability of each item on a scale from 1 (*very unreliable*) to 7 (*very reliable*). Figure 2 shows an example of a survey item within the game environment.

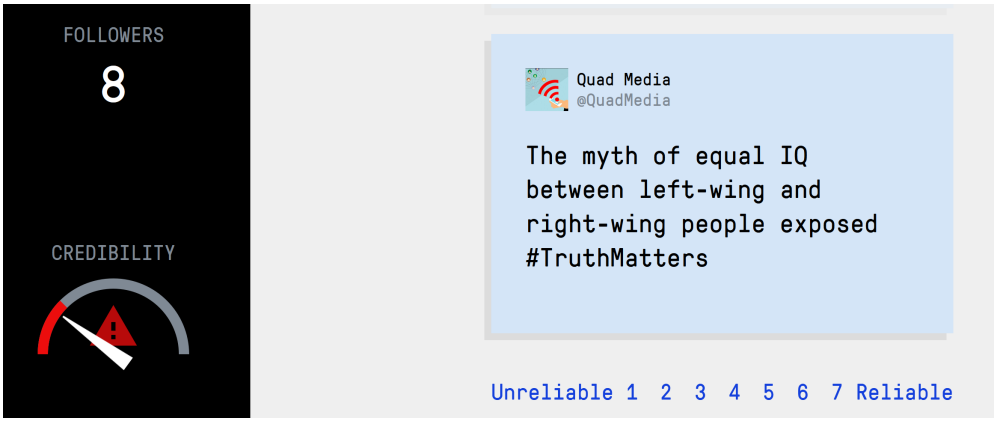


Figure 2. Example of a survey item implemented in Bad News (https://www.getbadnews.com/), with the "followers" and "credibility" meters on the left.

After answering these questions, participants proceeded through the game as normal. At the end of the game, they were asked to rate the same eight items from the pretest again (the posttest). Participants also answered a series of multiple-choice demographic and other questions: age group (participants under 18 were excluded as per our ethics approval); gender; political ideology, from 1 (*very left-wing*) to 7 (*very right-wing*); education level; whether they had played *Bad News* before; social media use, from 1 (*never*) to 5 (*daily*); and the "ball and bat" question from the Cognitive Reflection Test (CRT; Frederick, 2005). Finally, we also recorded participants' in-game choice of the news topic to spread misinformation about the polarization scenario (one of four possible options: large corporations, police brutality, rising crime rates, or the government). Figure 3 shows an overview of all three studies' design in a flowchart; in contrast to the within-subject design of Study 1, Studies 2 and 3 used a randomized design versus a control group to assign people to the game.

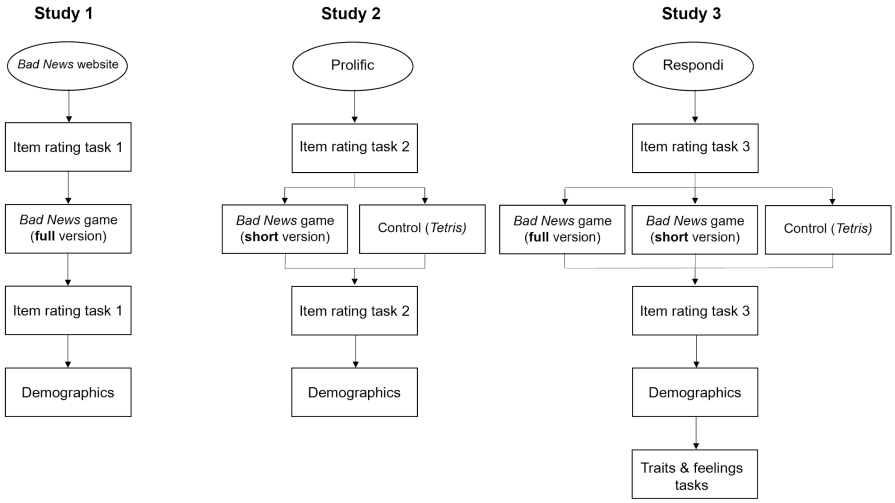


Figure 3. Study design flowchart.

Results

The results of the paired-samples *t* test comparing the pre- and post reliability scores for the social media posts from the in-game item-rating task (H1, H2, and H0) are shown in Table 1. See Table S2 for item-level statistics.

Table 1. Paired-Samples t tests for Pre–Post Reliability Ratings of Polarizing, Conspiratorial, and Real News Social Media Content.

Variables		*t*	*df*	*p*	M_{pre}	M_{post}	M_{diff}	*95% CI*	*Cohen's d*
Hypotheses									
H1: Polarization (Pre)	Polarization (Post)	4.403	471	**< .001**	2.94	2.58	0.36	[0.20, 0.52]	0.20
H2: Conspiracy (Pre)	Conspiracy (Post)	3.746	471	**< .001**	2.88	2.55	0.33	[0.16, 0.51]	0.17
H0: Real News (Pre)	Real News (Post)	0.510	471	0.610	5.21	5.17	0.04	[−0.13, 0.22]	0.02
Exploratory variables									
Misinformation (Pre)	Misinformation (Post)	4.580	471	**< .001**	2.92	2.57	0.35	[0.20, 0.50]	0.21
Discernment (Pre)	Discernment (Post)	−3.013	471	**0.003**	2.29	2.60	−0.31	[−0.50, −0.11]	−0.14

Note. The table shows paired-samples *t* tests for the four polarization items ("polarization"), the two conspiracy items ("conspiracy"), the three non-misinformation items ("real news"), all six misinformation items ("misinformation"), and discernment ("real news" minus "misinformation"), before (pre) and after (post) playing *Bad News*.

We found that participants rated polarizing social media content as significantly less reliable after playing ($p < .001$, $d = .20$), in support of H1. In addition, participants also found conspiratorial content significantly less reliable post-gameplay ($p < .001$, $d = .17$), supporting H2. Finally, we found that participants did not rate real news as significantly more or less reliable after playing, as compared with before ($p = .510$,

d = .02), with a TOST equivalence test with a smallest effect size of interest (SESOI) of *d* = ± .10 and α = .05, confirming statistical equivalence to 0, *t*(471) = −1.66, *p* = .049. These results support H0.

Overall, we found that participants' "truth discernment"—that is, the ability to distinguish misinformation (i.e., the average of all misinformation items, conspiracy + polarization) from non-misinformation (the average of the three real news items)—improved significantly post-gameplay (*p* < .003, *d* = .14). The same effects were observed for all six individual misinformation items used in this study (all *p* values < .019), indicating that the observed inoculation effect was not due to a reduction in perceived reliability for a few items, but for all items in the survey; see Table S4. The results are visualized in Figure 4.

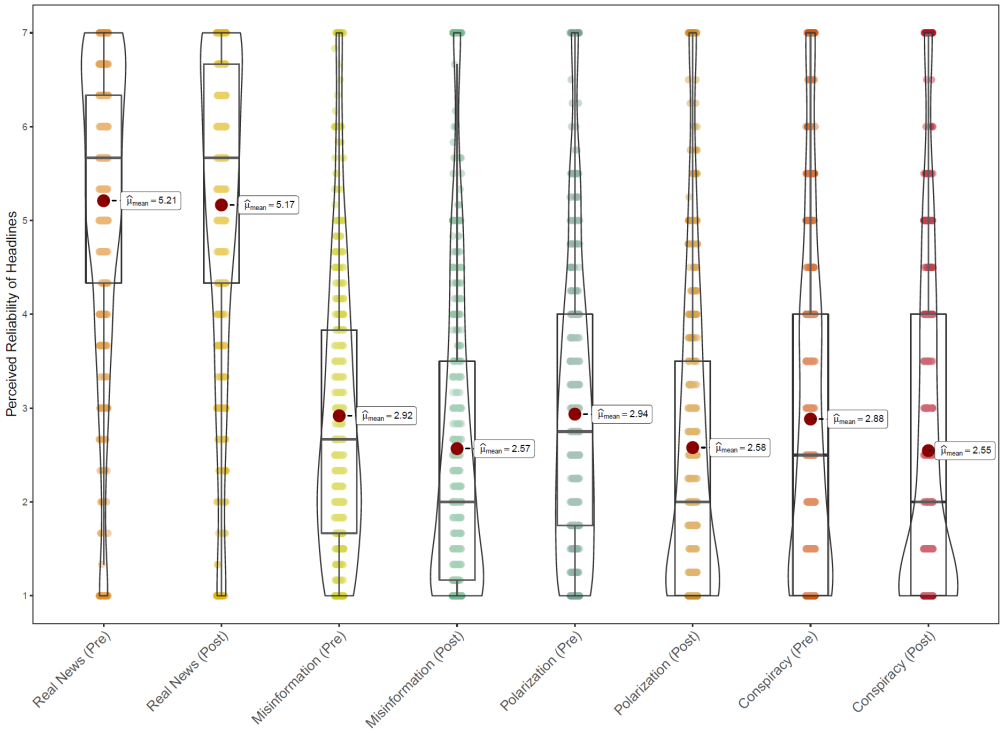


Figure 4. Violin plots (with boxplots and data jitter) for the pre- and postgame reliability scores of real news/non-misinformation: two leftmost plots, t(471) = .510, p = .610, d = .02; misinformation, all six misinformation items combined, t(471) = 4.580, p < .001, d = .21; polarizing content, four polarization items combined, t(471) = 4.403, p < .001, d = .20; and conspiratorial content, two conspiracy items combined, t(471) = 3.746, p < .001, d = .17.

To test H3, we conducted a logistic regression with the in-game choice of news topic that participants spread misinformation about (0 = a predominantly left-wing topic, that is, large corporations or police brutality, and 1 = a predominantly right-wing topic, that is, rising crime rates or the government) as the dependent variable, and political ideology (1 = left-wing, 7 = right-wing), age, gender, education level, social media use, CRT performance, and whether participants had played Bad News before as independent variables. We found that when controlling for all covariates, political ideology was not a

significant predictor of in-game choices (p = .767): neither left-wing nor right-wing participants chose to spread misinformation about a topic that was incongruent with their political ideology, contradicting H3. On its own (without other covariates), political ideology was also not a significant predictor of news topic choice (p = .576). See Table S5.

Finally, to check whether our covariates (age, gender, education level, political ideology, social media use, whether participants played Bad News before, and CRT performance) influenced the inoculation effect, we conducted a linear regression with the pre–post difference in perceived reliability of the polarization items as the dependent variable, and the mentioned covariates as independent variables. We found no significant effects of the covariates on the pre–post difference score (all p values > .089; see Table S6), indicating that the inoculation effect was robust when controlling for these variables.

Discussion

We found that playing *Bad News* significantly reduced the perceived reliability of polarizing social media content, while participants' perception of non-misinformation (real news) did not change. We found no evidence that game players' political ideology influenced their in-game behavior.

There are several limitations to this study. First, we did not include a control group. Second, we relied on a convenience sample (participants were people who voluntarily navigated to the *Bad News* website). Third, participants were inoculated not only against the polarization technique, but also against other techniques, so we were unable to test whether playing *Bad News* conferred cross-protection against untreated misinformation techniques. We addressed these limitations in Study 2.

Study 2

Following Basol et al. (2020), we conducted a 2 (pre–post) × 2 (control–treatment) mixed-between preregistered randomized controlled trial on the online recruitment platform Prolific Academic, with two conditions: a treatment condition (in which participants played the shortened version of *Bad News*, featuring only the polarization scenario) and a control condition (in which participants played *Tetris*). We preregistered the following hypotheses (see https://aspredicted.org/s5n7c.pdf)[2]:

H1: Participants playing [a shortened version of] Bad News are more likely to choose to spread misinformation [about topics that are] ideologically incongruent with their political beliefs than misinformation [about] ideologically congruent topics.

H2: Participants who play [a shortened version of] Bad News rate misinformation making use of the polarization technique as significantly less reliable post-gameplay, compared with a control group.

[2] We slightly rephrased these hypotheses to improve their clarity. Rephrased words are marked in [brackets].

H3: *Both left-wing and right-wing participants who play [a shortened version of] Bad News rate left-leaning and right-leaning headlines containing misinformation as significantly less reliable post-gameplay, as compared with a control group.*

H4: *Participants who play [a shortened version of] Bad News rate headlines using the "impersonation" technique as significantly less reliable post-gameplay, as compared with a control group.*

Method, Sample, and Procedure

At the start of the study, participants (U.S. residents; n = 193; 110 control, 83 treatment; 72.5% female, M_{Age} = 26.0, SD_{Age} = 9.09; left-leaning politically, M = 3.09, SD = 1.62 on a 7-point scale; see Supplement S1 and Table S1) performed an item-rating task similar, but not identical, to the item-rating task from Study 1, with 10 items (social media posts) instead of nine; see Table S2 for item-level statistics. We included two additional polarization items in addition to the four used in Study 1. To examine cross-protection, we included two items that made use of the impersonation technique instead of the two conspiracy items from Study 1. Finally, we again included two real news posts that did not contain any misinformation. As in Study 1, participants were asked to rate the reliability of each item on a scale from 1 (*very unreliable*) to 7 (*very reliable*; the pretest).

Next, participants were randomly assigned to play either the shortened *Bad News* game or *Tetris* for approximately the same amount of time. Participants in the *Bad News* condition were required to provide a password (which they could obtain at the end of the game) before proceeding with the rest of the study. After the game, participants rated the same items from the item-rating task (the posttest). Finally, participants were asked a series of demographic questions: age; gender; education level; political ideology, from 1(*very left-wing*) to 7 (*very right-wing*); the "ball-and-bat" question from the CRT (Frederick, 2005); and social media use and Twitter use, both from 1 (*never*) to 5 (*daily*). Finally, to examine whether participants' political ideology influenced their in-game choices (partisan self-selection), we asked participants whether they chose to spread fake news about a predominantly left-wing news topic (large corporations or police brutality) or a predominantly right-wing news topic (the government or rising crime rates); unlike in Study 1, this was a self-reported and not a behavioral variable.

Results

We did not find any support for H1, H2, H3, and H4 (all p values > .07). Because of space limitations, we report our full findings for this study (as well as our exploratory analyses) in Supplement S2. See Table S2 for item-level statistics.

Discussion

In this preregistered experiment, we failed to find support for our hypotheses: Participants' political ideology was not associated with their self-reported in-game choices. Like in Study 1, we thus found no support for political attitudes influencing how people made choices within the intervention. Furthermore, although treatment group participants found polarizing social media content descriptively less reliable than a control group, this difference was not significant (p = .07). There were also no meaningful differences between left-wing and right-wing participants in terms of the inoculation effect. Finally, we found no evidence of cross-protection (Parker et al., 2016), as treatment group participants did not become significantly better at recognizing content making use of the impersonation technique.

However, although we did not find support for our hypotheses, equivalence tests mostly failed to rule out the presence of meaningful effects. It is therefore possible that our sample size was too low to detect these effects.[3] This idea is further buttressed by the significant results from Study 1 and the fact that other mixed-design studies using *Bad News* found smaller than average effects for the polarization items compared with the other misinformation techniques (Basol et al., 2020; Maertens et al., 2021).

Another interpretation of our findings is that our sample was very young (M_{age} = 26.0), female (72.5%), and politically left-leaning (M = 3.09 on a 7-point scale), in comparison with Studies 1 and 3 (see Supplement S1). Previous research has found that younger people and, in the United States, left-wing individuals are comparatively good at identifying manipulative content online (Guess, Nagler, & Tucker, 2019; Rathje et al., 2022; Roozenbeek et al., 2022), and gender may play a small role in the inoculation effect (Roozenbeek & van der Linden, 2019). The observed effects may therefore be smaller than if the sample had been more balanced. Finally, the impersonation technique is conceptually quite different from the polarization technique: The former plays into people's (lack of) ability to spot manipulated text, whereas the latter makes use of divisive language to fuel intergroup tensions. It is possible that cross-protection does not apply to misinformation techniques that are too dissimilar from the one that participants were inoculated against, but it may nonetheless be observed for techniques that are conceptually closer. We address these possibilities in Study 3.

Study 3

We conducted a preregistered 2 (pre–post) × 3 (control–treatment 1–treatment 2) mixed randomized controlled trial on the ISO-certified platform Respondi, with three conditions: two treatment conditions (in which participants played the shortened version of *Bad News* featuring only the polarization scenario, or the full version of the game) and a control condition (in which participants played *Tetris*). We preregistered the following hypotheses (see https://aspredicted.org/s9fq9.pdf):

[3] A sensitivity analysis with 95% power and α = .05 shows that this study was sufficiently powered to detect effects with a Cohen's d of 0.53. With 80% power (α = .05), the study was powered for an effect size of d = 0.41. Because our obtained effect size was approximately d = 0.29, it appears likely that a lack of statistical power played a role in the observed absence of significant effects.

H1: *Participants playing [a shortened version of] Bad News are more likely to choose to spread misinformation [about topics that are] ideologically incongruent with their political beliefs than misinformation [about] ideologically congruent topics.*

H2: *Participants who play both the long and short version of Bad News rate misinformation making use of the "polarization" technique as significantly less reliable, are significantly more confident in their assessment, and are significantly less likely to be willing to share such misinformation post-gameplay as compared with a control group.*

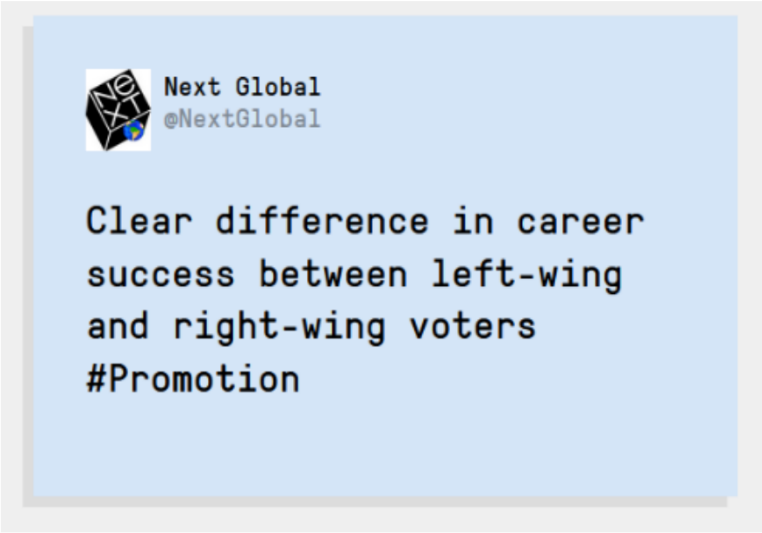
H3: *Both left-wing and right-wing participants who play both the long and short version of Bad News rate left-leaning and right-leaning headlines containing misinformation as significantly less reliable post-gameplay, are significantly more confident in their assessment, and are significantly less likely to be willing to share such misinformation as compared with a control group.*

H4: *Participants who play both the long and short version of Bad News rate misinformation as significantly less reliable post-gameplay, are significantly more confident in their assessment, and are significantly less likely to be willing to share such misinformation as compared with a control group.*

H0: *Participants who play both the long and short version of Bad News rate non-misinformation as equally reliable post-gameplay, are [not] significantly more confident in their assessment, and are [not] significantly less likely to be willing to share such misinformation as compared with a control group.*

Method, Sample, and Procedure

At the start of the study, participants (U.S. residents; n = 772; 319 control, 203 for the full *Bad News* condition, 256 for the short *Bad News* condition; 63.6% female, 56.3% over 45 years of age; balanced politically, M = 4.04, SD = 1.70 on a 7-point scale; see Supplement S1 and Table S1) performed an item-rating task with 18 social media posts displayed in a random order. In addition to the 10 social media posts from Study 2 (six polarization items, two impersonation items, and two real news items), participants also rated two items that made use of the "emotional language" technique, two that used conspiratorial reasoning, two that used ad hominem attacks (discrediting), and two posts that contained trolling (van der Linden & Roozenbeek, 2020). Participants were asked to rate the reliability of each item on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*) in response to the statement, "This post is reliable." In addition, we included two measures that have been used in past research to test the effectiveness of anti-misinformation interventions: confidence in their ability to identify misinformation (Basol et al., 2020), and willingness to share (mis)information with other people in their network (Basol et al., 2021; Pennycook et al., 2021). Both measures were assessed on the same 1–7 scale as the reliability measure (the pretest). See Figure 5. See Tables S2 and S3 for item-level statistics.



	Strongly disagree			Neutral			Strongly agree
	1	2	3	4	5	6	7
This post is reliable	○	○	○	○	○	○	○
I am confident in my assessment of this post's reliability	○	○	○	○	○	○	○
I would share this post with people in my network	○	○	○	○	○	○	○

Figure 5. Example of an item from the item-rating tasks from Study 3 (https://www.getbadnews.com/).

Next, participants were randomly assigned to play either the shortened *Bad News* game, the full *Bad News* game, or *Tetris*. Participants in both *Bad News* conditions were required to provide a password (obtained at the end of the game) before proceeding with the rest of the study. After the game, participants rated the same items from the item-rating task again (the posttest). Participants were then asked a series of other questions: age group; gender; education level; political ideology, from 1 (*very left-wing*) to 7 (*very right-wing*); political party affiliation (Democrat/Republican/Independent/Other); the "ball-and-bat" question from the CRT (Frederick, 2005); social media use, from 1 (*never*) to 5 (*daily*); and Twitter use, from 1 (*never*) to 5 (*daily*). Participants who played the shortened *Bad News* game were asked the same question about what news topic they chose to spread misinformation about from Study 2.

Finally, to study the games' effect on affective polarization (an exploratory analysis), participants who identified as Democrats or Republicans were asked two sets of questions about the opposing party. These questions were taken from Druckman and Levendusky (2019) and included the "feeling thermometer" (how people feel toward the opposing party, with 0 = *most unfavorable/cold* and 100 = *most*

favorable/warm), and a series of questions asking respondents to rate how well four positive and three negative traits (e.g., patriotic, honest, generous, hypocritical, selfish) applied to the opposing party.

Results

Next, we report the results for H1–H4 and H0 in this order. See Supplement S3 for our exploratory analyses, as well as Tables S2 and S3 for item-level statistics.

In-Game Choices and Congruence With Political Beliefs

To test H1, we conducted a logistic regression with political ideology predicting the choice of news type that people would choose to spread misinformation about in the shortened *Bad News* game (see Studies 1 and 2). We found a significant effect of political ideology on the type of news that people reported engaging with in the game (*OR* = 1.60, 95% CI [1.34, 1.94], $p < .001$), so that left-wing participants engaged more with left-wing topics, and vice versa. This effect was robust when controlling for age, gender, education level, social media use, and Twitter use. We thus found support for H1. See Figure 6 and Table S8.

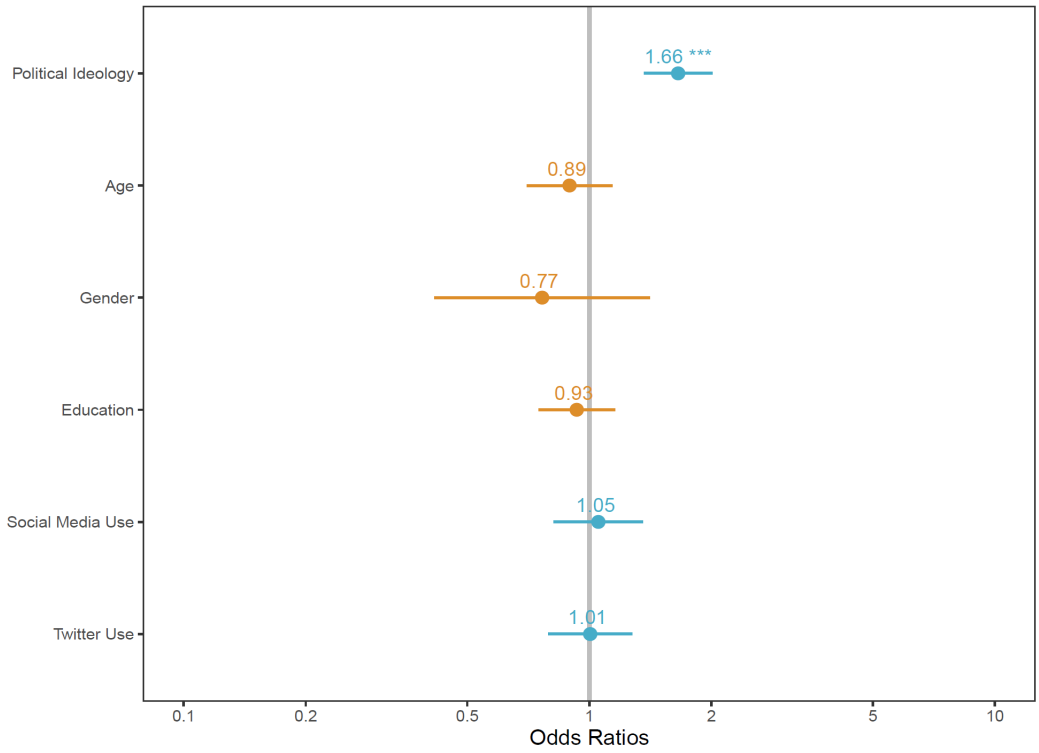


Figure 6. Odds ratios for political ideology, age, gender, education, social media use, and Twitter use, predicting the type of news that participants would choose to spread in the shortened Bad News game (0 = predominantly left-leaning topics; 1 = predominantly right-leaning topics). See also Table S8.

Perceived Reliability, Confidence, and Sharing Willingness of Polarizing Social Media Content

To test H2, we conducted a series of one-way Welch's analyses of variance (ANOVAs) on the averaged pre–post difference scores in the perceived reliability of items making use of the polarization technique, as well as people's attitudinal certainty (confidence) and willingness to share scores, by condition.[4]

Reliability

For the reliability measure, we found a significant effect of condition on the difference in perceived reliability of polarizing social media content, $F(2, 433.43) = 5.81$, $p = .003$, $\eta^2 = .01$. A Games-Howell post hoc test shows that the pre–post difference in perceived reliability was significantly higher in both the full *Bad News* condition compared with the control condition ($M = -.28$ vs. $M = -.08$, $M_{diff} = .20$, 95% CI [.03, .36], $p = .014$, $d = .26$) and the short *Bad News* condition compared with the control condition ($M = -.23$ vs. $M = -.08$, $M_{diff} = .15$, 95% CI [.01, .29], $p = .027$, $d = .22$). However, the difference between the full and short versions of *Bad News* is not significant ($M = -.28$ vs. $M = -.23$, $M_{diff} = .05$, 95% CI [−.14, .23], $p = .828$). Figure 7 shows the results.[5]

[4] Bartlett's test is significant for the reliability (Bartlett's $k^2 = 41.993$, $p < .001$), confidence (Bartlett's $k^2 = 15.307$, $p = .004$), and willingness to share (Bartlett's $k^2 = 33.552$, $p < .001$) measures, indicating that the assumption of equal variances is violated. We therefore report Welch's instead of Fisher's ANOVAs.

[5] As a robustness check, see Figure S1 for the same figure with Bonferroni-corrected p values.

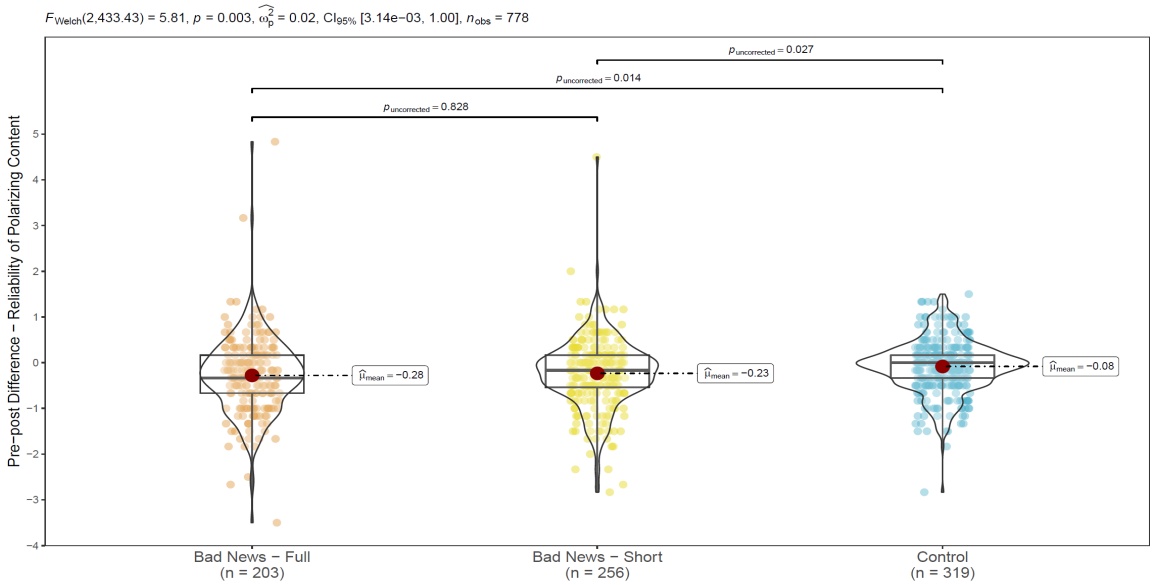


Figure 7. Violin plots (with boxplot and data jitter) for the averaged pre–post difference in the perceived reliability of polarizing social media content, for the full and shortened versions of Bad News and the control group (Tetris). Large black dots indicate the mean pre–post difference scores. Smaller colored dots show the data jitter. See also Figure S1.

Confidence

For the confidence measure, we found no significant effect of condition on the difference in participants' confidence in their assessment before and after gameplay, $F(2, 455.28) = .87$, $p = .422$. Furthermore, TOST equivalence tests confirmed statistical equivalence to 0 between the control condition and both the full *Bad News* condition, $t(370.68) = 1.96$, $p = .025$ and the short *Bad News* condition, $t(485.08) = 3.18$, $p < .001$.

Sharing

For the willingness to share measure, we found no significant effect of condition on the difference in sharing willingness before and after gameplay, $F(2, 438.75) = .14$, $p = .869$. TOST equivalence tests further confirmed statistical equivalence to zero—full version–control: $t(328.33) = 2.78$, $p = .003$; short version–control: $t(476.62) = 3.13$, $p < .001$.

We thus found partial support for H2: While participants in both treatment groups found polarizing social media content significantly less reliable post-gameplay as compared with the control group, they were not significantly more confident in their assessment and were not significantly less willing to share polarizing content with others in their network.

Perceived Reliability of Polarizing Social Media Content Across the Political Spectrum

To test H3, we first conducted a series of two-way ANOVAs to determine the effect of condition and political ideology (1 = *very left-wing*, 7 = *very right-wing*) on the pre–post difference score of the perceived reliability of polarizing social media content, as well as participants' confidence and their willingness to share such content. We found no significant two-way interaction between political ideology and condition for the reliability, $F(2, 772) = 2.16$, $p = .116$; confidence, $F(2, 772) = 2.35$, $p = .062$; and sharing measures, $F(2, 772) = .04$, $p = .906$. As supplementary analyses, we conducted a series of one-way ANOVAs on the pre–post reliability, confidence, and willingness to share scores for left-wing and right-wing participants separately. We found that the reduction in perceived reliability of polarizing content appeared to be somewhat larger for right-wing compared with left-wing participants, although we note that the two-way interactions reported earlier constituted the most appropriate way to test for the hypothesized conditional effects (see Supplement S3 for more details). Overall, we found support for H3: There was no significant two-way interaction between political ideology and condition. However, it is possible that the reduction in the perceived reliability of polarizing content was larger for right-wing than left-wing participants.

Cross-Protection

To test H4, we conducted a one-way Welch's ANOVA on the averaged pre–post difference score in the perceived reliability of misinformation items that did *not* make use of the polarization technique (but used a different technique instead), by condition (treatment–control), and did the same for the confidence and sharing measures.[6]

For the reliability measure, we found a significant effect of condition on the difference in perceived reliability of misinformation, $F(2, 426.13) = 23.69$, $p < .001$, $\eta^2 = .06$. A Games-Howell post hoc test shows that the pre–post difference in perceived reliability was significantly higher in both the full *Bad News* condition compared with the control condition ($M = -.38$ vs. $M = .01$, $M_{diff} = .40$, 95% CI [.25, .54], $p < .001$, $d = .62$) and the short *Bad News* condition compared with the control condition ($M = -.16$ vs. $M = .01$, $M_{diff} = .17$, 95% CI [.06, .28], $p = .001$, $d = .32$). In addition, the pre–post difference in reliability is significantly higher in the full *Bad News* condition than the short *Bad News* condition ($M = -.38$ vs. $M = -.16$, $M_{diff} = .23$, 95% CI [.07, .38], $p = .002$, $d = .33$). Figure 8 shows the results.[7]

[6] Bartlett's test is significant for the reliability (Bartlett's $k^2 = 64.068$, $p < .001$), confidence (Bartlett's $k^2 = 9.453$, $p = .009$), and willingness to share (Bartlett's $k^2 = 27.539$, $p < .001$) measures, indicating that the assumption of equal variances is violated.

[7] As a robustness check, see Figure S2 for the same figure with Bonferroni-corrected p values.

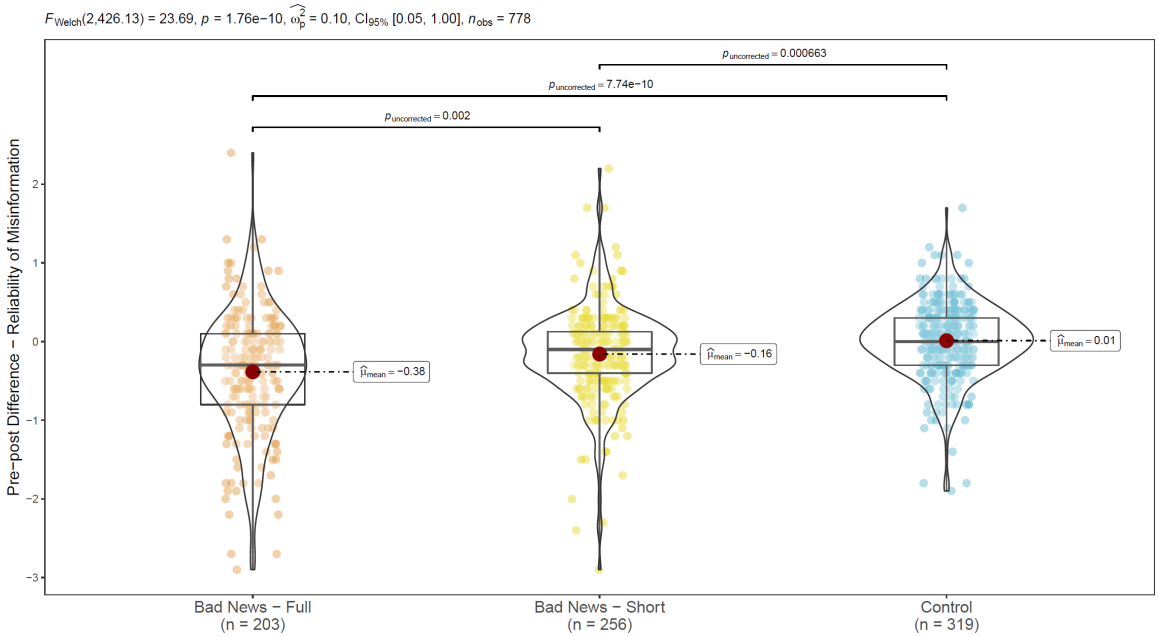


Figure 8. Violin plots (with boxplot and data jitter) for the averaged pre–post difference in the perceived reliability of misinformation (not including test items using the "polarization" technique), for the full and shortened versions of Bad News and the control group (Tetris). Large black dots indicate the mean pre–post difference scores. Smaller colored dots show the data jitter. See also Figure S2.

For the confidence, $F(2, 459.97) = .81$, $p = .447$, and sharing measures, $F(2, 442.85) = 1.26$, $p = .284$, we found no significant effects of condition on the pre–post difference scores. TOST equivalence tests confirmed statistical equivalence to 0 for the confidence measure, full version–control: $t(378.55) = -2.74$, $p = .003$, short version–control: $t(503.14) = 2.64$, $p = .004$; and for the sharing measure, full version–control: $t(339.12) = 2.37$, $p = .009$, short version–control: $t(478.56) = -2.51$, $p = .006$.

We thus found partial support for H4: Participants who played the short version of *Bad News* (which only covers the polarization technique) found misinformation that did *not* make use of this technique to be significantly less reliable post-gameplay as compared with a control group. However, this effect was significantly weaker than for participants who played the full version of the game, indicating partial but not full cross-protection against misinformation that participants were not inoculated against. In addition, we found no significant differences for the confidence and sharing measures. See Tables S2 and S3 for the item-level results.

Effects of the Inoculation Treatment on Perceptions of "Real News"

To test H0, we conducted a one-way Welch's ANOVA on the averaged pre–post difference score in the perceived reliability of non-misinformation items, by condition (treatment–control), and did the same for the confidence and sharing measures.[8]

For the reliability measure, we found a significant effect of condition on the difference in perceived reliability of non-misinformation, $F(2, 454.48) = 3.33$, $p = .037$, $\eta^2 = .01$. A Games-Howell post hoc test shows that the pre–post difference in perceived reliability was significantly higher in the full *Bad News* condition compared with the control condition ($M = -.29$ vs. $M = -.07$, $M_{diff} = .22$, 95% CI [.02, .42], $p = .028$, $d = .24$). However, the difference between the short version of *Bad News* and the control group was not significant ($M = -.12$ vs. $M = -.07$, $M_{diff} = .05$, 95% CI [−.12, .22], $p = .745$).

For the confidence measure, $F(2, 464.65) = .93$, $p = .395$, and the sharing measure, $F(2, 458.19) = 1.67$, $p = .190$, we found no significant effects of condition on the pre–post difference scores. TOST equivalence tests confirmed statistical equivalence to 0 for the confidence measure, full version–control: $t(383.82) = 2.46$, $p = .007$, short version–control: $t(561.49) = 2.27$, $p = .012$; and for the sharing measure, full version–control: $t(374.65) = 1.66$, $p = .049$, short version–control: $t(498.64) = -3.28$, $p < .001$.

We thus found partial support for H0: Participants who played the short version of *Bad News* had no significantly different perceptions of non-misinformation post-gameplay compared with a control group. However, participants who played the full version of the game rated non-misinformation as significantly less reliable post-gameplay compared with the control group, although they were not significantly more or less confident or willing to share such content with others.

Discussion

In this second preregistered randomized controlled experiment, unlike in Studies 1 and 2, we found that political ideology was a significant predictor of the choices that people reported to make in the *Bad News* game. In line with our hypothesis, we found that left-wing participants tended to choose to spread misinformation about topics relating to left-wing out-groups (i.e., large corporations and police brutality), and right-wing participants tended to do so about right-wing out-groups (the government and rising crime rates).

Furthermore, playing both the long and short *Bad News* games significantly decreased the perceived reliability of polarizing social media content. Although the two-way interaction between political ideology and reliability was not significant, there was some indication that right-wing participants showed a higher post-gameplay reduction in the perceived reliability of polarizing content than left-wing participants (see Supplement S3).

[8] Bartlett's test is significant for the reliability (Bartlett's $k^2 = 13.984$, $p < .001$), confidence (Bartlett's $k^2 = 12.024$, $p = .002$), and willingness to share (Bartlett's $k^2 = 11.124$, $p = .004$) measures, indicating that the assumption of equal variances is violated.

Unlike in previous studies (Basol et al., 2021; Roozenbeek, van der Linden, & Nygren, 2020), we found no significant pre–post differences for our confidence and willingness to share measures. An explanation for this finding may have to do with players' in-game choices: If left-wing players choose to focus on left-wing topics, and right-wing players on right-wing topics, partisans may create their own "echo chamber" within the game. This lack of exposure to polarizing content from the opposite side of the political spectrum may limit the inoculation effect. However, this does not explain why we did find significant results for the reliability measure. We therefore encourage further research to explore these findings in more detail.

We also found partial support for a cross-protection effect (Parker et al., 2016): Participants who played the shortened version of *Bad News*, which only contained the polarization scenario, found social media content that made use of the emotion, conspiracy, trolling, or discrediting techniques to be significantly less reliable post-gameplay compared with a control group, although this effect was substantially smaller than for the full *Bad News* game (d = 0.32 vs. d = 0.62). This indicates that it is possible to (partially) inoculate people against untreated attacks on their beliefs (Compton et al., 2021).

Finally, in line with previous research (Basol et al., 2021; Guess et al., 2020), we found that anti-misinformation interventions may impact people's perceptions of non-misinformation (real news) as well as misinformation. This effect, however, appears to vary between studies and items (and, notably, was not observed in Study 1).

General Discussion

In this article, we explored three research questions: (1) whether inoculation interventions improve people's ability to recognize polarizing social media content (Kubin & von Sikorski, 2021); (2) whether inoculating against one misinformation technique confers psychological resistance against misinformation that makes use of other techniques (Parker et al., 2016); and (3) whether people's political ideology influences how people engage with anti-misinformation interventions (Roozenbeek et al., 2022; Van Bavel et al., 2021), as measured by their (self-reported) behavior in a gamified inoculation intervention.

In Studies 1 and 3, we found that playing *Bad News* reduced the perceived reliability of polarizing social media content. In Study 2, which used a shortened version of the game, this effect was not observed, although it was close to significant (p = .07). Our findings show that it is feasible to improve people's ability to recognize polarizing social media content through inoculation. We found no evidence for "side effects," in the sense that playing *Bad News* did not inadvertently increase affective polarization (see Supplement S3). However, we note that, unlike in previous studies on gamified inoculation interventions (Basol et al., 2021; Roozenbeek & van der Linden, 2020), we did not find that people's confidence in their ability to spot polarizing content improved post-gameplay, and their willingness to share such content was not significantly reduced either.

In Study 3, we found some indication that the reduction in the perceived reliability of polarizing content was most present for right-wing participants, although the two-way interaction between political ideology and perceived reliability was not significant, and so we do not make strong statements about whether there are differences between left-wing and right-wing participants in this respect. These findings

suggest that polarization may be a particularly pernicious tactic to address. Nonetheless, our results show that playing *Bad News* is effective in conferring psychological resistance against polarizing content.

With respect to cross-protection (Parker et al., 2016), we found no evidence for our hypothesis (H4) in Study 2. In Study 3, participants who played the short version of *Bad News* (covering only the polarization technique) found misinformation that did *not* evoke polarization to be significantly less reliable post-gameplay compared with a control group, indicating initial support for cross-protection. However, this inoculation effect was weaker than for participants who played the full version of the game, indicating partial but not full cross-protection.

Our findings are somewhat ambiguous as to how inoculation affects people's perception of real news. When playing the short version of the *Bad News* game, participants in Study 3 did not significantly differ in their perceptions of non-misinformation (real news) post-gameplay compared with a control group. Participants who played the full version of *Bad News* rated non-misinformation as significantly less reliable post-gameplay compared with the control group, indicating heightened skepticism of all information rather than exclusively of misinformation. In Study 1, conversely, we did not find meaningful pre–post differences in perceived reliability for real news. Similar findings were reported by Basol et al. (2021), although in that study this "heightened skepticism" of real news dissipated one week after the intervention. These contradictory findings may be an artifact of the fact that Study 3 was conducted online via a survey, whereas data for Study 1 were collected within *Bad News* via an in-game survey. Further research is needed to explore whether this relatively small "heightened skepticism" effect is related to the items, study design, or a common "side effect" of anti-misinformation interventions (Guess et al., 2020).

Studies 1 and 2 showed no significant effect of political ideology on the news topic that people chose to engage with within the game. However, Study 3 showed that both left-wing and right-wing players chose to engage with politically congruent news topics. This is in line with expectations of out-group animosity: Left-wing game players may choose to attack a group with misinformation that they consider to be part of an out-group, such as large corporations, whereas right-wing players did the same (e.g., attacking the government). However, in Study 1, we measured behavioral data (tracking people's actual, as opposed to self-reported, in-game choices), whereas we were only able to record self-reported data in Studies 2 and 3. Because patterns in Studies 1, 2, and 3 diverged, we do not make strong predictions about sharing habits here. We acknowledge that although people were randomly assigned to the treatment or control group, the game scenarios themselves allowed for a degree of self-selection, so the treatment was not held perfectly constant across individuals (Trilling, Van Klingeren, & Tsfati, 2017).

Conclusion

Across three studies, we found that technique-based inoculation can successfully reduce the perceived reliability of polarizing social media content, although attitudinal certainty and sharing intentions are not affected. We also found support for partial "cross-protection" (increased psychological resilience against manipulation techniques that people were not inoculated against). We found preliminary evidence that political ideology plays a role in how people engage with anti-misinformation interventions, influencing

the choices they make within an inoculation game. However, these choices may not reflect the sharing decisions that people make in their own social networks.

References

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review, 110*(3), 629–676. doi:10.1257/aer.20190658

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., . . . Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences, 115*(37), 9216–9221. doi:10.1073/pnas.1804840115

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

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). doi:10.5334/joc.91

Compton, J., Van der Linden, S., Cook, J., & Basol, M. (2021). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Social and Personality Psychology Compass, 15*(6), e12602. doi:10.1111/spc3.12602

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE, 12*(5), 1–21. doi:10.1371/journal.pone.0175799

Druckman, J. N., & Levendusky, M. S. (2019). What do we measure when we measure affective polarization? *Public Opinion Quarterly, 83*(1), 114–122. doi:10.1093/poq/nfz003

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42. doi:10.1257/089533005775196732

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. doi:10.1073/pnas.1920498117

Guess, A. M., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances, 5*(1). doi:10.1126/sciadv.aau4586

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. doi:10.1111/jasp.12453

Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication, 37*(2), 256–280. doi:10.1080/10584609.2019.1661888

Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association, 45*(3), 188–206. doi:10.1080/23808985.2021.1976070

Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied, 27*(1), 1–16. doi:10.1037/xap0000315

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

Parker, K. A., Rains, S. A., & Ivanov, B. (2016). Examining the "blanket of protection" conferred by inoculation: The effects of inoculation messages on the cross-protection of related attitudes. *Communication Monographs, 83*(1), 49–68. doi:10.1080/03637751.2015.1030681

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature, 592*, 590–595. doi:10.1038/s41586-021-03344-2

Pew Research Center. (2019, December 17). *In a politically polarized era, sharp divides in both partisan coalitions*. Retrieved from https://www.pewresearch.org/politics/2019/12/17/in-a-politically-polarized-era-sharp-divides-in-both-partisan-coalitions/

Pew Research Center. (2020, July 14). *As the U.S. copes with multiple crises, partisans disagree sharply on severity of problems facing the nation*. Retrieved from https://www.pewresearch.org/fact-tank/2020/07/14/as-the-u-s-copes-with-multiple-crises-partisans-disagree-sharply-on-severity-of-problems-facing-the-nation/

Pew Research Center. (2021, May 17). *Public trust in government: 1958–2021*. Retrieved from https://www.pewresearch.org/politics/2021/05/17/public-trust-in-government-1958-2021/

Rathje, S., Roozenbeek, J., Traberg, C. S., Van Bavel, J. J., & van der Linden, S. (2022). Letter to the editors of *Psychological Science*: Meta-analysis reveals that accuracy nudges have little to no effect for U.S. conservatives: Regarding Pennycook et al. (2020). *Psychological Science*. doi:10.25384/SAGE.12594110.v2

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Outgroup animosity drives engagement on social media. *Proceedings of the National Academy of Sciences, 118*(26), e2024292118. doi:10.1073/pnas.2024292118

Roozenbeek, J., Maertens, R., Herzog, S., Geers, M., Kurvers, R., Sultan, M., & van der Linden, S. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgment and Decision Making, 17*(3). Advance online publication.

Roozenbeek, J., Traberg, C. S., & van der Linden, S. (2022). Technique-based inoculation against real-world misinformation. *Royal Society Open Science, 9*(211719). doi:10.1098/rsos.211719

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

Roozenbeek, J., & van der Linden, S. (2018). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research, 22*(5), 570–580. doi:10.1080/13669877.2018.1443491

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Humanities and Social Sciences Communications, 5*(65), 1–10. doi: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 (HKS) Misinformation Review, 1*(8). doi:10.37016/mr-2020-47

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

Simchon, A., Brady, W. J., & Van Bavel, J. J. (2021). *Troll and divide: The language of online polarization*. PsyArxiv Preprints. doi:10.31234/osf.io/xjd64

Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. *The ANNALS of the American Academy of Political and Social Science, 700*(1), 136–151. doi:10.1177/00027162221087936

Trilling, D., Van Klingeren, M., & Tsfati, Y. (2017). Selective exposure, political polarization, and possible mediators: Evidence from the Netherlands. *International Journal of Public Opinion Research, 29*(2), 189–213. doi:10.1093/ijpor/edw003

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., . . . Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour, 4*(5), 460–471. doi:10.1038/s41562-020-0884-z

Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political psychology in the digital (mis)information age: A model of news belief and sharing. *Social Issues and Policy Review, 15*(1), 84–113. doi:10.1111/sipr.12077

van der Linden, S., & Roozenbeek, J. (2020). Psychological inoculation against fake news. In R. Greifeneder, M. Jaffé, E. Newman, & N. Schwarz (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation* (pp. 147–170). New York, NY: Psychology Press. doi:10.4324/9780429295379-11