

Research



Cite this article: Roozenbeek J, Traber C, van der Linden S. 2022 Technique-based inoculation against real-world misinformation. *R. Soc. Open Sci.* **9**: 211719.
<https://doi.org/10.1098/rsos.211719>

Received: 1 November 2021

Accepted: 19 April 2022

Subject Category:

Psychology and cognitive neuroscience

Subject Areas:

psychology

Keywords:

misinformation, inoculation theory, gamification, fake news, cross-protection

Author for correspondence:

Jon Roozenbeek

e-mail: jjr51@cam.ac.uk

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5978996>.

Technique-based inoculation against real-world misinformation

Jon Roozenbeek, Cecilie S. Traber and
Sander van der Linden

Department of Psychology, School of the Biological Sciences, University of Cambridge, Downing Street, CB2 3EB Cambridge, UK

JR, 0000-0002-8150-9305; CST, 0000-0002-5640-9273;
SvdL, 0000-0002-0269-1744

In recent years, numerous psychological interventions have been developed to reduce susceptibility to misinformation. Inoculation theory has become an increasingly common framework for reducing susceptibility to both individual examples of misinformation (issue-based inoculation) and to the techniques and strategies that are commonly used to mislead or misinform people (technique-based inoculation). In this study, we address two open questions related to technique-based inoculation in two separate experiments (total $n = 2188$; convenience sample recruited via the *Bad News* online game platform): (i) can technique-based inoculation effectively reduce susceptibility to real-world misinformation that went viral on social media? and (ii) can technique-based inoculation confer cross-protection against misinformation that does not make use of any of the techniques against which people were inoculated? We find that playing a 15 min game confers psychological resistance against real-world misinformation that makes use of manipulation techniques against which people were inoculated (Cohen's $d = 0.37$, Cohen's $U_3 = 64.4\%$, $p < 0.001$), and that cross-protection is achieved but at a reduced effect size ($d = 0.10$, $U_3 = 54.0\%$, $p = 0.001$).

1. Introduction

Online misinformation is a pernicious problem that has proven to be difficult to eradicate with the help of detection algorithms, legislation, and fact-checking initiatives [1,2]. During the COVID-19 pandemic, misinformation and conspiracy theories have proliferated widely online and have been linked to diminished willingness to follow health guidance measures and reduced intentions to get vaccinated [3,4]. In response, researchers have sought to investigate how insights from psychological and behavioural science may be used to address

the issue [5,6]. Simple measures such as infographics [7] and pausing for a few seconds to consider the accuracy of a particular news item [8] have shown promise. Recent work on psychological ‘inoculation’ against misinformation has also been fruitful in demonstrating a reduction in individual susceptibility to misinformation [9–11].

Inoculation theory [12] posits that pre-emptively exposing individuals to a weakened form of a misleading argument as well as teaching individuals how to refute those arguments triggers the production of ‘mental antibodies’. This process has been shown to confer psychological resistance against future manipulation attempts, much like a medical vaccine induces resistance against a particular pathogen [13,14]. Inoculation messages typically consist of motivational threat and refutational pre-emption: the threat component forewarns individuals that they may be exposed to a persuasive attack, and refutational pre-emption either entails directly providing individuals with the counterarguments that refute incoming (mis)information, known as *passive* inoculation, or it actively involves the participant in the generation of those counterarguments, known as *active* inoculation [15,16]. Active inoculation may confer a comparatively stronger inoculation effect because people are encouraged to generate their own antibodies and counterarguments [17,18]. That is, as the participant takes an active role in the inoculation process, this allows them to generate internal refutations themselves, which may lead to longer lasting effects owing to increased cognitive involvement [19].

Although early inoculation research focused primarily on ‘cultural truisms’ where participants all had the desired attitude on a given topic, later research has evidenced that inoculation can bolster resistance to persuasion even when people are already familiar with the topic or have been exposed to the misinformation before [20,21], so-called ‘therapeutic inoculation’. Therapeutic inoculation mirrors recent advances in medicine where therapeutic vaccines can still boost the production of antibodies even when people have already been infected [22,23]. A seminal meta-analysis by Banas & Rains [24] highlights the ability of inoculation interventions to confer resistance against persuasive attacks with an average intervention effect size of $d=0.43$. Several recent reviews have also highlighted the efficacy of inoculation in the context of misinformation specifically [15,16,25].

A final and important distinction has been made between refutational-same versus refutational-different messages [18], where refutational-*same* messages inoculate against specific material individuals will later be exposed to, whereas refutational-*different* messages tackle arguments that individuals may not directly be directly exposed to later, but that may help bolster resistance towards various different persuasive attacks [16]. In recent years, researchers have started to focus on inoculating people against misinformation *techniques*, rather than individual examples of misinformation, representing a form of refutational-different inoculation [16].

The benefit of technique-based and active inoculations lie in both their potential scalability and applicability [16,26,27], as they prepare individuals to resist being persuaded by messages that may be different in *content*, but use the same underlying persuasion *strategy*. While there is some evidence that traditional inoculation messages offer cross-protection (that is, protect the individual against persuasion attempts on untreated, but related attitudes; see [28]), technique-based inoculation potentiates a much a wider level of protection: the fast-paced nature of online misinformation means that the misleading *content* is topically varied and constantly evolving, and therefore it is arguably more effective to target the persuasive techniques which largely remain the same.

One example of an active, technique-based inoculation intervention is *Bad News*,¹ a free online browser game in which players learn about six common misinformation techniques, a categorization known as **DEPICT**: Discrediting opponents, Emotional language use, increasing intergroup Polarization, Impersonating people through fake accounts, spreading Conspiracy theories and evoking outrage through Trolling [1,9]. In the game, players take on the role of a fake news creator and are tasked with building a fake news empire by (i) gaining as many followers as they can and (ii) maximizing their credibility. During the game, players are forewarned about the threat of misinformation and exposed to weakened doses of the strategies used in its production, consistent with the mechanisms of inoculation theory [9,29]. Figure 1 shows a screenshot of the game environment.

In a series of studies, *Bad News* was shown to improve people’s ability to spot manipulative social media posts [27,30], confer similar effects across different language versions of the game [31], improve people’s confidence in their ability to spot misinformation [32] and induce detectable inoculation effects for up to 13 weeks after gameplay if players are given regular reminders or ‘booster shots’ [9].

The efficacy of *Bad News* and other technique-based inoculation interventions has been primarily evaluated by asking people to rate the reliability of fictitious social media posts, designed by

¹Playable at www.getbadnews.com.

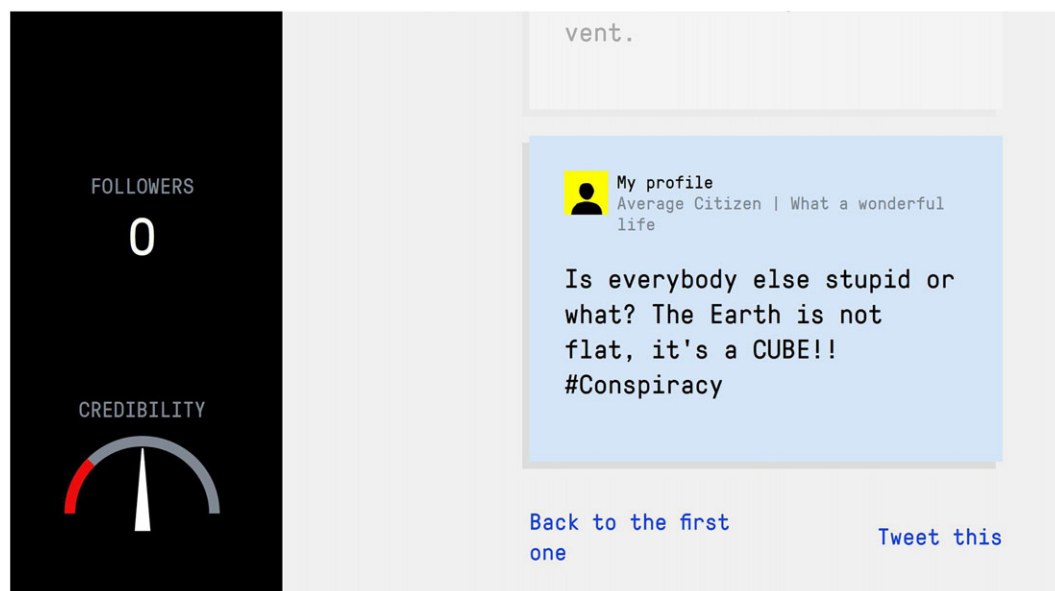


Figure 1. *Bad News* screenshot, with the 'Followers' and 'Credibility' meters on the left.

the researchers themselves [33]. Participants were shown a series of posts that were designed to make use of one of the manipulation techniques featured in the game (or not, to examine the effect of the inoculation on people's perceptions of 'real news'), before and after gameplay. Importantly, the misinformation posts were not selected for their truth value *per se*: while some may contain a kernel of truth, the primary selection criterion has been whether a post uses a particular misinformation technique.

However, several questions about the effectiveness of technique-based inoculation remain unanswered, which we address in this study. First, there is limited data available about whether active inoculation interventions also confer psychological resistance against real-life (as opposed to fictitious) examples of misinformation [17,26]. Second, no studies have directly compared the efficacy of technique-based inoculation for reducing susceptibility to misinformation that is simply *false* (i.e. a headline makes a claim of fact that is probably incorrect but not otherwise manipulative) versus misinformation that makes use of a specific misinformation *technique* that people are inoculated against (but may not contain a claim of fact, such as conspiratorial reasoning). Third, the most common outcome measures for testing technique-based inoculation interventions have been the perceived reliability [9,32] or manipulateness [17,34] of both misinformation and real news items. As Pennycook *et al.* [35] point out, it is also important to assess whether inoculation improves people's ability to assess the *accuracy* of real and false news. Asking participants to assess the *reliability* or *manipulativeness* of an item does not necessarily mean asking them to assess its veracity (i.e. if it is true or false), whereas this is the case for eliciting a headline's *accuracy*. Importantly, Roozenbeek *et al.* showed that eliciting different questions and response modes should not yield major variations in response patterns when assessing misinformation susceptibility [36].

Resolving these matters is important for several reasons; while using realistic yet fictitious items as outcome measures was crucial in early research to preclude the possibility of memory confounds (i.e. it is impossible for study participants to have seen the item set before, which could otherwise bias their assessment), it also reduces the ecological validity of the evaluation method [9,27]. Testing whether an intervention improves people's ability to spot *real-world* misinformation should therefore be a key indicator of its effectiveness. Furthermore, in a meta-analysis, Walter & Murphy [2] found that correcting real-world misinformation is more challenging than constructed misinformation. Therefore, investigating whether active inoculation interventions can also be effective at preventing unwanted persuasion attempts found in real-world misinformation gives insight into how inoculation interventions can be used to pre-emptively counter such misinformation.

In addition, although researchers have suggested that the benefit of these interventions is their potential to offer broader 'blanket' or 'umbrella' protection [13,15,28] against a wider array of misinformation, this has yet to be formally tested. As this test is crucial for knowing the extent to which inoculation treatments can confer psychological resistance even against entirely unknown examples of misinformation, the current study aims to examine whether and to what extent

technique-based inoculation interventions can offer cross-protection against misinformation that inoculated individuals were not trained to recognize.

Finally, different question framings (e.g. *reliability* versus *accuracy*) should not be expected to measure substantially different constructs, and so we may expect broadly similar response patterns. Assessing whether the effects of technique-based inoculation interventions are robust to different methods of measurement is an important indicator of the overall efficacy of such interventions.

2. The present research

In the light of the above questions, this study is to our knowledge, the first to evaluate *Bad News* using stimuli found ‘in the wild’ in two separate experiments. Following previous research designs by Roozenbeek & van der Linden [27] and Roozenbeek *et al.* [30], we implemented two voluntary pre-post survey experiments within the *Bad News* game environment. In experiment 1, we use an item set with social media posts that make use of one of the **DEPICT** techniques, to test if the *Bad News* game successfully reduces susceptibility to real-world misinformation that uses a misinformation technique they learned about in the game. In experiment 2, we use a different item set of false and true headlines used in previous studies [37,38] that do not explicitly make use of a **DEPICT** technique, to test whether cross-protection against untreated misinformation is achieved [28]. In addition, in experiment 2, we ask about the *accuracy* rather than the *reliability* of a set of headlines.

Between 15 March and 7 September 2020, a total of 2188 valid completed survey entries were collected ($n_1 = 1,216$, $n_2 = 972$). The rationale for the data collection period was twofold. First, based on previously known effect sizes in similar studies (Cohen’s $d = 0.30$), a power analysis determined a minimum of 483 participants would be necessary to capture previously known effect sizes of approximately $d = 0.30$, and second, the availability of the online platform hosting the study. This study was approved by the Cambridge Psychology Research Ethics Committee (PRE.2019.103). The datasets, measures, items and our visualization scripts are available on the OSF: <https://osf.io/59pjk/>. We discuss both experiments separately below.

3. Experiment 1

3.1. Method

Following Roozenbeek & van der Linden [27], experiment 1 is a simple within-subjects pre-post survey, administered at the start and at the end of the *Bad News* game. Survey participants were asked to rate the reliability of a series of nine social media posts on a 1–7 Likert scale, with 1 being ‘unreliable’ and 7 being ‘reliable’, at the start and end of the game. They were also asked several demographic questions (age, gender, education level and political affiliation). Figure 2 shows screenshots of the in-game survey environment.

Two real news (control) posts were taken from credible news sources, Reuters and Associated Press [39,40], and did not contain any misinformation, such as ‘Trump appeals U.S. judge’s border wall funding ruling’ (figure 2). The other six (misinformation) items were real examples of misinformation and matched so that each made use of one of the manipulation techniques people are inoculated against in the game: discrediting, emotion, polarization, impersonation, conspiracy, and trolling [1]. For example, the item making use of emotionally evocative language read ‘Horrific TV show inspiring suicide, says filmmaker’, and the item making use of the conspiracy technique read ‘Exposing the shadow ‘elite’ controlling the world’ (figure 2). In addition, we also included one false item about the novel coronavirus [4,41,42].² See the electronic supplementary material, table S2 for the full item wordings.

The misinformation items were a combination of previously used headlines by Traberg & van der Linden [43] and new headlines taken from an online fact-checking platform developed by researchers at Indiana University known as Hoaxy [44]. All items were selected based on their use of the **DEPICT** misinformation techniques mentioned above and were converted into a Tweet-format to facilitate integration in the game environment (figure 2). For the impersonation and trolling techniques, no

²Although the items were not selected based on their conceptual correlation (given that they rely on different misinformation techniques), we do note that the seven misinformation items show internal consistency ($M = 2.59$, $s.d. = 1.16$, Cronbach’s $\alpha = 0.74$). The real news (control) items have relatively low internal consistency ($M = 5.35$, $s.d. = 1.58$, Cronbach’s $\alpha = 0.52$), most likely owing to the low number of items, so we report item-level as well as averaged results.



Figure 2. Examples of the *Bad News* in-game survey environment for experiment 1. Clockwise from top left: start of the survey; a misinformation post (using the ‘conspiracy’ technique); a real news (control) post and a demographic question.

ecologically valid examples were found through Hoaxy. The reason for this is that impersonating people or organizations online [45] and trolling (i.e. baiting people on social media into responding emotionally or lashing out at others, see [46]), while common, are not featured on Hoaxy as examples of misinformation. We therefore relied on a manual Twitter search as well as previous (credible) news reporting of online trolling scams to find examples of social media content that made use of these techniques. In total, participants were thus shown nine social media posts before and after playing *Bad News* (two control items, one item for each of the six techniques, and one item about COVID-19). Participants were shown the same items pre- and post-gameplay. The items and their sources can be found in the electronic supplementary material, table S2. In experiment 1, we test the following hypotheses:

- H₁:** people who play *Bad News* rate real-life examples of misinformation [**H_{1a}**], but not real news [**H_{1b}**] as significantly less reliable after playing; and
- H₂:** people who play *Bad News* improve in their ability to discern the reliability of real-life examples of manipulative and non-manipulative news.

3.2. Sample

A total of 1 216 valid responses were collected. The sample consists of 55.1% men, with 48.0% of survey participants reporting being between 18 and 29 years old,³ with 58.0% having obtained a higher degree. Participants were also somewhat left-leaning ($M = 3.29$, $s.d. = 1.43$ on a 7-point Likert scale, 1 being ‘very left-wing’ and 7 being ‘very right-wing’). See the electronic supplementary material, table S1 for the sample composition.

3.3. Results

Following Roozenbeek & van der Linden [27], we conduct a series of paired-samples *t*-tests on the pre- and post-scores for each social media post (item), as well as on the aggregated scores for the

³Participants under 18 were excluded, as per our ethics approval.

misinformation and real news (control) items (see the electronic supplementary material, tables S3 and S4). For the misinformation items (averaged across all seven items), we find a significant reduction in perceived reliability post-gameplay ($M_{\text{pre}} = 2.58$ versus $M_{\text{post}} = 2.22$, $M_{\text{diff}} = -0.35$, $t_{1215} = -12.79$, $p < 0.001$, $d = -0.37$, 95% confidence interval (CI) $[-0.43, -0.31]$). This translates to 64.4% of the post-gameplay reliability scores being lower than the mean of pre-gameplay reliability scores (i.e. Cohen's $U_3 = 64.4$). We also find that participants rate all six misinformation posts that made use of one of the manipulation techniques learned in *Bad News* as significantly less reliable after playing (all p 's < 0.001 , with Cohen's d ranging between $d = -0.12$ and $d = -0.27$). In addition, participants rated the post about coronavirus as significantly less reliable after gameplay ($M_{\text{pre}} = 1.88$ versus $M_{\text{post}} = 1.78$, $M_{\text{diff}} = -0.10$, $t_{1215} = -2.05$, $p = 0.04$, $d = -0.06$, 95% CI $[-0.12, -0.003]$), albeit with a substantially lower effect size. We note that the pre-score for the coronavirus post is the lowest out of all the items and is even lower than the lowest post-gameplay score for the other items ($M_{\text{coronavirus,pre}} = 1.88$ versus $M_{\text{discredit,post}} = 1.90$), indicating possible flooring effects (see the electronic supplementary material, table S4).

In order to test whether the inoculation effect is different across different levels of initial (i.e. pre-gameplay) levels of misinformation susceptibility, we also grouped people (in terciles) by their performance on the pre-test. Comparing across reliability judgements on the pre-test, we find that the effect of inoculation was most pronounced for those who were most susceptible to misinformation prior to gameplay ($F_{2,2426} = 64.8$, $p < 0.001$, $\eta^2 = 0.025$), a finding consistent with results from Roozenbeek & van der Linden [27]. The full overview of this analysis can be found in the electronic supplementary material, table S5; see also figures S2 and S3. Overall, our results thus support hypothesis **H_{1a}**.

Second, we find that participants also rate 'real' news as significantly less reliable after playing, albeit with a lower effect size than misinformation ($M_{\text{pre}} = 5.35$ versus $M_{\text{post}} = 5.18$, $M_{\text{diff}} = -0.17$, $t_{1215} = -3.55$, $p < 0.001$, $d = -0.10$, 95% CI $[-0.16, -0.045]$). However, we note that we only used two real news posts, and that only one post shows a significant pre-post difference ($p_{\text{real,Brexit}} < 0.001$ versus $p_{\text{real,Trump}} = 0.15$), potentially indicating that item effects are at play. Nonetheless, we do not find support for hypothesis **H_{1b}**.

Third, we find a significant difference between the difference scores (real news minus misinformation reliability scores) before and after gameplay ($M_{\text{diff,pre}} = 2.78$ versus $M_{\text{diff,post}} = 2.96$, $M_{\text{diff,diff}} = 0.18$, $t_{1215} = 3.45$, $p < 0.001$, $d = 0.10$, 95% CI $[0.042, 0.15]$), indicating improved truth discernment. These results support hypothesis **H₂**.

To see to what extent demographic variables influence the inoculation effect, we conduct a linear regression with the average pre-post different scores for the misinformation (false news) items as the dependent variable and gender, age, education and political affiliation as independent variables. We find that none of these variables significantly predict the pre-post-inoculation effect for the false news items (all p 's > 0.387), except for education level ($p = 0.007$), so that people with a higher education level also display a larger inoculation effect.

Finally, we also conducted a linear regression with the same independent variables, but with the difference in discernment (i.e. mean reliability scores for the real news (control) posts minus mean reliability scores for the misinformation posts) before and after playing *Bad News* as the dependent variable. We find that none of the independent variables are significant predictors of improved discernment (all p 's > 0.206); see the electronic supplementary material, table S6.

Figure 3 shows the results from experiment 1 in a raw data, description and inference (RDI) plot.

3.4. Discussion

Overall, we find that playing *Bad News* significantly decreases the perceived reliability of real-world misinformation that made use of one of the **DEPICT** misinformation techniques. The effect sizes are in line with the initial within-subject study on the effectiveness of *Bad News* [27], which used a similar methodology to the present study; Roozenbeek & van der Linden [27] report pre-post within-subjects effect sizes between $d = -0.16$ and $d = -0.36$ for their fictitious social media posts, whereas we find slightly smaller effects in this study: between $d = -0.06$ for the coronavirus item, which suffers from flooring effects in the sense that almost all participants found this post highly unreliable already in the pre-test ($M_{\text{pre}} = 1.88$ versus $M_{\text{post}} = 1.78$), and $d = -0.27$ for the 'emotion' item. Nonetheless, the average effect size for all misinformation posts of $d = -0.37$ is encouraging.

Second, in line with recent findings on the effectiveness of media literacy interventions [7], we find that playing *Bad News* slightly decreases the perceived reliability of real news alongside misinformation, albeit to a much smaller extent ($d = -0.37$ versus $d = -0.10$). However, we note that we used a smaller number of real news than misinformation items and are therefore unable to rule

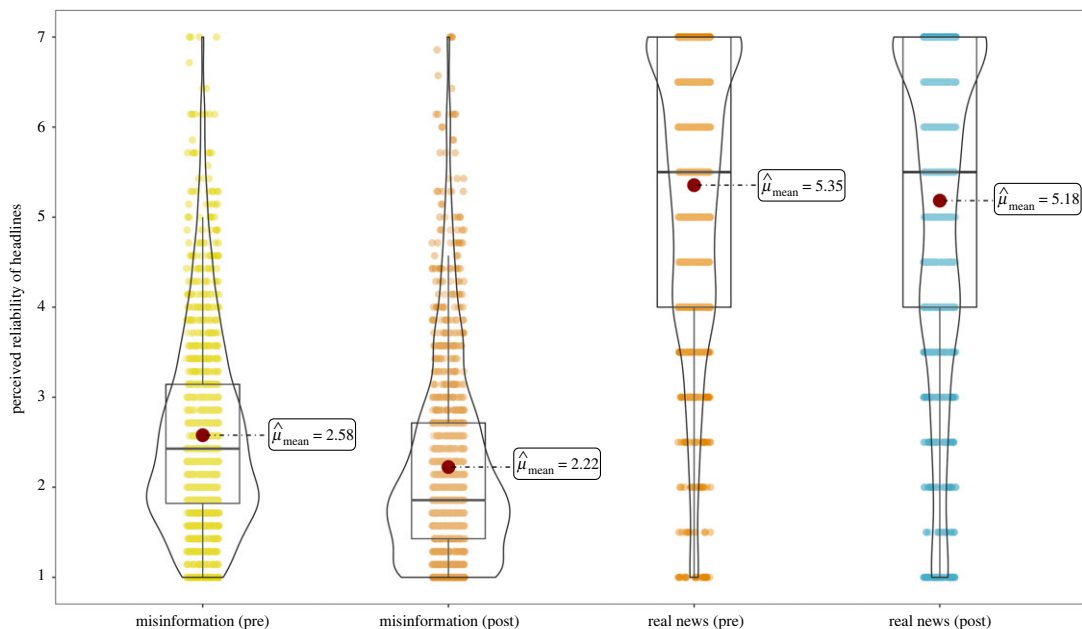


Figure 3. Pre-post RDI plot for misinformation items pre- and post-gameplay, and the real news (control) headlines pre- and post-gameplay. Pre-post differences are significant for both misinformation ($t_{1215} = -12.79$, $p < 0.001$, $d = -0.37$) and real news ($t_{1215} = -3.55$, $p < 0.001$, $d = -0.10$). See the electronic supplementary material, table S3 and figure S1 for item-level results.

out that the observed findings are owing to item effects [30]. The fact that only one out of two real news items showed a significant reduction in perceived reliability post-gameplay supports this notion. In addition, truth discernment [37,38] increased significantly after gameplay, indicating that *Bad News* players improve in their ability to discern real information from misinformation.

4. Experiment 2

4.1. Method

In experiment 2, we used the two-group within-subject design developed by Roozenbeek *et al.* [30] in order to address several open questions not answered in experiment 1. First, while the items in experiment 1 consisted of social media posts, in experiment 2, participants were shown a series of real and false news headlines (including headline source) as they may appear on someone's social media feed. These headlines were selected from a larger popular set of true and false headlines used in previous research on misinformation (see [37,38]). The headlines were selected based on their representativeness of different forms of misinformation that individuals may be exposed to online, and not necessarily their internal consistency.⁴

Second, in experiment 1, participants were shown the same items before and after gameplay. Following Roozenbeek *et al.* [30], participants in experiment 2 were therefore shown different sets of real and false headlines before and after playing (we will call these sets A and B). At the start of the survey, participants were randomly assigned to a group (A-B or B-A). The A-B group was shown item set A before gameplay, and item set B after gameplay. The B-A group was first shown set B and then set A. Doing so allows us to (i) compare the accuracy ratings of real and false headlines for the *same* item sets (A-A and B-B) by *different* participants before and after gameplay; (ii) compare the accuracy ratings for *different* item sets (A-B and B-A) for the *same* participants and (iii) examine the overall effect (averaged across both item sets) for the real and false headlines, as well as for truth discernment.

Third, following Pennycook *et al.* [35,37,47], we asked participants to rate the *accuracy* (rather than the reliability) of real and false news headlines on a 1–7 Likert scale (1 being 'inaccurate' and 7 being 'accurate'). Fourth, as pointed out by Maertens *et al.* [9], in order to more accurately assess whether playing *Bad News* improves truth discernment (i.e. people's ability to discern true from false news), it

⁴We find low internal consistency ($M_{\text{SetA,Real}} = 4.77$, $\text{s.d.}_{\text{SetA,Real}} = 1.15$, Cronbach's $\alpha = 0.38$; $M_{\text{SetA,Fake}} = 2.67$, $\text{s.d.}_{\text{SetA,Fake}} = 1.26$, Cronbach's $\alpha = 0.62$; $M_{\text{SetB,Real}} = 4.36$, $\text{s.d.}_{\text{SetB,Real}} = 1.19$, Cronbach's $\alpha = 0.59$; $M_{\text{SetB,Fake}} = 2.36$, $\text{s.d.}_{\text{SetB,Fake}} = 1.11$, Cronbach's $\alpha = 0.55$).

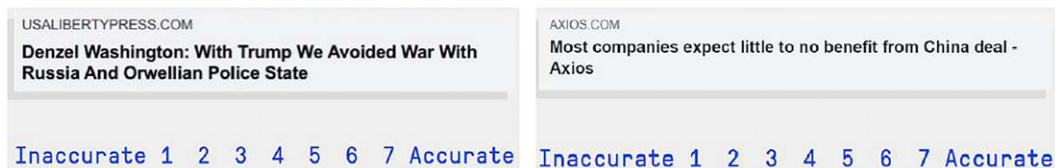


Figure 4. Examples of the *Bad News* in-game survey environment for experiment 2, with a false headline (left) and a true headline (right) adapted from Pennycook & Rand [37,38].

is useful to present an equal number of real and false items. Participants were therefore shown four real and four false headlines before gameplay, and a different set of four real and four false headlines after gameplay, for a total of eight real and eight false headlines. Figure 4 shows screenshots of what these headlines looked like in the game environment (see the electronic supplementary material, table S7 for the full item sets).

Again following Roozenbeek *et al.* [30], we test the following hypotheses:

- H₁:** when comparing an index of the same items, pre-test (group x) with post-test (group y), there is a significant decrease in the perceived accuracy of false headlines, both for item set A-A [**H_{1a}**] and item set B-B [**H_{1b}**];
- H₂:** when comparing an index of the same items, pre-test (group x) with post-test (group y), there is no significant decrease in the perceived accuracy of real headlines, both for item set A-A [**H_{2a}**] and item set B-B [**H_{2b}**];
- H₃:** when comparing different item sets, pre-test (group x) with post-test (group x), there is a significant decrease in the perceived accuracy of false headlines, both for group A-B [**H_{3a}**] and group B-A [**H_{3b}**];
- H₄:** when comparing different item sets, pre-test (group x) with post-test (group x), there is no significant difference in the perceived accuracy of real headlines, both for group A-B [**H_{4a}**] and group B-A [**H_{4b}**]; and
- H₅:** people who play *Bad News* significantly improve in their ability to discern the accuracy of real-life examples of real and false headlines.

4.2. Sample

A total of 968 valid pre-post survey responses were collected from the in-game survey. The sample consisted of 50.7% men, with 55.5% of respondents being between 18 and 29 years old, and 51.8% having obtained a higher degree. As in experiment 1, participants were left-leaning ($M = 3.17$, $s.d. = 1.39$). In addition, we also asked participants about their region of origin (46.4% being from the USA, 26.9% from Europe), and whether participants had played *Bad News* before (10.2% yes, 89.8% no). See the electronic supplementary material, table S1 for the full sample composition.

4.3. Results

We first look at whether an inoculation effect is found when crossing the item sets between groups (hypotheses **H₁** and **H₂**), comparing item set A (B) pre-test scores with item set A (B) post-test scores. Doing so yields a significant effect for the false news items for item set A (pre-test for the A-B group) versus item set A (post-test for the B-A group) ($M_{\text{Fake,SetA,pre}} = 2.66$ versus $M_{\text{Fake,SetA,post}} = 2.41$, $M_{\text{diff}} = 0.25$, $t_{966} = 3.109$, $p = 0.002$, $d = 0.20$, 95% CI [0.073, 0.33]). These results support hypothesis **H_{1a}**. However, we find no significant difference for the false news items for item set B (pre-test for the B-A group) versus item set B (post-test for the A-B group) ($M_{\text{Fake,SetB,pre}} = 2.36$ versus $M_{\text{Fake,SetB,post}} = 2.40$, $M_{\text{diff}} = 0.04$, $t_{966} = 0.518$, $p = 0.605$, $d = 0.03$, 95% CI [-0.093, 0.16]); see the electronic supplementary material, tables S8 and S9. To check whether this result implies the absence of an effect of interest, we performed a two one-sided *t*-test (TOST) equivalence test with a smallest effect size of interest of $d = (-)0.20$ and $\alpha = 0.01$ [48].⁵ Doing so confirms statistical equivalence to zero: $t_{964} = -2.60$, $p = 0.005$. Our data thus does not support hypothesis **H_{1b}**.

⁵We used the TOSTER package in Jamovi for equivalence testing. The SESOI of $d = +0.20$ was based on the effect size of $d = 0.20$ obtained when comparing item set A (pre) to item set A (post). See also: <https://daniellakens.blogspot.com/2017/03/equivalence-testing-in-jamovi.html>.

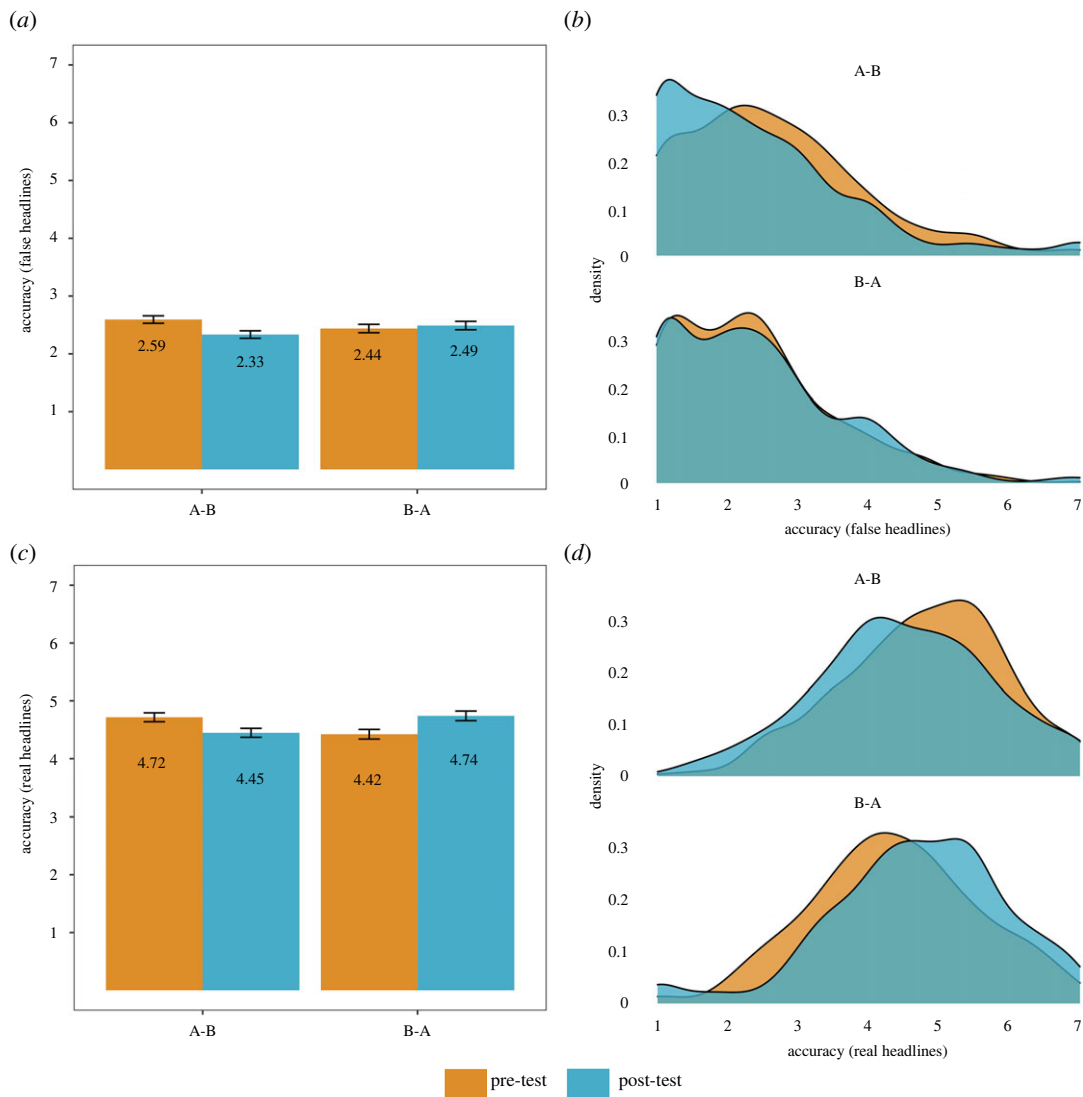


Figure 5. Bar graphs and density plots for false headlines (a,b) and real headlines (c,d). Note: $n = 968$. Error bars represent the 95% confidence interval.

We perform the same analyses for the real news items. Doing so yields no significant difference for both item set A ($M_{\text{Real,SetA,pre}} = 4.77$ versus $M_{\text{Real,SetA,post}} = 4.68$, $M_{\text{diff}} = 0.09$, $t_{966} = 1.176$, $p = 0.002$, $d = 0.076$, 95% CI $[-0.051, 0.20]$) and item set B ($M_{\text{Real,SetB,pre}} = 4.36$ versus $M_{\text{Real,SetB,post}} = 4.50$, $M_{\text{diff}} = 0.14$, $t_{966} = 1.766$, $p = 0.078$, $d = 0.11$, 95% CI $[-0.013, 0.24]$); see the electronic supplementary material, tables S8 and S9. However, A TOST equivalence test does not confirm statistical equivalence to zero for item set A ($t_{924} = -1.93$, $p = 0.027$) or item set B ($t_{961} = -1.34$, $p = 0.09$), which means that we cannot rule out that there is an effect (i.e. an *increase* in the perceived accuracy of real news post-gameplay) that is larger than $d = 0.20$. These results support hypothesis H_{2a} , and partially support hypothesis H_{2b} .

Next, we examine whether inoculation effects can be detected when comparing different item sets within the same groups (hypotheses H_3 and H_4). When looking at groups A-B and B-A together, we find a significant overall reduction in perceived accuracy post-gameplay for the false news items ($M_{\text{Fake,pre}} = 2.52$ versus $M_{\text{Fake,post}} = 2.41$, $M_{\text{diff}} = 0.11$, $t_{966} = -3.223$, $p = 0.001$, $d = -0.10$, 95% CI $[-0.17, -0.04]$). This means that 54.0% of the post-gameplay false headline accuracy scores are lower than the mean of pre-gameplay accuracy scores (i.e. Cohen's $U_3 = 54.0$). However, this effect is only significant for the A-B group ($M_{\text{Fake,A-B,pre}} = 2.67$ versus $M_{\text{Fake,A-B,post}} = 2.41$, $M_{\text{diff}} = 0.26$, $t_{512} = -5.544$, $p < 0.001$, $d = -0.24$, 95% CI $[-0.33, -0.16]$), but not for the B-A group ($M_{\text{Fake,B-A,pre}} = 2.36$ versus $M_{\text{Fake,B-A,post}} = 2.41$, $M_{\text{diff}} = 0.05$, $t_{454} = 0.956$, $p = 0.34$, $d = 0.05$, 95% CI $[-0.047, 0.14]$); see the electronic supplementary material, tables S8 and S9. A TOST equivalence test confirms statistical equivalence to zero for the B-A group: $t_{454} = 3.31$, $p < 0.001$. We thus find support for hypothesis H_{3a} , but not for hypothesis H_{3b} .

For the real news items, we find no significant overall difference in perceived reliability post-gameplay ($M_{\text{Real,pre}} = 4.58$ versus $M_{\text{Real,post}} = 4.58$, $M_{\text{diff}} = 0.007$, $t_{966} = 0.173$, $p = 0.86$, $d = 0.01$, 95% CI $[-0.057, 0.069]$). However, for the A-B group, we find a significant *decrease* in perceived reliability of real news ($M_{\text{Real A-B,pre}} = 4.77$ versus $M_{\text{Real,A-B,post}} = 4.50$, $M_{\text{diff}} = 0.27$, $t_{512} = -4.789$, $p < 0.001$, $d = -0.21$, 95% CI $[-0.23, -0.12]$), whereas for the B-A group, we find the opposite, namely a significant and descriptively larger *increase* in the perceived reliability of real news ($M_{\text{Real B-A,pre}} = 4.36$ versus $M_{\text{Real,B-A,post}} = 4.68$, $M_{\text{diff}} = 0.32$, $t_{454} = 5.279$, $p < 0.001$, $d = 0.25$, 95% CI $[0.15, 0.34]$); see the electronic supplementary material, tables S8 and S9. We thus find no support for either hypothesis **H_{4a}** or hypothesis **H_{4b}**.

In terms of participants' ability to distinguish real and false news headlines, we find a significant overall increase in truth discernment post-gameplay, indicating that *Bad News* players generally improve in their ability to discern true from false information ($M_{\text{Discernment,pre}} = 2.06$ versus $M_{\text{Discernment,post}} = 2.18$, $M_{\text{diff}} = 0.12$, $t_{966} = 2.421$, $p = 0.016$, $d = 0.08$, 95% CI $[0.015, 0.14]$). These results support hypothesis **H₅**; see the electronic supplementary material, tables S8 and S9. Figure 5 shows the results in a bar graph and density plot for both the false and real headlines.

Finally, as in experiment 1, we check whether any covariates influence the observed inoculation effects. To do so, we again first conduct a linear regression with the average pre-post different scores for the misinformation (false news) items as the dependent variable, group (A-B or B-A) as a factor variable, and gender, age, education and political affiliation as independent variables. Doing so shows that none of our covariates significantly predicts the pre-post-inoculation effect for false headlines (all p 's > 0.063). Second, we conduct a linear regression with the same factor and independent variables, but with the difference in truth discernment before and after gameplay as the dependent variable. Doing so shows that none of the covariates predict the pre-post-inoculation effect for truth discernment (all p 's > 0.14), except age ($p = 0.005$), so that being older is associated with a higher pre-post difference in truth discernment; see the electronic supplementary material, table S10.

4.4. Discussion

Overall, we find that playing the *Bad News* game significantly decreases the perceived accuracy of false headlines, with no effect on real headlines, leading to improved truth discernment. However, this effect was only significant for the A-B group, and not for the B-A group. Furthermore, while the overall effect for real news headlines was not significant, we found a significant pre-post *reduction* in the perceived accuracy of real news for the A-B group and a significant *increase* in perceived accuracy for the B-A group. These findings indicate that the experiment might suffer from item effects, in the sense that the observed results may be skewed by the items that were used: rather than the *Bad News* game influencing the perceived accuracy of real headlines, we find that the observed pre-post differences are owing to imbalances in the *initial* perceived accuracy of headline sets A and B [30]. This is to be expected, as the item sets were not psychometrically validated *a priori*. Nonetheless, our findings are highly encouraging for the general applicability of *Bad News* as a tool to learn how to spot misinformation, especially because the false headlines used in this experiment did not explicitly make use of the misinformation techniques learned in the game. Overall, our results show a reduced (but still significant) inoculation effect even for misinformation that *Bad News* players were not inoculated against, indicating partial cross-protection [28].

5. General discussion and conclusion

Across two large-scale experiments, we find that the *Bad News* game confers psychological resistance against misinformation techniques used in real-life examples of online misinformation. In line with previous research, we find that players' truth discernment, or their ability to distinguish manipulative from non-manipulative information, significantly improves after gameplay. Importantly, this effect is primarily driven by a decreased perceived reliability (or accuracy) of misinformation. For real news, the findings are somewhat more ambiguous; while in experiment 1, we find a small but significant decrease in perceived reliability, we find the same effect for one item set used in experiment 2, but the opposite in the second item set, namely an *increase* in the perceived accuracy of real news (in line with findings by, for example, Pennycook *et al.* [49], who reported a post-intervention increase in participants' willingness to share real news headlines). In general, these results align with the findings by Roozenbeek *et al.* [30].

Although not a consistent finding (and one that may be owing to item effects), the observed decrease in the perceived reliability of real news warrants further discussion, as this is not an explicit goal of the inoculation intervention. A similar phenomenon was observed for a different anti-misinformation intervention, so-called ‘digital literacy tips’ [7]. We argue that this finding is not a major concern, for two reasons. First, so long as (i) a more accurate identification of misinformation is the main psychological target (that is, that we see a significant difference between the reduced reliability of misinformation versus real news, which is the case here) and (ii) the overall reduction in perceived reliability of real news remains small, this finding is arguably not alarming. For example, while participants may have reduced their reliability judgements of real news in experiments 1 and 2, their overall judgement was that real headlines were still on the very high end of the reliability scale, meaning they did not change their judgement entirely from reliable to unreliable. Second, believing any news headline from a single source to be 100% reliable may not be the ideal way to make accurate judgements about news headlines in general, as sometimes even mainstream media sites publish misleading information, perhaps owing to human error [50]. That is, when it comes to news media consumption in the real world, it may in fact be beneficial to rely on multiple sources before accepting a news headline to be entirely true, as all sources may have their own spin on current events [51].

With respect to false news, while our findings in experiment 1 are straightforward, the results from experiment 2 are more ambiguous: effect sizes reported in this experiment are descriptively smaller than in experiment 1, and group B-A showed no significant decrease in the perceived accuracy of false headlines [30]. That said, the overall results still point towards an improved ability to distinguish real from false headlines. These results are in line with findings by Walter & Murphy [2], who reported that real-world misinformation is more difficult to correct than constructed misinformation.

One reason for this non-predicted effect may be that the false headlines we used in experiment 2 did not make use of the **DEPICT** misinformation techniques learned while playing *Bad News* (e.g. using emotional language, trolling or impersonation), and players were not specifically inoculated against these types of false headlines (as the intervention does not seek to improve people’s ability to distinguish between false and true information, but rather aims to improve people’s ability to recognize the use of specific manipulation techniques). This may explain the differences between the present study and those reported in Roozenbeek *et al.* [30], who did include items that made use of misinformation techniques featured in the game. Nonetheless, the fact that we still find a significant improvement in truth discernment and reduced overall accuracy for false headlines is encouraging, as it shows that the inoculation effect conferred by playing the *Bad News* game is at least partially transferrable to previously unseen types of misinformation. In other words, active inoculation treatments appear to be at least somewhat effective at conferring cross-protection against related but untreated persuasive attacks [13,15,28]. Though we note that there is ongoing discussion on the boundaries between the ‘umbrella’ or ‘blanket’ of protection (which confers resistance against different examples of the *same* misinformation technique) versus true *cross-protection*, which could occur when resistance is formed against novel misinformation that either does not make use of the same strategies (as in the current study) or makes more explicit use of entirely different strategies (akin to different topics) that people were not inoculated against. Future research will have to examine—in a more controlled way—whether inoculation games can also confer *cross-protection* in the sense that inoculating people against one misinformation technique (e.g. conspiracy theories) also confers protection against unmentioned and entirely different techniques (e.g. impersonation).

Although we conducted two studies, we were still unable to account for several key limitations. First, with respect to covariates, although we do not find political ideology to be an independent predictor of the inoculation effect (see the electronic supplementary material, table S9), previous research has found political similarity with sources to impact misinformation susceptibility [43]. As such, it is still possible that more complex political factors involving the political slant of sources or news content may have impacted the inoculation effect found in experiment 2 (as these headlines included source information).

Second, our item sets were not psychometrically validated *a priori*, and we note the presence of item effects that may complicate our findings. We were only able to use a limited number of items in each experiment, and the low internal consistency for our item sets, especially in experiment 2, limits the generalizability of our findings. Third, this study also explored whether eliciting different question framings (e.g. the reliability or accuracy of an item set) can be expected to yield different response patterns when testing the efficacy of anti-misinformation interventions. Our results are inconclusive: although we find that both question framings yield significant results in the hypothesized direction, the effect sizes in experiments 1 and 2 are substantially different. However, because we used different item sets, we are unable to directly compare both experiments’ response patterns. Thus, while this

study is to our knowledge, the first to test two separate outcome measures (reliability and accuracy) on the same intervention, it is unclear to what extent these measures are interpreted differently by participants, and how this may affect their item ratings; see Roozenbeek *et al.* [36].

Finally, more can be said about the mechanisms underlying the efficacy of active inoculation games. In related research, scholars have provided some early evidence for the role of motivational threat [52], which proved significantly higher in active than passive inoculation treatments [17]. Other fruitful avenues include the potential role of confidence and attitude certainty [32] as well as higher post-inoculation sharing following active versus passive inoculation [17]. We suggest that future research further explores the potential cognitive and affective mechanisms by which active inoculation interventions can confer resistance to misinformation, particularly insofar it (i) makes people aware of their own vulnerability and (ii) elicits greater motivation to protect oneself from manipulation [53].

Overall, we corroborate previous findings that playing *Bad News* improves people's ability to spot misinformation [27,30,32]. This is especially important in light of the game's relatively easy scalability (being free-to-play in a browser on a phone, computer or tablet) and availability in approximately 20 languages. The game may be implemented as part of media literacy curricula in schools, or played as a standalone game in-browser. In addition, the game may be deployed in conjunction with other anti-misinformation tools such as videos [54], 'prebunking' infographics [17], accuracy-based interventions [8,47] or media literacy interventions [7], to improve resilience against online misinformation at scale.

Ethics. This study was approved by the University of Cambridge Psychology Research Ethics Committee (PRE.2019.103).

Data accessibility. The datasets, measures, items and our visualization scripts are available on the OSF: <https://osf.io/59pjk/>.

Electronic supplementary material is available online at [55].

Authors' contributions. J.R.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—original draft, writing—review and editing; C.S.T.: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing—original draft, writing—review and editing; S.v.d.L.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. J.R. and S.v.d.L. were involved in the development of the *Bad News* game.

Funding. We received no funding for this study.

Acknowledgements. We would like to thank David G Rand and Gordon Pennycook for their help with experiment 2.

References

- van der Linden S, Roozenbeek J. 2020 Psychological inoculation against fake news. In *The psychology of fake news: accepting, sharing, and correcting misinformation* (eds R Greifeneder, M Jaffé, E Newman, N Schwarz), pp. 147–170. Hove, UK: Psychology Press.
- Walter N, Murphy ST. 2018 How to unring the bell: a meta-analytic approach to correction of misinformation. *Commun. Monogr.* **85**, 423–441. (doi:10.1080/03637751.2018.1467564)
- Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. 2021 Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat. Hum. Behav.* **5**, 337–348. (doi:10.1038/s41562-021-01056-1)
- Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, van der Bles AM, van der Linden S. 2020 Susceptibility to misinformation about COVID-19 around the world. *R. Soc. Open Sci.* **7**, 201199. (doi:10.1098/rsos.201199)
- Van Bavel JJ *et al.* 2020 Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471. (doi:10.1038/s41562-020-0884-z)
- van der Linden S. 2022 Misinformation: susceptibility, spread, and interventions to immunize the public. *Nat. Med.* **28**, 460–467. (doi:10.1038/s41591-022-01713-6)
- Guess AM, Lerner M, Lyons B, Montgomery JM, 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. *Proc. Natl Acad. Sci. USA* **117**, 15 536–15 545. (doi:10.1073/pnas.1920498117)
- Fazio L. 2020 Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Misinformation Rev.* **1**. (doi:10.37016/mr-2020-009)
- Maertens R, Roozenbeek J, Basol M, van der Linden S. 2021 Long-term effectiveness of inoculation against misinformation: three longitudinal experiments. *J. Exp. Psychol.* **27**, 1–16. (doi:10.1037/xap0000315)
- Roozenbeek J, van der Linden S. 2020 Breaking harmony square: a game that 'inoculates' against political misinformation. *The Harvard Kennedy School (HKS) Misinformation Rev.* **1**. (doi:10.37016/mr-2020-47)
- van der Linden S, Panagopoulos C, Roozenbeek J. 2020 You are fake news: the emergence of political bias in perceptions of fake news. *Media Culture Soc.* **42**, 460–470. (doi:10.1177/0163443720906992)
- McGuire WJ. 1964 Some contemporary approaches. *Adv. Exp. Soc. Psychol.* **1**, 191–229. (doi:10.1016/S0065-2601(08)60052-0)
- Compton J, Pfau M. 2005 Inoculation theory of resistance to influence at maturity: recent progress in theory development and application and suggestions for future research. *Ann. Int. Commun. Assoc.* **29**, 97–145. (doi:10.1207/s15567419cy2901_4)
- McGuire WJ, Papageorgis D. 1961 The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *J. Abnormal Soc. Psychol.* **62**, 327–337. (doi:10.1037/h0042026)
- 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. *Soc. Pers. Psychol. Compass* **15**, e12602. (doi:10.1111/spc3.12602)
16. Traber CS, Roozenbeek J, van der Linden S. 2022 Psychological inoculation against misinformation: current evidence and future directions. *Ann. Am. Acad. Political Soc. Sci.* **700**, 136–151. (doi:10.1177/00027162221087936)
 17. 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 Soc.* **8**. (doi:10.1177/20539517211013868)
 18. McGuire WJ, Papageorgis D. 1961 Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *J. Abnormal Soc. Psychol.* **63**, 326–332. (doi:10.1037/h0048344)
 19. Tyler SW, Hertel PT, McCallum MC, Ellis HC. 1979 Cognitive effort and memory. *J. Exp. Psychol.* **5**, 607–617. (doi:10.1037/0278-7393.5.6.607)
 20. van der Linden S, Maibach E, Cook J, Leiserowitz A, Lewandowsky S. 2017 Inoculating against misinformation. *Science* **358**, 1141–1142. (doi:10.1126/science.aar4533)
 21. Wood MLM. 2007 Rethinking the inoculation analogy: effects on subjects with differing preexisting attitudes. *Hum. Commun. Res.* **33**, 357–378. (doi:10.1111/j.1468-2958.2007.00303.x)
 22. Compton J. 2020 Prophylactic versus therapeutic inoculation treatments for resistance to influence. *Commun. Theory* **30**, 330–343. (doi:10.1093/ct/qtz004)
 23. Ivanov B, Rains SA, Dillingham LL, Parker KA, Geegan SA, Barbati JL. 2021 The role of threat and counterarguing in therapeutic inoculation. *Southern Commun. J.* **87**, 15–27. (doi:10.1080/1041794X.2021.1983012)
 24. Banas JA, Rains SA. 2010 A meta-analysis of research on inoculation theory. *Commun. Monogr.* **77**, 281–311. (doi:10.1080/03637751003758193)
 25. Lewandowsky S, van der Linden S. 2021 Countering misinformation and fake news through inoculation and prebunking. *Europ. Rev. Soc. Psychol.* **32**, 348–384. (doi:10.1080/10463283.2021.1876983)
 26. Cook J, Lewandowsky S, Ecker UKH. 2017 Neutralizing misinformation through inoculation: exposing misleading argumentation techniques reduces their influence. *PLoS ONE* **12**, 1–21. (doi:10.1371/journal.pone.0175799)
 27. Roozenbeek J, van der Linden S. 2019 Fake news game confers psychological resistance against online misinformation. *Humanities Soc. Sci. Commun.* **5**, 1–10. (doi:10.1057/s41599-019-0279-9)
 28. Parker KA, Rains SA, Ivanov B. 2016 Examining the ‘Blanket of Protection’ conferred by inoculation: the effects of inoculation messages on the cross-protection of related attitudes. *Commun. Monogr.* **83**, 49–68. (doi:10.1080/03637751.2015.1030681)
 29. Compton J. 2013 Inoculation theory. In *The SAGE handbook of persuasion: developments in theory and practice* (eds JP Dillard, L Shen), pp. 220–236, 2nd edn. Newbury Park, CA: SAGE Publications, Inc.
 30. Roozenbeek J, Maertens R, McClanahan W, van der Linden S. 2021 Disentangling item and testing effects in inoculation research on online misinformation. *Educ. Psychol. Measurement* **81**, 340–362. (doi:10.1177/0013164420940378)
 31. 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 Rev.* **1**, 8. (doi:10.37016/mr-2020-008)
 32. Basol M, Roozenbeek J, van der Linden S. 2020 Good news about Bad News: gamified inoculation boosts confidence and cognitive immunity against fake news. *J. Cogn.* **3**, 1–9. (doi:10.5334/joc.91)
 33. Roozenbeek J, Maertens R, Herzog S M, 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**.
 34. Roozenbeek J, van der Linden S. 2022 How to combat health misinformation: a psychological approach. *Am. J. Health Promotion* **36**, 569–575. (doi:10.1177/08901171211070958)
 35. Saleh N, Roozenbeek J, Makki F, McClanahan W, van der Linden S. 2021 Active inoculation boosts attitudinal resistance against extremist persuasion techniques – a novel approach towards the prevention of violent extremism. *Behav. Public Policy* 1–24. (doi:10.1017/bpp.2020.60)
 36. Pennycook G, Cannon T, Rand DG. 2018 Prior exposure increases perceived accuracy of fake news. *J. Exp. Psychol.* **147**, 1865–1880. (doi:10.1037/xge0000465)
 37. Pennycook G, Rand DG. 2019 Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50. (doi:10.1016/j.cognition.2018.06.011)
 38. Pennycook G, Rand DG. 2020 Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *J. Pers.* **88**, 185–200. (doi:10.1111/jopy.12476)
 39. Casert R. 2020 *Brexit deal cleared by EU parliament; UK set to leave Friday*. New York, NY: Associated Press. See <https://apnews.com/article/b22e257a1fb0c86f221b20304f2233d4>.
 40. Wolfe J. 2019 *Trump appeals U.S. Judge's border wall funding ruling*. London, UK: Reuters. See <https://www.reuters.com/article/us-usa-immigration-court-idUSKCN1TU0R1>.
 41. Reuters. 2020 False claim: gargling with warm water and salt or vinegar will kill the coronavirus. www.uk.reuters.com/article/uk-factcheck-coronavirus-salt-vinegar/false-claim-gargling-with-warm-water-and-salt-or-vinegar-will-kill-the-coronavirus-idUSKBN2142MW.
 42. World Health Organization. 2020 Coronavirus disease (COVID-19) advice for the public: mythbusters. www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters.
 43. Traber CS, van der Linden S. 2022 Birds of a feather are persuaded together: perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Pers. Ind. Diff.* **185**, 111269. (doi:10.1016/j.paid.2021.111269)
 44. Shao C, Ciampaglia GL, Flammini A, Menczer F. 2016 Hoaxy: a platform for tracking online misinformation. In *Proc. of the 25th Int. Conf. Companion on World Wide Web. WWW '16 Companion*, pp. 745–750. Geneva, Switzerland: Int. World Wide Web Steering Committee.
 45. BBC News. 2018 A fake billionaire is fooling people on twitter. www.bbc.co.uk/news/world-us-canada-45331781.
 46. Griffiths MD. 2014 Adolescent trolling in online environments: a brief overview. *Educ. Health* **32**, 85–87.
 47. Pennycook G, Epstein Z, Mosleh M, Arechar A, Eckles D, Rand DG. 2021 Shifting attention to accuracy can reduce misinformation online. *Nature* **592**, 590–595. (doi:10.1038/s41586-021-03344-2)
 48. Lakens D, Scheel AM, Isager PM. 2018 Equivalence testing for psychological research: a tutorial. *Adv. Methods Practices Psychol. Sci.* **1**, 259–269. (doi:10.1177/2515245918770963)
 49. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. 2020 Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. *Psychol. Sci.* **31**, 770–780. (doi:10.1177/0956797620939054)
 50. Sharockman A. 2014 Fact-checking fox, MSNBC and CNN: PunditFact's network scorecards. <https://www.politifact.com/article/2014/sep/16/fact-checking-fox-msnbc-and-cnn-punditfacts-network/>.
 51. D'Alessio D, Allen M. 2020 Media bias in presidential elections: a meta-analysis. *J. Commun.* **50**, 133–156. (doi:10.1111/j.1460-2466.2000.tb02866.x)
 52. Banas JA, Richards AS. 2017 Apprehension or motivation to defend attitudes? Exploring the underlying threat mechanism in inoculation-induced resistance to persuasion. *Commun. Monogr.* **84**, 164–178. (doi:10.1080/03637751.2017.1307999)
 53. Sagarin BJ, Cialdini RB, Rice WE, Serna SB. 2002 Dispelling the illusion of invulnerability: the motivations and mechanisms of resistance to persuasion. *J. Pers. Soc. Psychol.* **83**, 526–541. (doi:10.1037/0022-3514.83.3.526)
 54. Lewandowsky S, Yesilada M. 2021 Inoculating against the spread of Islamophobic and radical-Islamist disinformation. *Cogn. Res.* **6**, 57. (doi:10.1186/s41235-021-00323-z)
 55. Roozenbeek J, Traber CS, van der Linden S. 2022 Technique-based inoculation against real-world misinformation. [FigShare](https://figshare.com/collections/m9.figshare.c.5978996). (doi:10.6084/m9.figshare.c.5978996)