

**Long-Term Effectiveness of Inoculation Against Misinformation:
Three Longitudinal Experiments**

Rakoen Maertens¹, Jon Roozenbeek^{1,2}, Melisa Basol¹, and Sander van der Linden¹

¹ Department of Psychology, University of Cambridge

² Department of Slavonic Studies, University of Cambridge

Author Note

Rakoen Maertens  <https://orcid.org/0000-0001-8507-5359>, Jon Roozenbeek 
<http://orcid.org/0000-0002-8150-9305>, Melisa Basol  <https://orcid.org/0000-0003-1480-6462>,
Sander van der Linden  <https://orcid.org/0000-0002-0269-1744>

We have no conflict of interest to disclose.

This work was supported by the UK *Economic and Social Research Council (ESRC)* and the *Cambridge Trust*.

The data that support the findings of this study are openly available on the OSF at <https://osf.io/2dtkb/> (DOI: <https://doi.org/10.17605/OSF.IO/2DTKB>).

Correspondence concerning this article should be addressed to Rakoen Maertens, Department of Psychology, University of Cambridge, Downing Street, CB2 3EB Cambridge, United Kingdom. Email: rm938@cam.ac.uk

Abstract

This study investigates the long-term effectiveness of active psychological inoculation to build resistance against misinformation. Using three longitudinal experiments (two pre-registered), we tested the effectiveness of *Bad News*, a real-world intervention in which participants develop resistance against misinformation through exposure to misinformation techniques. In three experiments ($N_{\text{Exp1}} = 151$, $N_{\text{Exp2}} = 194$, $N_{\text{Exp3}} = 170$), participants played either *Bad News* (inoculation group) or *Tetris* (gamified control group) and rated the reliability of news headlines that either used a misinformation technique or not. We found that participants rate fake news as significantly less reliable after intervention. In Experiment 1, we assessed participants at regular intervals to explore the longevity of this effect and found that the inoculation effect remains stable for at least three months. With Experiment 2, we sought to replicate these findings without regular testing and found significant decay over a two-month time period so that the long-term inoculation effect was no longer significant. In Experiment 3, we replicated the inoculation effect and investigated whether long-term effects could be due to item-response memorisation or the fake-to-real ratio of items presented, but found that this is not the case. We discuss implications for inoculation theory and psychological research on misinformation.

Keywords: inoculation theory, fake news, misinformation, resistance to persuasion

Public Significance Statement

This study shows that inoculation-based media and information literacy interventions such as the *Bad News Game* can confer protection against the influence of misinformation over time. With regular assessment, the positive effects can be maintained for at least three months. Without regular “boosting,” the effects dissipate within two months.

Long-Term Effectiveness of the Fake News Vaccine: Three Longitudinal Experiments

Fake news can pose a serious threat to science, society, and democracy (Lewandowsky et al., 2017) with false content spreading faster and deeper on social networks than accurate or factual news (Petersen et al., 2019; Vosoughi et al., 2018). Although fake news may not usually constitute a majority of people's media diet (Allen et al., 2020), including during elections (Allcott & Gentzkow, 2017; Bovet & Makse, 2019; Grinberg et al., 2019), the risk can nonetheless be substantial. For example, the spread of false child abduction rumours on WhatsApp has led to deadly mob lynchings (Arun, 2019). Recent viral misinformation about the COVID-19 pandemic has led to the spread of dangerous health recommendations such as drinking bleach (Frenkel et al., 2020) and conspiracies about 5G networks worsening or causing COVID-19 symptoms have been associated with violent intentions (Jolley & Paterson, 2020) and contributed to people vandalising at least 50 phone masts in the UK alone (K. Chan et al., 2020). Accordingly, psychological research has seen a renewed interest in evaluating effective methods to counteract persuasion by (online) misinformation (Lazer et al., 2018).

Research on the social and cognitive correlates of belief in fake news has flourished, finding that although ideological motivations play a role in the perception and dissemination of misinformation (Grinberg et al., 2019; Guess et al., 2019; Jost et al., 2018; Swire, Berinsky, et al., 2017; van der Linden et al., 2020), higher cognitive ability and analytical thinking are generally associated with reduced belief in fake news (Bago et al., 2020; Bronstein et al., 2019; De keersmaecker & Roets, 2017; Pennycook & Rand, 2019, 2020; Swire, Berinsky, et al., 2017). Specifically, the finding that cognitive ability is strongly associated with susceptibility to misinformation opens up opportunities for the development of interventions. Accordingly, over

the past years, researchers across disciplines have focused on creating solutions to effectively combat misinformation. One predominant approach focuses on the efficacy of debunking and debiasing (Lewandowsky et al., 2012). Debunking, however, can be difficult, as fact-checks and corrections about contested issues may fail in light of (politically) motivated cognition (Flynn et al., 2017). Although the prevalence of the worldview backfire-effect is now increasingly debated (see Ecker, Lewandowsky, Fenton, et al., 2014; Ecker & Ang, 2019; Guess & Coppock, 2018; Wood & Porter, 2019), the continued influence effect (CIE) of misinformation can still limit the effectiveness of debunking techniques (Lewandowsky et al., 2012). Once exposed to a falsehood, it is difficult to correct, as people will often continue to rely on debunked information even when they acknowledge a correction (M.-P. S. Chan et al., 2017; Swire, Ecker, et al., 2017). Moreover, even when debunking is successful, regular exposure to misinformation can increase its perceived accuracy (Pennycook et al., 2018; Swire, Ecker, et al., 2017). Lastly, because fake news tends to spread faster and deeper than other types of news, fact-checkers continually remain behind the curve (Vosoughi et al., 2018).

As such, an attractive alternative approach to debunking is *prebunking* or “inoculating” individuals against future persuasion attempts. Hornsey and Fielding (2017) propose a “jiu-jitsu” analogy of defence against persuasion attacks, which involves using the weight of an opponent against themselves. Inoculation follows a similar approach: by becoming familiar with persuasion techniques, people can protect themselves from being persuaded by misinformation.

Inoculation Theory

A Vaccine for Brainwash – William J. McGuire (1970, p. 36)

The “grandparent theory of resistance to attitude change” is inoculation theory (Eagly & Chaiken, 1993, p. 561). The process of inoculation follows a biomedical immunisation analogy, where exposure to a weakened strain of a pathogen triggers the production of antibodies to confer protection against future infection. In a similar fashion, inoculation theory posits that people can build up cognitive resistance against unwanted persuasion attempts through “prebunking”, i.e. by pre-emptively exposing people to weakened doses of persuasive arguments (Compton, 2013; McGuire, 1961, 1973; McGuire & Papageorgis, 1962). Over 50 years of research has shown that inoculation is among the most effective frameworks to help people resist persuasion attempts (Banas & Rains, 2010; Compton & Pfau, 2005). The inoculation procedure includes two components: *forewarning* and *refutational preemption*, which influence both cognitive and affective processes. Participants build up a set of skills to refute counterarguments and are made aware that their attitudes are vulnerable to more attacks in the future (creating a sense of *threat*; Compton & Pfau, 2005). The operationalisation of threat was traditionally left implicit (“inherent threat”) and was theorised to be elicited by refutational preemption of counterarguments (Pfau et al., 1997), while the explicit forewarning (“extrinsic threat”) was only introduced at a later stage (McGuire, 1964). More recent developments point towards an affective response to forewarning-induced threat which enhances resistance (Compton & Ivanov, 2014). Whether threat is a vital component for inoculation or not is actively debated (Banas & Rains, 2010; Banas & Richards, 2017; Compton, 2009). Originally it was argued that threat was essential (McGuire, 1964; McGuire & Papageorgis, 1962), especially to distinguish inoculation from two-sided messages (Compton & Pfau, 2005; Miller et al., 2013). Although some scholars have indeed demonstrated the importance of the role of threat (Compton & Ivanov, 2012;

Richards & Banas, 2018), others have argued that this may not be a crucial component for conferring resistance to persuasion (Banas & Rains, 2010; Compton, 2009).

Although McGuire's own interpretation of the inoculation theory focused primarily on bolstering (existing) positive attitudes toward cultural truisms (e.g., brushing your teeth after a meal), contemporary inoculation scholarship now distinguishes between purely prophylactic and *therapeutic* inoculation approaches (Compton, 2019; van der Linden & Roozenbeek, 2020). In fact, scholars have argued that the inoculation analogy should be "more instructive than prescriptive" (Compton, 2013, p. 233), and that "the therapeutic inoculation analogy can inspire a new generation of inoculation research" (Compton, 2019, p. 10). Just as therapeutic vaccines can still suppress infection by boosting the immune response, research has shown that people can also be inoculated against misinformation even when the message is not congenial to their prior attitudes, such as in the context of misinformation about climate change (Cook et al., 2017; Maertens et al., 2020; van der Linden et al., 2017). In a recent therapeutic intervention, Roozenbeek and van der Linden (2019) found that the largest inoculation effects were observed among those who were most susceptible to fake news prior to the intervention. Moreover, in an attempt to make inoculation theory scalable beyond specific issues, a second innovation has been a move away from argument-specific *narrow-spectrum* inoculations to *broad-spectrum* inoculations that focus on conferring resistance against a range of common *techniques* used in the production of misinformation (Basol et al., 2020; Cook et al., 2017; Roozenbeek, van der Linden, et al., 2020; Roozenbeek & van der Linden, 2018; van der Linden & Roozenbeek, 2020). In fact, McGuire (1961) himself hypothesised that one important factor in increasing the scope of protection was the notion of "active" rather than passive inoculation. In the active form of

inoculation, participants have to generate their own “antibodies” or counterarguments. One example of active inoculation in the context of fake news is the *Bad News* game, a popular intervention which has been played by over a million people worldwide, and has been translated into more than 17 languages in collaboration with the UK Foreign and Commonwealth Office (Roozenbeek, van der Linden, et al., 2020).¹

Bad News Game

The *Bad News Game* is a real-world online intervention designed by Roozenbeek and van der Linden (2019) in collaboration with the Dutch media platform DROG, based on the principles of inoculation theory. In this free browser game, players enter a simulated social media environment and take on the role of a fake news producer. They design Twitter posts, news article headlines, and memes to gain popularity as a news publisher (see Figure 1 for an in-game screenshot). Players must gain followers while maintaining a sufficiently high level of credibility. If the credibility meter drops too low, the player loses, and the game ends. This way, the player is forced to think actively about how one can be deceived.

Often using a combination of humour and entertainment, the purpose of the intervention is to expose people to severely weakened doses of the techniques commonly used in the production of online misinformation. The game features six specific misinformation techniques known as DEPICT (the “six degrees of manipulation”), including *Discrediting opponents* (e.g., creating a cloud of doubt around your opponent), *appealing to Emotion* (e.g., the use of outrage or highly emotive language to manipulate people), *Polarising audiences* (e.g., using hot-button issues to drive a wedge between two groups), *Impersonation* (e.g., misusing the identity of

¹ The online game is free and publicly available at www.getbadnews.com.

politicians, experts, or celebrities online), *floating Conspiracy theories* (e.g., casting doubt on mainstream narratives by providing an attractive story in which a small sinister group of people is responsible for doing harm to many), and *Trolling* (e.g., eliciting reactions from people by provoking them online). See Roozenbeek and van der Linden (2019) and van der Linden and Roozenbeek (2020) for a detailed background and overview of these techniques.

The game was designed to incorporate the components necessary for inoculation (Roozenbeek, van der Linden, et al., 2020). During gameplay, the player is required to imagine how misinformation techniques could be refuted, which serves as the active *refutational* element of the inoculation. The scenarios were designed to provide participants with a slightly uncomfortable feeling (as they are responsible for creating and sharing fake news), thus eliciting a sense of threat.² As opposed to being issue-based, threat in broad-spectrum inoculation is understood as making the dangers of the spread of fake news salient (by exposure to weakened doses). In fact, the game scenarios themselves incorporate a strong *forewarning* component to foreshadow how fake news can have damaging consequences. For example, participants are explicitly warned about how emotions can be exploited in the media or that “conspiracy theories can be a great way of spreading disinformation”. Threat is also elicited directly by attacks from other simulated “users” through a wide range of social media content. For example, when players choose options that are not in line with the purpose of the game, motivation to do so is boosted by issuing a warning that elevates the threat level; “Whoops, we’re running into a bit of a problem, some ‘fact-checker’ has taken notice ... seriously you need to have a look at this.”

² A prior study analysed open-ended responses as part of a post-gameplay survey and found that the game elicits more (negative) affect compared to a control group (Roozenbeek & van der Linden, 2018).

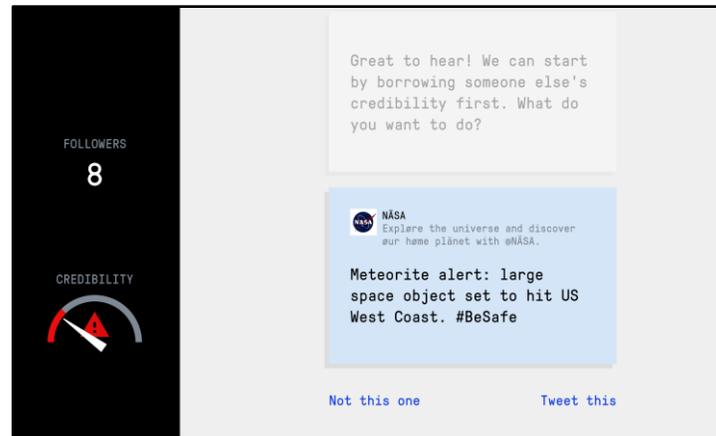


Figure 1. Screenshot of *Bad News* Game Environment.

In their original study, Roozenbeek and van der Linden (2019) used a within-subjects design to evaluate the efficacy of the *Bad News Game* as a “broad-spectrum vaccine” against fake news. In their study, about $N = 15,000$ participants rated the reliability of several fake and real news items (in the form of fictitious Twitter posts) pre and post gameplay. Notably, these were different items than people were trained on in the game (i.e., a refutational-*different* approach to inoculation). The researchers found that while the *fake news* items corresponding to the misinformation techniques were rated as significantly less reliable after playing the game ($d_{\text{average}} = -0.52$), people did not meaningfully adjust their ratings of the *real news* items. Subsequent experiments have shown that the *Bad News* intervention also boosts *confidence* in people’s truth-discernment abilities (Basol et al., 2020) and that the inoculation effect generalises across different cultural contexts (Roozenbeek, van der Linden, et al., 2020). Yet, importantly, nothing is currently known about the duration of the inoculation effect conferred through *Bad*

News, which is a crucial factor in not only determining the long-term efficacy of the intervention, but also in advancing our understanding of inoculation theory and immunity to persuasion.

Longevity

Although the effectiveness of inoculation theory has been well established, research on its long-term effectiveness remains an area with many open questions (Banas & Rains, 2010). Importantly, the rate of the decay of the treatment effect of the *Bad News* intervention is therefore not only of practical utility: it is also a question of high theoretical significance. However, there is currently no coherent theoretical framework that accurately predicts a specific decay³ function of resistance to persuasion.

Although McGuire (1964) initially argued that a delay of a few days between inoculation and attack is needed in order to build up sufficient “mental antibodies”, the more consistent finding points towards the opposite: decay in the inoculation effect over time (Banas & Rains, 2010; Ivanov, 2012). Research indicates that this decay might be slower than the decay found when using other methods such as narrative messaging (Niederdeppe et al., 2015) or consensus messaging (Maertens et al., 2020). A recent study into the effectiveness of a digital media literacy intervention found over 50% decay of the effect over three weeks (Guess et al., 2020). While some studies show that the inoculation effect decays within one or two weeks (Zerback et al., 2020), other findings suggest that inoculation effects can remain detectable for up to six weeks (Pfau et al., 2004, 2006). A meta-analysis suggested an unchanged (stable) effect with a

³ Throughout this paper, we use the word *decay* as a theory-neutral description of the decrease in the inoculation effect over time, unless otherwise specified. The term ‘decay’ is often used in inoculation research in general terms without making claims about whether the decay is *due to* the mere passage of time, or related to memory function. In contrast, in memory research, *decay* often refers to *trace decay*, a specific memory theory (Brown, 1958).

duration of at least two weeks, followed by a decay of the effect after this period of stability. The most recent study to systematically explore inoculation decay at multiple time points found that decay started between four to six weeks after intervention (Ivanov et al., 2018). The longest retention figures suggest that some inoculation effects could sustain over a period of 33 weeks (Pfau et al, 1992). Yet, it remains unclear whether the inoculation decay function is continuous or intermittent; linear, curvilinear or exponential; and if the decay function can take different forms under specific circumstances (Banas & Rains, 2010; Compton & Pfau, 2005; Ivanov, 2017).

The *Bad News* intervention is a particularly interesting test case as active inoculation is meant to stimulate analytical thinking and strengthen linkages between nodes in associative memory networks, which are thought to both facilitate resistance to persuasion and improve the longevity of the inoculation effect (Banas & Rains, 2010; Pfau et al., 1997, 2005). For example, based on network models of memory (Anderson, 1983; Forgas, 2001), Pfau et al. (2005) theorised that resistance to persuasion might nest itself in long-term memory networks. They argued that an attitude could be represented as an associative memory network with cognitive and affective nodes. Based on Petty et al. (1994), they posited that a more dense network would be more resistant to change. Using concept mapping as a method to represent mental structures, they found increases in relevant nodes and linkages after an inoculation message, which in turn led to more resistance to persuasion attacks at a later date (Pfau et al., 2005).⁴

⁴ With concept mapping participants have to draw a map similar to a mind map. Participants are asked to think about and write down everything related to a central topic (i.e., the inoculation topic). The different nodes (circles with concepts) and the links between the nodes they draw, count as the density of the memory network.

More generally, to counter the decay of the inoculation effect, evidence has been found for the effectiveness of *booster treatments* (Ivanov et al., 2018). It is theorised that similar to biomedical inoculations, a regular “booster shot” may be needed to top up the cognitive immune system (McGuire, 1961). Examples of booster messages include a weakened attack message, a repetition of the inoculation procedure (in full or shortened form), or a new warning message to elicit a fresh sense of threat (Ivanov et al., 2018). While evidence on the effectiveness of booster treatments is mixed (Compton & Pfau, 2005; Ivanov et al., 2009; Pfau, 1992), the general conclusion is that boosters work when administered in the right form at the right time (Ivanov, 2012; Ivanov et al., 2018; McGuire, 1961; Pfau et al., 2004). Further, the concept of booster treatments could be interpreted through a memory lens as *relearning*, leading to stronger memory representations (Ebbinghaus, 1885; Murre & Dros, 2015). Ivanov et al. (2018, p. 661) stress that “the book on boosters is not ready to be closed” and that we need to “reignite the research interest in inoculation booster messages.”

The Present Research

Since prior evaluations of the *Bad News* intervention were not pre-registered, the first goal of the current study was to replicate the original effect of the *Bad News* intervention in a pre-registered experimental study with a larger battery of fake news test items. Based on prior work, we expected to replicate the main effect of the intervention.

H₁: *On average, participants in the inoculation group rate fake news (post - pre) as significantly less reliable compared to (post - pre) ratings of the same items in the control group.*

Given a paucity of research on the longevity of inoculation interventions, we advance the literature by measuring the effectiveness of the inoculation intervention over time. Although there is no clear theory that would predict the longevity of the inoculation effect, based on the work reviewed above, we can conclude that the inoculation effect decays over time. Based on the meta-analysis finding of decay starting at some point after two weeks (Banas & Rains, 2010), and a recent study showing decay setting in between four to six weeks (Ivanov et al., 2018), we hypothesise that the decay process should happen within the timeframe of two months. This led to our second (decay) hypothesis.

H₂: *After two months, participants show a significant decrease in the inoculation effect.*

General Method

Design and Procedure

For Experiment 1 and Experiment 2, we utilised a randomised pretest-posttest design (Campbell, 1957; Huck & McLean, 1975). Participants were randomly allocated to either the inoculation group or the control group. In Experiment 3, all participants received the inoculation intervention. In all three experiments, participants started with a pretest survey. In this survey, participants had to judge the reliability of news items (21 in Exp 1-2, seven in Exp 3) that were either factual news headlines (three in Exp 1-2, one in Exp 3) or headlines featuring a misinformation technique (18 in Exp 1-2, six in Exp 3). All participants had to rate the reliability of the news headlines on a Likert scale from 1 (very unreliable) to 7 (very reliable). All items were presented in random order. After rating the news items at pretest (T1), participants were asked to complete *Bad News* (inoculation group), or to play ~15 minutes of *Tetris* (control

group). The 15-minute time slot was chosen to match the completion time of the *Bad News* game. After the intervention, participants were asked to rate the reliability of the same headlines again (T2). In Experiment 1 and Experiment 3, participants were directed to a demographics survey after the posttest, and answered questions about their year of birth, gender, political affiliation (from 1-7, very left-wing to very right-wing), country of residence, first language, social media usage (from 1-5, never to daily), and had to respond to a single-item cognitive reflection test: “*A ball and a bat cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?*” (Frederick, 2005). All participants received a debriefing at the end.

The same participants were then contacted again (T3) to participate in the follow-up where they had to rate the reliability of the same (Exp 1-2) or different (Exp 3) news headlines. T3 took place one week after the initial test date for Experiments 1 and 3, and two months after the initial test date in Experiment 2. Participants then received an extended debriefing. Participants in Experiment 1 were asked to participate in another follow-up four weeks after the second posttest (T4), and then again eight weeks after this (three months after the initial test date; T5). For consistency, all items were kept the same throughout the different test dates in Experiment 1 and Experiment 2, while Experiment 3 specifically investigated whether using the same (versus different) items produces a confound. See Figure 2 for a visualisation of the study flowchart depicting the experiments’ design.

All three experiments were approved by the Cambridge Psychology Research Ethics Committee (ref. PRE.2018.085, PRE.2019.103). Our pre-registered hypotheses can be found on the AsPredicted platform (Experiment 1: <https://aspredicted.org/xn6qy.pdf>, Experiment 3: <https://aspredicted.org/ka2at.pdf>). All deviations from the original pre-registration can be found

in Supplementary Declaration S1. All datasets, measurement items, and analysis scripts for Experiment 1, 2, and 3 are publicly available on our OSF repository:

<https://doi.org/10.17605/OSF.IO/2DTKB>.

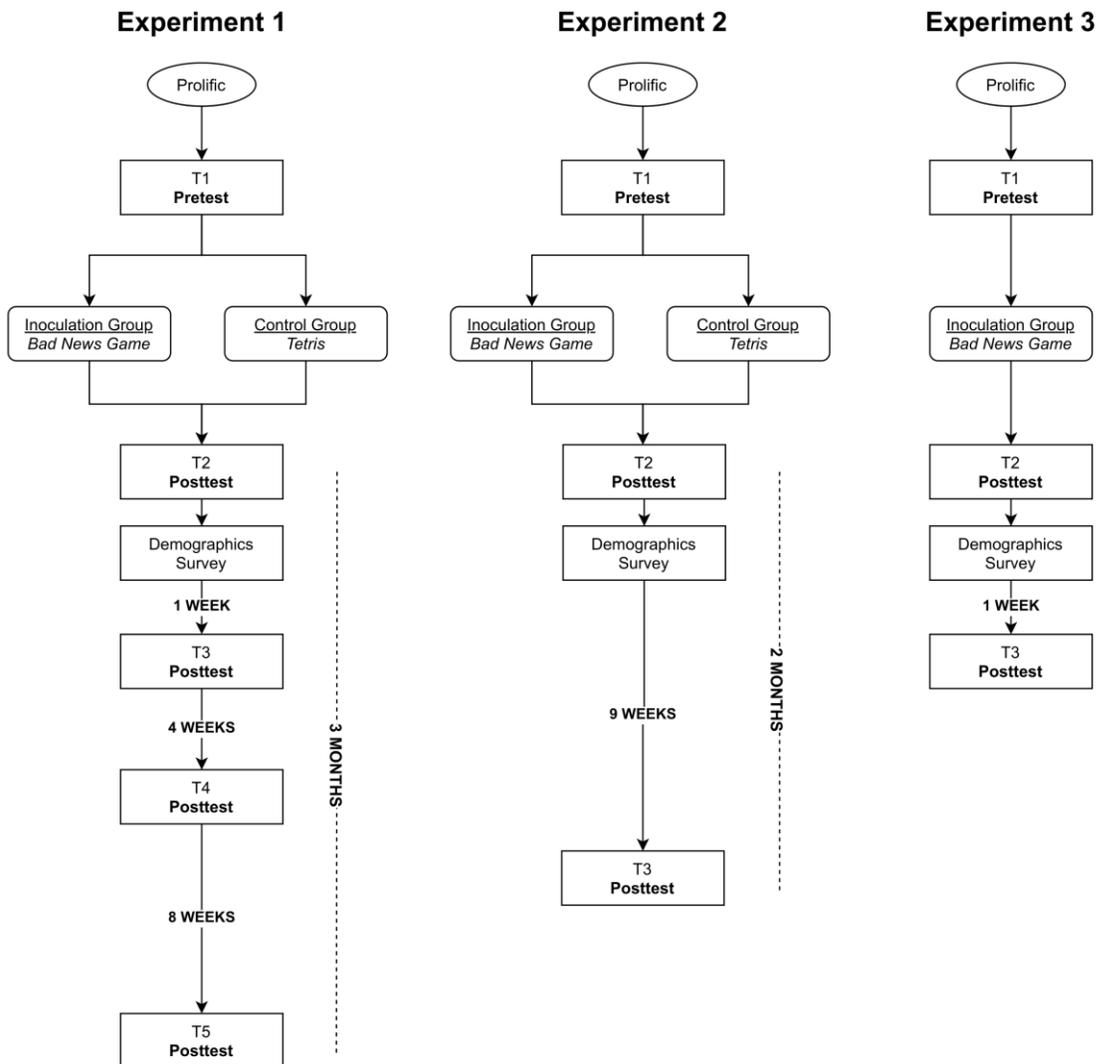


Figure 2. Overview of Experiment flowchart of Experiment 1, 2, and 3.

Measures

In the initial *Bad News* study by Roozenbeek and van der Linden (2019), only one fake news item per misinformation category was evaluated by the participants, and the items were not

randomised. To make the measure stronger and to avoid the possibility of item-effects (Roozenbeek, Maertens, et al., 2020), we developed three manipulative news headlines per misinformation technique, plus three factual ‘real news’⁵ control headlines. Importantly, whilst modelled after real-world events, the test items were a) fictional and b) different from those used in the inoculation training itself. Participants could, therefore, not just rely on recognition but were required to detect the misinformation strategy in a new setting. Using fictional items also maximises experimental control over isolating the manipulation techniques and avoids familiarity confounds with real fake news (Roozenbeek, Maertens, et al., 2020; Roozenbeek, van der Linden, et al., 2020). The items were designed based on the most commonly used misinformation techniques (using the definitions provided in the original study; Roozenbeek & van der Linden, 2019). An example of a real news item would be “*Physical fitness keeps your brain in good shape,*” posted by “@PsychologyToday.” This item does not contain any misleading strategy, while the fake news items each have their own misinformation element. Examples include “*Scientists discovered solution to greenhouse effect years ago but aren’t allowed to publish it, report claims*” (conspiracy) and “*New study shows right-wing people lie much more than left-wing people*” (polarisation). See Figure 3 for an example of how the items were presented. Consistent with Basol et al. (2020) all analyses were done with an aggregate index of all fake news items ($M = 3.39$, $SD = 0.81$, Cronbach’s $\alpha = 0.83$).⁶

⁵ These were not of primary interest here but included to remain consistent with the approach of Roozenbeek and van der Linden (2019). An overview of all items can be found in Supplementary Information S1.

⁶ This is a deviation from the pre-registration to prevent multiple testing artefacts and increase internal consistency of the measurement. For these reasons, we only use the combined fake news index (18 items) and neither analyse subcategories (3 items per category only) nor real news (3 items only). A scree plot provides evidence for the unidimensionality of the 18 fake news items (only one factor with eigenvalue > 1; see Supplementary Figure S5). Nonetheless, results per subcategory are provided in Supplementary Tables S3-S8 and real news items are plotted in Supplementary Figures S1-S4.

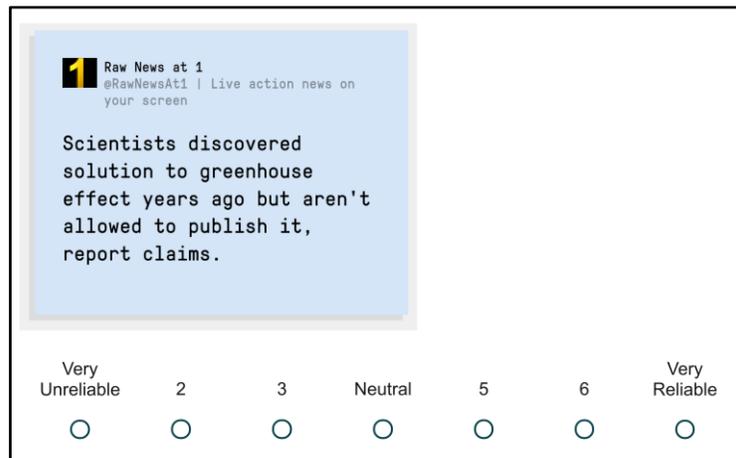


Figure 3. Example test item using the conspiracy technique.

Empirical Strategy

We follow our pre-registered empirical strategy with one change. To help make a clearer distinction between our two hypotheses and make the coefficients interpretable, we separated the analyses for both Experiment 1 and Experiment 2 into a standard analysis of covariance (ANCOVA) for Hypothesis 1 and a repeated-measures analysis of covariance (rmANCOVA) for Hypothesis 2.⁷

To test H_1 , we performed an ANCOVA with *intervention* as a between-subjects factor with two levels: *Bad News* (Inoculation Group) vs. *Tetris* (Control Group), and posttest reliability rating (T2/T3/T4/T5) as the dependent variable. As our primary concern was measuring the difference between the control group and the inoculation group after intervention

⁷ The pre-registration only mentions the rmANCOVA model. As this model is confounded with the time variable, and the pure inoculation effect hypothesis H_1 is unrelated to the time variable, we decided to split these analyses. All deviations from the pre-registration can be found in Supplementary Declaration S1.

without confounds by potential group differences, we modelled the pretest (T1) measure as a covariate (Coppock, 2019).

To test **H₂**, we used the same ANCOVA but with the repeated measures variable time added as a within-subjects factor with two levels: T2 (day 1, after intervention) and T3, T4, or T5, and the interaction between time and the intervention.

In Experiments 1 and 2, the fake-to-real item ratio (18:3) was unbalanced, and the same item sets were used for each assessment. We, therefore, added Experiment 3, a replication of Experiment 1 (up to T3) with different item sets and a balanced item ratio (6:6). To test the validity of the hypothesis tests after taking into account these potential confounds, we followed the pre-registered empirical strategy of using within and between-subject *t*-tests to compare the inoculation effects within and between the two experiments.

Experiment 1

Method

Participants and Sample Composition

In the original large-sample *Bad News* study, an average inoculation effect size of $d = 0.52$ was found for the fake news scale (Rozenbeek & van der Linden, 2019). A power analysis with an effect size of $d = 0.52$, a significance level of .05, .80 power, and taking in account potential participant attrition of 20% over the test period, indicated a required sample size of 148 participants. Based on this, we recruited a total of 151 participants.

Participants were recruited through the online platform Prolific (<https://prolific.ac/>) and were rewarded 2.35 GBP if they participated in both the initial experiment (T1/T2) and the

follow-up one week later (T3). They were unaware of the later follow-ups (T4, T5), but were later invited to participate for 0.25 GBP per bonus follow-up.⁸ All participants gave informed consent before participation. The sample consisted of 151 participants (81 control, 70 intervention), 52% identifying as female, slightly skewed towards younger age (19-66, *Mdn* = 28). The sample was culturally diverse with 23 different countries of which the largest are the United Kingdom (29%), Italy (17%), and Poland (15%). Political ideology was skewed towards left-wing (49% left-wing, 19% right-wing; $M = 3.50$, $SD = 1.25$), and 50% with a higher education level diploma. For a complete overview of the sample demographics separated by the T1 sample and the complete-cases T5 sample, see Supplementary Table S1.

As pre-registered and accounted for in our power analysis, we only use the dataset with complete cases ($N_{\text{all}} = 151$, $N_{\text{complete}} = 118$, attrition = 22%) for our hypothesis tests. We did this to have comparable results between the different test dates. An attrition analysis indicated that no specific demographic factors (e.g., age, education, ideology) could predict the pattern of attrition (see Supplementary Table S9), and pretest reliability could not predict completeness ($b = -0.04$, $OR = 0.96$, 95% $CI_{OR} [0.59, 1.54]$, $p = 0.87$).

Results

Inoculation Effect

The average inoculation effect was significantly stronger in the treatment vs. control group, manifested by a significant main effect of *intervention* in our ANCOVA analysis ($F(1, 114) = 29.86$, $p < .001$, $\eta^2_p = 0.21$, $\eta^2 = 0.07$). Difference-in-differences analysis ($M_{\text{diffT2T1,control}} =$

⁸ The bonus follow-ups were not pre-registered because we originally did not plan them. See Supplementary Declaration S1 for an overview of all deviations from the pre-registered procedure.

-0.08, $SD_{diffT2T1,control} = 0.34$; $M_{diffT2T1,inoculation} = -0.61$, $SD_{diffT2T1,inoculation} = 0.66$) using a post-hoc t -test indicated a significant mean difference of $M_{diff-in-diffs} = -0.52^9$ ($t(84) = -5.41$, $p < .001$, 95% CI [-0.72, -0.33], $d = -1.00$), indicating that participants who played the *Bad News Game* rated the fake news items as significantly less reliable than those who played *Tetris*.¹⁰

Decay: One Week

To test H₂, we first used the same ANCOVA model to verify if the inoculation effect was still significant, but now with the one-week-later posttest (T3) as the dependent variable. We found that the inoculation effect was still significant ($F(1, 114) = 23.91$, $p < .001$, $\eta^2_p = 0.17$, $\eta^2 = 0.07$; see Figure 4, panel A, T3). Using an rmANCOVA with both the immediate posttest (T2) and the posttest one week later (T3) to test the decay hypothesis, no indication for an interaction effect between time and intervention was found ($F(1, 229) = 0.02$, $p = .88$, $\eta^2_p = 0.00$, $\eta^2 = 0.00$).

Decay: Five Weeks

To extend our original hypothesis, we looked at potential decay five weeks after the intervention (T4). Using the same ANCOVA, we found that the inoculation effect was still significant ($F(1, 114) = 27.68$, $p < .001$, $\eta^2_p = 0.20$, $\eta^2 = 0.10$; see Figure 4, panel A, T4). When looking at the rmANCOVA, we again found no significant interaction effect between time and intervention ($F(1, 229) = 0.44$, $p = .51$, $\eta^2_p = 0.00$, $\eta^2 = 0.00$).

Decay: 13 Weeks

We once more extend our scope with a follow-up three months after the initial intervention (T5). The inoculation effect still remained significant ($F(1, 114) = 13.27$, $p < .001$,

⁹ The difference (0.61 - 0.08) was 0.52 and not 0.53 due to rounding. Raw values (no rounding) were used for all calculations.

¹⁰ All raw means and confidence intervals for Experiment 1 are available in Supplementary Table S3-S6.

$\eta^2_p = 0.10$, $\eta^2 = 0.06$; see Figure 4, panel A, T5). The rmANCOVA result showed no significant interaction effect between intervention and time ($F(1, 229) < 0.01$, $p = .98$, $\eta^2_p = 0.00$, $\eta^2 = 0.00$). Using a post-hoc t -test difference-in-differences calculation, we found an inoculation effect retention of 100%¹¹ between T2 and T5 ($M_{diffT5T2,control} = -0.19$, $SD_{diffT5T2,control} = 0.81$; $M_{diffT5T2,inoculation} = -0.19$, $SD = 0.69$; $M_{diffT5T2,inoculation} = 0.00$, $t(114) = 0.00$, $p > .999$, 95% CI [-0.27, 0.27], $d = 0.00$). See Figure 4, panel B, for a summarising density plot for each test date.

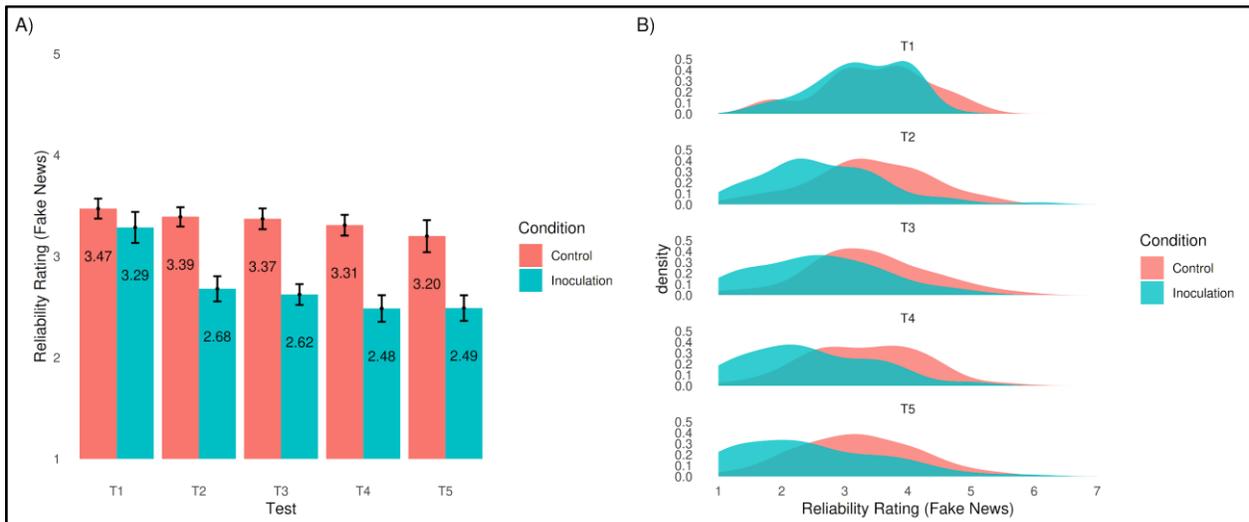


Figure 4.¹² Reliability ratings of fake news items, separated by time and condition in Experiment 1. Panel A: average reliability rating of fake news items over time. Panel B: density plots of these results. T1 = pretest. T2 = posttest (0 weeks). T3 = posttest (1 week). T4 = posttest (5 weeks). T5 = posttest (13 weeks). $N = 118$. Error bars represent 95% Confidence Intervals.

¹¹ Retention formula: $1 - ((Control\ T5 - Control\ T1) - (Inoculation\ T5 - Inoculation\ T1)) / ((Control\ T2 - Control\ T1) - (Inoculation\ T2 - Inoculation\ T1))$.

¹² Alternative plots for Experiment 1 are available in Supplementary Figure S1-S2.

Exploratory

For exploratory purposes, we analysed and visualised a range of extra variables and examined the robustness of the inoculation effect when controlling for individual differences.¹³ Using a linear regression model with the T2-T1 difference score as outcome variable we found that the inoculation effect remained significant, $t(106) = -5.29$, $p < .001$, $\beta = -0.46$, 95% CI [-0.66, -0.26], after controlling for gender, age, country of residence, education level, political ideology, social media use, Twitter use, cognitive reflection test performance, and primary language, consistent with previous results by Roozenbeek and van der Linden (2019).

Discussion

The lack of decay observed in Experiment 1 is contrary to our expectations. We hypothesise that the repeated tests might have confounded the result as they could function as booster sessions or simply testing effects. The regular exposure to weakened doses of each fake news technique (in this case, the measurement items) could serve as reminders that reinstate the inoculation effect. Ivanov et al. (2018) found that for longer time intervals booster messages can indeed prolong the inoculation effect, leading to significant inoculation effects for at least six weeks. These findings can be related to memory research, where a robust literature shows the memory-strengthening impact of repeated testing (Karpicke & Roediger, 2008; Linton, 1975; Roediger & Karpicke, 2006a, 2006b). By repeatedly requiring people to recall what they learned, they relearn these lessons (Nader & Hardt, 2009), potentially leading to an increase in inoculation effect stability over time.

¹³ See Supplementary Table S10 for linear model estimates and Supplementary Analysis S1 for visual plots.

To investigate the inoculation effect retention while eliminating the possibility of learning or boosting effects, we re-run the experiment, removing all intermediate tests to allow for a full two-month decay period without follow-ups (Experiment 2).

Experiment 2

Method

Design and Participants

In a parallel research project, a similar experiment was conducted with the same test items, but without any follow-up after the initial test date. We decided to leverage this opportunity to re-examine the potential for decay. We re-contacted this study's participants two months after their initial test and asked them to participate in an unexpected bonus follow-up, which functions as our Experiment 2. Importantly, the main procedures and reliability measures are the same between Experiment 1 and Experiment 2, with one crucial difference: the different time interval between T2 and T3 (two months instead of one week).¹⁴

In total, $N = 194$ participants (107 control, 87 inoculation) were recruited through Prolific. In the unexpected follow-up two months later (T3), the number of participants was reduced to 110 (56 control, 54 inoculation), a 57% retention of the initial sample. In terms of the sample composition,¹⁵ 57% were male, skewed towards younger age (18-44, modal bracket = 18-24). Political ideology was skewed towards left-wing (59% left-wing, 21% right-wing; $M = 2.40$,

¹⁴ One other notable difference is that two out of three real news items were actually different in Experiment 2 (see Supplementary Information S1 for a comparison).

¹⁵ Note that, as participation in the demographics survey was optional in this experiment and conducted at a different date, education level, age, and gender were only answered by 39% of participants.

$SD = 1.34$), and 26% had a higher education level diploma (modal bracket = *High school diploma*, 51%).

We used the same empirical strategy as in Experiment 1. Similarly, as in Experiment 1, all hypothesis tests were done using complete cases ($N_{\text{all}} = 194$, $N_{\text{complete}} = 110$, attrition = 43%) and on the 18-item fake news scale ($M = 3.20$, $SD = 0.85$, $\alpha = 0.83$).¹⁶

Results

Inoculation Effect

An inoculation effect was found on the initial test date (T2) for the inoculation group ($F(1, 106) = 11.65$, $p < .001$, $\eta^2_p = 0.10$, $\eta^2 = 0.03$). A post-hoc *t*-test for difference-in-differences analysis ($M_{\text{diffT2T1,control}} = -0.08$, $SD_{\text{diffT2T1,control}} = 0.32$; $M_{\text{diffT2T1,inoculation}} = -0.50$, $SD_{\text{diffT2T1,inoculation}} = 0.80$) showed that this difference was a significant decrease in reliability ratings for the inoculation group compared to the control group ($M_{\text{diff-in-diffs}} = -0.42$, $t(69) = -3.62$, $p < .001$, 95% CI [-0.19, -0.65], $d = -0.69$).¹⁷ See Figure 5, panel A, T1 and T2 for a visual comparison.

Decay: Nine Weeks

Using the same analyses as for Experiment 1, we found no inoculation effect for the standalone ANCOVA ($F(1, 106) = 2.17$, $p = .14$, $\eta^2_p = 0.02$, $\eta^2 = 0.01$), and no interaction effect between time and intervention for the rmANCOVA ($F(1, 213) = 2.18$, $p = .14$, $\eta^2_p = 0.01$, $\eta^2 = 0.00$). Using a post-hoc *t*-test difference-in-differences analysis we found a non-significant

¹⁶ As 61% of participants did not complete the extra demographics survey, we performed neither attrition analyses nor exploratory analyses for Experiment 2. However, model estimates and visual plotting of these analyses in Experiment 1 can be found in Supplementary Table S10 and Supplementary Analysis S1, respectively.

¹⁷ All raw means and confidence intervals for Experiment 2 are available in Supplementary Table S7-S8.

inoculation effect retention of 36%¹⁸ between T2 and T3 ($M_{\text{diffT3T2,control}} = -0.01$, $SD_{\text{diffT3T2,control}} = 0.78$; $M_{\text{diffT3T2,inoculation}} = 0.26$, $SD_{\text{diffT3T2,inoculation}} = 0.75$; $M_{\text{diff-in-diffs}} = 0.27$, $t(108) = 1.85$, $p = .07$, 95% CI [-0.56, 0.02], $d = -0.35$).

Plotting these results pointed towards a partial decay hypothesis (see Figure 5, Panel A, T1-T3). We found further evidence for this by visually analysing the distribution of the reliability ratings in the inoculation group (see Figure 5, Panel B, inoculation group), manifested by a dent in the plot indicating that some reverted to baseline while for others the inoculation benefits persisted.

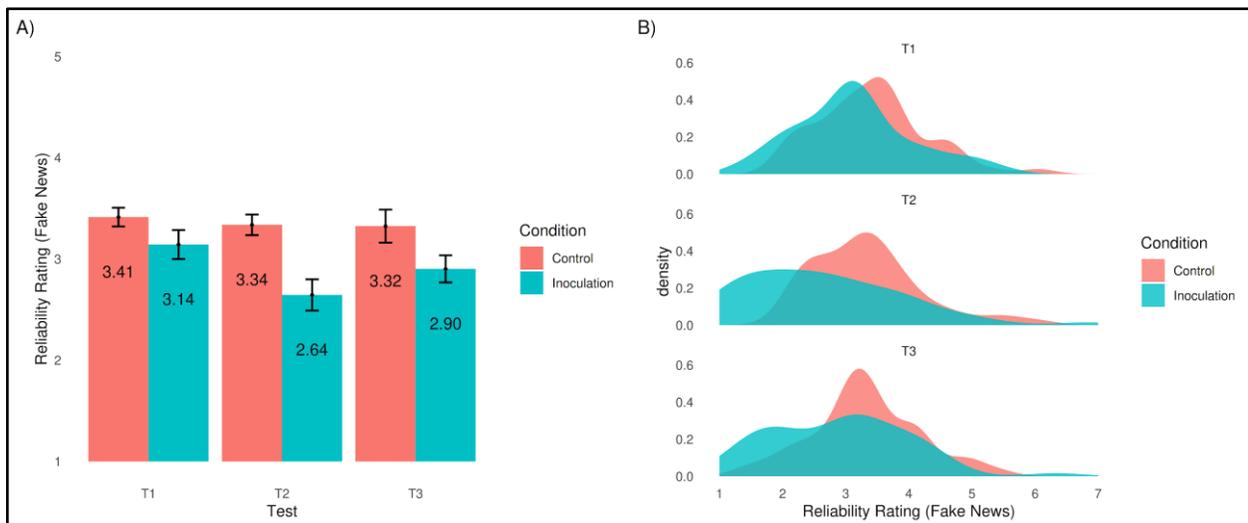


Figure 5.¹⁹ Reliability ratings of fake news items, separated by time and condition in Experiment 2. Panel A: plot of average fake news reliability ratings. Panel B: density plots of the same results. T1 = pretest. T2 = posttest (0 months). T3 = posttest (2 months). $N = 110$. Error bars represent 95% confidence intervals.

¹⁸ Retention formula: $1 - ((\text{Control T3} - \text{Control T1}) - (\text{Inoculation T3} - \text{Inoculation T1})) / ((\text{Control T2} - \text{Control T1}) - (\text{Inoculation T2} - \text{Inoculation T1}))$.

¹⁹ Alternative plots for Experiment 2 are available in Supplementary Figures S3-S4.

Discussion

In Experiment 2, we eliminated the confound of repeated measurement by removing all follow-ups between the direct posttest and the posttest two months later. In line with the original hypothesis, we find that the inoculation effect indeed decays over the course of two months, rendering the effect no longer significant. The analyses also show that the decay is only partial, with density plots suggesting that the effect might still linger on for some participants. A final set of concerns left unanswered by the previous two experiments is whether confounds are introduced because of (a) the unbalanced fake-to-real ratio (18:3) of the presented news items (Aird et al., 2018) and (b) the fact that the same items were used at each follow-up, which may lead to item-response memorisation effects. Experiment 3 aims to rule out these alternative explanations.

Experiment 3

Method

Design and Participants

In Experiment 3, we explored whether the sustained effects in Experiment 1 could be due to either the memorisation of responses to the items (all items were the same for each test date) or due to the skewed ratio (18:3) of fake-to-real news items (Aird et al., 2018). To accomplish this, we designed an experiment that was identical to Experiment 1 (up to T3, the first follow-up) but changed both the item set and fake-to-real ratio for the follow-up measure. In this pre-registered experiment²⁰, we omitted the control group, as we wanted to maximise power and

²⁰ <https://aspredicted.org/ka2at.pdf>. Any deviations can be found in Supplementary Declaration S1.

because our core comparison of interest was the inoculation group. This design allowed us to compare the results of Experiment 3 to those of Experiment 1, to find out whether the T3 results are the same now that two confounds (item set repetition and fake-to-real ratio) have been eliminated. See Figure 6 for a comparison of the two experimental designs, and Supplementary Repository S1 for the precise item sets.

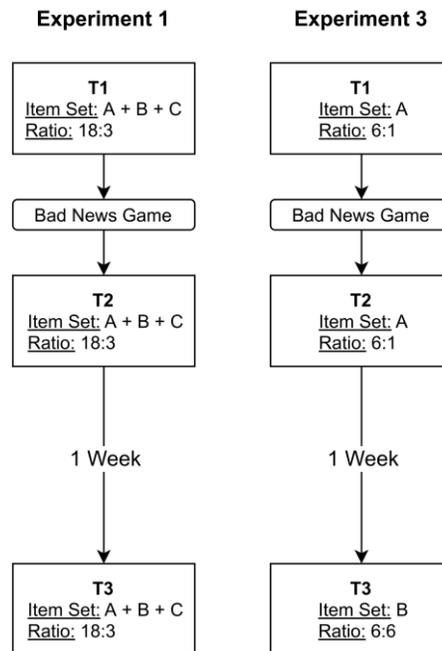


Figure 6. Flowcharts of Experiment 1 (up to T3, excluding control group) and Experiment 3, with item set information. Item sets are news sets with six fake news items and one to six real news items each. Ratio refers to fake-to-real ratio.

We conducted a power analysis with power = 0.80, $\alpha = 0.05$, $d = 0.45$ (SESOI), expected attrition = 10%, and $N_{\text{Exp1}} = 70$. We recruited 100 participants from Prolific. Participants in any

previous *Bad News* experiments were barred from participation. We followed the same data cleaning procedures as for Experiment 1. Thirteen people dropped out for T3, making the final sample $N_{\text{Exp3}} = 87$. Our final sample was younger ($Mdn = 22$, 84% between 18-29), predominantly male (75% male, 23% female), more left-wing ($M = 3.45$, $SD = 1.41$), educated (45% with higher education diploma), and most participants came from Poland (29%) or Portugal (28%).

Results

Within-Group

As pre-registered, we first looked at whether the inoculation effect is present for each time point. When comparing T2 ($M = 2.83$, $SD = 1.09$) to T1 ($M = 3.48$, $SD = 0.88$), we found a significant negative effect with $M_{\text{diff},T2T1} = -0.65$, 95% $CI_M [-0.84, -0.46]$, $t(86) = -6.70$, $p < .001$, $d = -0.72$, 95% $CI_d [-0.95, -0.48]$. This effect shows that a medium-to-large baseline effect was established using the same item set (Set A).

We also compared T3 ($M = 2.79$, $SD = 0.98$) to T1 ($M = 3.48$, $SD = 0.88$), and found a near-identical significant effect with $M_{\text{diff},T3T1} = -0.70$, 95% $CI_M [-0.90, -0.50]$, $t(86) = -6.87 = p < .001$, $d = -0.74$, 95% $CI_d [-0.97, -0.50]$. We thus found a significant medium-to-large effect of the inoculation intervention using Set B, indicating that the intervention was effective despite using a different item set and after equalising the fake-to-real ratio. See Supplementary Tables S12 and S13 for an overview of the raw means and difference scores for each time point and each item set.

Between-Groups

As pre-registered, the next step in our decision tree was to compare the within-group difference scores between both experiments, to explore if, despite the inoculation effect remaining significant, the altered experiment design influenced the treatment effect. We first looked at the T2-T1 difference in Exp 3 ($M_{\text{diff},T2T1} = -0.65$, $SE = 0.10$) compared to Exp 1 ($M_{\text{diff},T2T1} = -0.68$, $SE = 0.10$), and found no significant difference between the two groups with $M_{\text{diff-in-diffs}} = 0.03$, 95% CI_M [-0.24, 0.31], $t(153) = 0.23$, $p = .82$, $d = -0.04$, 95% CI_d [-0.28, 0.35]. This difference was also statistically equivalent to zero ($t(153) = -2.59$, $p = .005$);²¹ we could therefore conclude that the baseline effect was the same between the two samples.

We then compared T3-T1 difference in Exp 3 ($M_{\text{diff},T3T1} = -0.70$, $SE = 0.10$) to Exp 1 ($M_{\text{diff},T3T1} = -0.91$, $SE = 0.14$), and found no significant difference between the two experiments with $M_{\text{diff-in-diffs}} = 0.21$, 95% CI_M [-0.12, 0.55], $t(134) = 1.25$, $p = .21$, $d = -0.20$, 95% CI_d [-0.11, 0.52]. Although this effect was not significant, it was not statistically equivalent to zero at the traditional α level ($t(133) = 1.53$, $p = 0.06$). These findings indicated that there was no significant increase in reliability ratings of fake news by changing the experimental design for the T3 follow-up one week later (although a small increase could not be ruled out).

Finally, looking at the T3-T2 difference in Exp 3 ($M_{\text{diff},T3T2} = -0.06$, $SE = 0.10$) compared to Exp 1 ($M_{\text{diff},T3T2} = -0.23$, $SE = 0.11$), we did not find a significant difference with $M_{\text{diff-in-diffs}} = 0.18$, 95% CI_M [-0.11, -0.47], $t(152) = 1.21$, $p = 0.23$, $d = 0.19$, CI_d [-0.12, 0.51]. Equally, although the comparison was not significantly different, it was not statistically equivalent to zero

²¹ We used Two One-Sided Tests (TOST) Equivalence Testing using the TOSTER package in R with $\alpha = 0.05$ and as Smallest Effect Size of Interest (SESOI) $d = (-)0.45$.

($t(152) = -1.60, p = 0.06$). This indicated that the inoculation retention over a one-week period was similar between the two experimental setups, thereby finding no evidence for item ratio or item set specific retention effects. See Figure 7 (Panel A) for a bar chart comparing the two experiments.

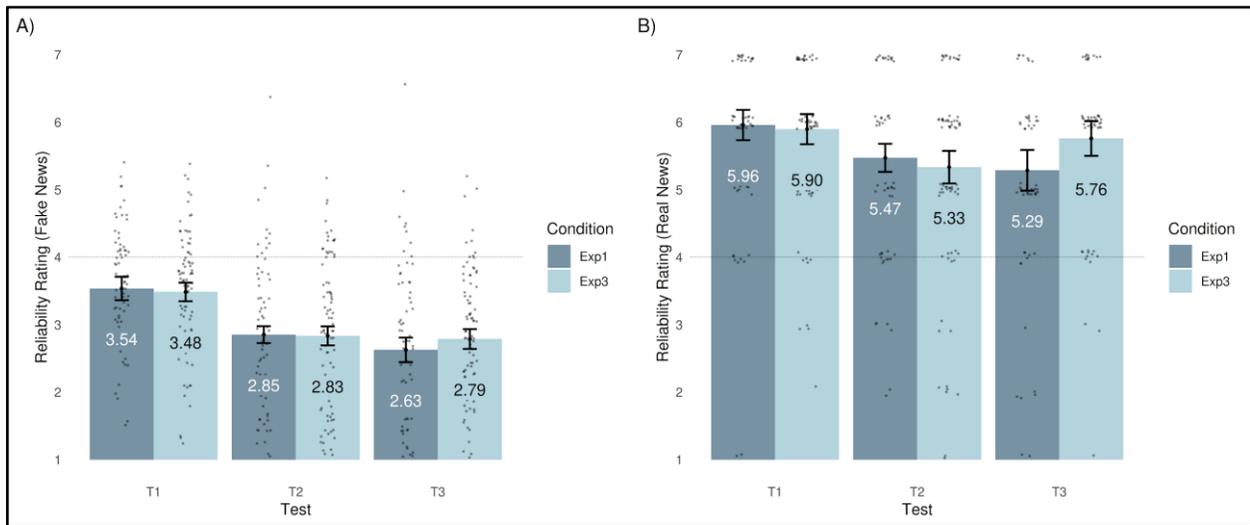


Figure 7. Comparison of reliability ratings of Experiment 3 to Experiment 1. Panel A depicts fake news ratings; Panel B depicts real news ratings. Only items overlapping between both experiments are shown. Horizontal line reflects binary fake (< 4) or real (> 4) classification threshold. Error bars represent 95% confidence intervals. $N = 157$.

Exploratory

Although not pre-registered, we also looked at the real news items. All seven real news items were rated as very reliable (> 4/7) before intervention, immediately after intervention, and one week after intervention. We then compared the two overlapping items that were used in both

experiments and found no significant differences between T2-T1 difference scores in Exp 3 ($M_{\text{diff},T2T1} = -0.56$, $SE = 0.16$) and Exp 1 ($M_{\text{diff},T2T1} = -0.49$, $SE = 0.13$), with $M_{\text{diff-in-diffs}} = -0.08$, 95% CI_M [-0.48, 0.32], $t(153) = -0.38$, $p = .70$, $d = -0.06$, 95% CI_d [-0.37, 0.25], with statistical equivalence to zero ($t(153) = 2.47$, $p = .007$). This indicated that also for real news, the baseline effect between both experiments was the same.

When comparing the real news T3-T1 difference-in-differences scores between Exp 3 ($M_{\text{diff},T3T1} = -0.14$, $SE = 0.17$) and Exp 1 ($M_{\text{diff},T3T1} = -0.67$, $SE = 0.20$), we found a significant positive effect with $M_{\text{diff-in-diffs}} = 0.53$, 95% CI_M [0.01, 1.05], $t(145) = 2.02$, $p = .045$, $d = 0.33$, 95% CI_d [0.01, 0.64]. A comparable result was found when comparing T3-T2 difference-in-differences scores between Exp 3 ($M_{\text{diff},T3T2} = 0.43$, $SE = 0.18$) and Exp 1 ($M_{\text{diff},T3T2} = -0.19$, $SE = 0.19$), which showed a significant effect with $M_{\text{diff-in-diffs}} = 0.61$, 95% CI_M [0.09, 1.13], $t(151) = 2.31$, $p = .022$, $d = 0.37$, 95% CI_d [0.05, 0.69]. These analyses demonstrated higher reliability ratings for real news at T3 in Exp 3 (where design confounds were removed) as compared to Exp 1 (see Figure 7, Panel B).

Discussion

In Experiment 3, we investigated whether the effects found in Experiment 1 were confounded by the ratio of fake-to-real items, and the repeated use of the same item set. Although we only looked at a time period of one week after the intervention, the results show that there is no significant difference between the results of Experiment 3 and the results of Experiment 1 for fake news. Meanwhile, consistent with Roozenbeek, Maertens, et al. (2020), exploratory analyses indicated that removing the confounds had improved the reliability rating of

the real news item.²² We can, therefore, reasonably conclude that while there may be some longer-term effects of design choices that are not measured here, the findings presented in Experiment 1 and Experiment 2 are unlikely to be due to item-specific or item-ratio effects.

General Discussion

Overall, across the three experiments, we successfully replicate the inoculation treatment effect reported by Roozenbeek and van der Linden (2019), but with more rigorous experimental designs. We show that after playing *Bad News*, participants find fake news headlines significantly less reliable than before playing the game. The three inoculation effects ($d_{\text{Exp1}} = -1.00$, $d_{\text{Exp2}} = -0.69$, $d_{\text{Exp3}} = -0.72$) are descriptively larger than the $d = -0.52$ found in the original study by Roozenbeek and van der Linden (2019). In their broad meta-analysis of inoculation theory, Banas and Rains (2010) found a corrected average of $d = 0.43$ (95% CI = [.39, .48]) for inoculation interventions compared to control groups over 41 studies. A comparison of these results indicates that the *Bad News* inoculation intervention scores in the high range of inoculation effectiveness. In the broader context of resistance to persuasion research, these can be considered large effect sizes (Weber & Popova, 2012). Given that consequential recent elections have been decided on small margins, practically, these results are also potentially meaningful, especially when applied at population-level (Funder & Ozer, 2019).

Moreover, one potent criticism of such interventions could be that they are potentially less useful if the effects do not persist over time. The field of inoculation research lacks sufficient insights from longitudinal studies in order to accurately draw a decay function of the

²² It has to be taken into account that this was based on a comparison of one real news item that overlapped between Experiment 3 (T3) and Experiment 1 (T3), and not an index of items as is the case for the fake news analyses.

inoculation effect. Our study provides new insights into the long-term stability of inoculation interventions. Contrary to our expectations, with effects lasting up to at least three months, no evidence was found for the decay of the inoculation effect in Experiment 1. Accordingly, we theorised that regular testing in itself might have a positive “boosting” influence, and thus we leveraged insights from a second experiment (Experiment 2). When we excluded regular follow-ups, the inoculation effect was no longer significant two months after the intervention. These results demonstrate the limits of the longevity of the intervention and add new questions to the debate about the feasibility of long-term resistance against persuasion. The difference between T4 and T5 in Experiment 1 was two months, the same timeframe as between T2 and T3 in Experiment 2. Yet, whereas we find no decay in Experiment 1, we find 64% decay in Experiment 2. The possibility must be considered that in Experiment 1, little decay was observed because we used the same item sets for each test, meaning that participants may have remembered their responses from the previous test date. Although it is unlikely that they would remember the exact responses one week following the initial test (we had 21 items each on a 7-point Likert scale), the general response tendency could have been remembered. Since the fake-to-real ratio (18:3) was strongly balanced in favour of fake news, this is a valid concern (Aird et al., 2018). We, therefore, conducted a third experiment (Experiment 3), where we presented a different item set for the T3 (one week later) follow-up measure and balanced the fake-to-real item ratio (6:6). Here, we found that the inoculation effect and its decay are not influenced by item memorisation effects, thereby providing a stronger case for a broader “booster shot” or learning mechanism rather than item-specific or simple memory effects. However, these findings

cannot fully exclude the possibility that with more follow-up measures response memorisation could play a larger role.

It has been argued that the active inoculation method could be linked to longer retention of the inoculation effect, as, rather than passively reading material, participants are more cognitively involved in the intervention (McGuire, 1961; Rogers & Thistlethwaite, 1969). Researchers have found preliminary evidence that the effect could persist for six weeks up to 33 weeks (Pfau et al., 1992, 2006). In Experiment 1, we find full inoculation retention up to at least 13 weeks, thus pointing towards the potential long-term effectiveness of active inoculation interventions with regular assessment. In Experiment 2, however, we find decay after eight weeks, which may have started within the proposed six-week timeframe for inoculation intervention decay (Ivanov et al., 2018). Future research will have to look deeper into the links between memory strength and inoculation, and the potential of protecting against forgetting by implementing “booster shots.” Classical explanations for the decay of the inoculation effect include a decreasing motivation to protect the attacked attitudes and the lack of a fresh sense of threat (Ivanov, 2017; Miller et al., 2013; Pryor & Steinfatt, 1978). In the context of fake news, we deem it unlikely that the sense of “threat” has disappeared, as fake news has become a common and looming threat in the mainstream media. Moreover, Compton and Ivanov (2012) found that variable testing might boost threat levels and contribute to the effectiveness of inoculation. However, as threat has been shown to be an important contributor to motivation over time (Banas & Richards, 2017) and considering that we did not explicitly measure threat here, we cannot make any conclusions about its role in our study. In addition, linked to threat, Insko (1967, p. 316) stressed that with a decrease in motivation “the individual ceases to accumulate

belief-bolstering material..., [dropping] off over time like the ordinary forgetting curve.” A decreasing motivation is also possible in the context of information overload, as people might start to rely more on heuristics and have less energy to fight against attitudinal attacks (Laato et al., 2020).

However, we argue that an alternative theoretical model could be based on memory strength and forgetting. After an inoculation intervention participants have bolstered their psychological “immune system”, but the techniques used in the attacks have to be remembered, and are subject to interference and potentially catastrophic forgetting (Hardt et al., 2013). Indeed, we can link various key concepts of inoculation theory to a potential memory model to explain decreases in the inoculation effect. As associative networks have been linked to the long-term memory system (Collins & Loftus, 1975; Smith, 1998), inoculation effect decreases over time could be researched through the lens of neural network simulations of memory networks (Hardt et al., 2013). Over time, the memory network could suffer from *forgetting* (Frankland & Bontempi, 2005), with *interference* as the mechanism (Underwood, 1957).²³ Interference theory refers to forgetting taking place when other (similar or related) information conflicts with (or replaces) the initial memory. “Booster shots” could, therefore, be seen as relearning, protecting against interference by strengthening the memory representations (Ebbinghaus, 1885; Ivanov et al., 2018; McGuire, 1961). This leads to the question of whether the decay function can be depicted as a *forgetting curve* (Ebbinghaus, 1885; Murre & Dros, 2015)²⁴: an exponential

²³ We do not mention *trace decay* as an explanation as “there has been a long-standing consensus that [trace] decay plays no role in forgetting over the long term” (Brown & Lewandowsky, 2010, p. 51).

²⁴ The forgetting curve was proposed by Hermann Ebbinghaus (1885) in his treatise *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie* [*On Memory: A contribution to experimental psychology*]. About 130 years later, the forgetting curve was successfully replicated (see Murre & Dros, 2015).

function with the steepness of forgetting being a function of memory strength and time, suggesting that a stronger memory, which can be attained through relearning (cf. booster sessions), will be less susceptible to forgetting (i.e., less influenced by interference).

Accordingly, just like a real vaccine, it might be necessary to have several boosters before long-term immunisation can be established or come to its potential optimum (Compton & Pfau, 2005; Ivanov, 2017). We hypothesise that the tests themselves could have served as “booster shots,” being a potential reminder of the techniques and skills learned in the game as well as providing a refreshed sense of threat (Compton and Ivanov, 2012). These findings could also be explained through the lens of memory research, as researchers have shown the importance of repeated testing for memory strengthening (Karpicke & Roediger, 2008; Linton, 1975; Roediger & Karpicke, 2006a, 2006b). A future study could experiment with a shortened or passive version of *Bad News*, for example, to help refresh the cognitive skills participants have acquired during gameplay and to reactivate and strengthen associative memory networks (Pfau et al., 2005).

A different question that remains is how media literacy training can help teach people how to correctly signal real news, as well as fake news. We found that the real news indices used in our study were not reliable. The findings, reported in Supplementary Tables S3-S8 and S12-S13, Supplementary Figures S1-S4, and Figure 7, indicate that in all three experiments real news items remain rated as highly reliable (> 4 out of 7) both before and after intervention, while fake news is rated as low in reliability before and particularly after intervention (< 4 out of 7). In the original large-sample ($N = 15,000$) study on *Bad News* and its cross-cultural replication no meaningful change in real news reliability was found, but only two news items were used

(Roozenbeek & van der Linden, 2019). A recent methods paper indicates that negative effects for real news items in *Bad News* may be due to an interaction between the specific item set used and the intervention, and not generalisable to other items (Roozenbeek, Maertens, et al., 2020).

Compatible findings were seen in Experiment 3, where the same real news item was rated higher at T3 when the pretest items were different (Experiment 3) than when they were the same (Experiment 1).

As the measurement in our intervention is the change in reliability that people assign to news messages, we cannot be certain whether any changes in beliefs have occurred. We argue that the reliability rating is a proxy of the readiness to refute the fake item and with this the motivation to protect oneself against it, in line with inoculation theory. Moreover, we caution against the view that news is either “real” or “fake” and that people either “believe” or they “do not”, as most fake news is about subtle degrees of news manipulation (Ecker, Lewandowsky, Chang, et al., 2014; van der Linden & Roozenbeek, 2020). Thus, rather than informing people what is true or false, the *Bad News* intervention trains people to spot misinformation techniques so that people can calibrate their judgments accordingly (Basol et al., 2020). However, future research measuring shifts in beliefs could help further clarify this distinction.

This study does not come without limitations. The control group, in which people play *Tetris*, does not fully eliminate demand characteristics. In addition, in Experiment 1 and Experiment 2, the item sets used were not balanced in their fake-to-real ratios. Future research could introduce a control group which elicits demand effects and look into the development of a more balanced scale that is equally powerful and reliable for the correct signalling of real news as it is for fake news. Finally, while integrated into the design of the game, like McGuire, we did

not explicitly measure threat and motivation. These components have shown to be potentially important in eliciting and maintaining inoculation effects (Banas & Richards, 2017; Compton & Ivanov, 2012), and could provide useful insights into mechanisms behind the longevity of the effect. We recommend future longitudinal studies to explicitly measure these components.

In particular, to unveil the mechanisms of decay, one could consider integrating measures of threat and motivation (Compton & Ivanov, 2012) as it is possible that the treatment effect on fake news ratings is mediated by enhanced threat and motivation (Banas & Richards, 2017; Richards & Banas, 2018). Furthermore, recent best practices suggest the need to square the sample size when testing for interaction effects (Giner-Sorolla, 2018). We recommend that future studies recruit more participants per group, to enable more precise and more generalisable answers about the nature of the decay function. New insights could also be gained by replicating this experiment using more advanced longitudinal designs implementing more time points as well as varying assessment intervals.

In conclusion, with the results of this study, we gain novel insights into the effectiveness and longevity of a real-world fake news intervention based on inoculation theory. In times where the spread of (micro-targeted) misinformation is threatening public health and scientific and democratic discourse (Lewandowsky et al., 2017), inoculation based interventions could form a crucial part of the solution (Farrell et al., 2019; van der Linden & Roozenbeek, 2020). As the *Bad News* intervention is entertaining, easy to scale, adapt, and tailor, it can be put into action to protect specific groups of people who are most vulnerable to misinformation (Scheufele & Krause, 2019). In the words of the infamous defence against the dark arts teacher, Professor Severus Snape, “*Your defences must be as flexible and inventive as the arts you seek to undo.*”

References

- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14). <https://doi.org/10.1126/sciadv.aay3539>
- Aird, M. J., Ecker, U., Swire, B., Berinsky, A. J., & Lewandowsky, S. (2018). Does truth matter to voters? The effects of correcting political misinformation in an Australian sample. *Royal Society open science*, 5(12), 180593. <https://doi.org/10.1098/rsos.180593>
- Anderson, J. R. (1983). *Cognitive science series. The architecture of cognition*. Lawrence Erlbaum Associates, Inc.
- Arun, C. (2019). *On WhatsApp, rumours, lynchings, and the Indian Government*. <https://papers.ssrn.com/abstract=3336127>
- Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology. General*. <https://doi.org/10.1037/xge0000729>
- Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, 77(3), 281–311. <https://doi.org/10.1080/03637751003758193>
- Banas, J. A., & Richards, A. S. (2017). Apprehension or motivation to defend attitudes? Exploring the underlying threat mechanism in inoculation-induced resistance to persuasion. *Communication Monographs*, 84, 164–178.

<https://doi.org/10.1080/03637751.2017.1307999>

- Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about Bad News: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, 3(1), 2. <https://doi.org/10.5334/joc.91>
- Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, 10(1), 7. <https://doi.org/10.1038/s41467-018-07761-2>
- Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*, 8(1), 108–117. <https://doi.org/10.1016/j.jarmac.2018.09.005>
- Brown, G. D. A., & Lewandowsky, S. (2010). Forgetting in memory models: Arguments against trace decay and consolidation failure. In S. Della Sala (Ed.), *Forgetting* (pp. 49–75). Psychology Press. <https://doi.org/10.4324/9780203851647>
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *The Quarterly Journal of Experimental Psychology*, 10(1), 12–21. <https://doi.org/10.1080/17470215808416249>
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297–312. <https://doi.org/10.1037/h0040950>
- Chan, K., Dupuy, B., & Lajka, A. (2020, April 21). *Conspiracy theorists burn 5G towers claiming link to virus*. ABC News; ABC News. <https://abcnews.go.com/Health/wireStory/conspiracy-theorists-burn-5g-towers-claiming-link-virus-70258811>

- Chan, M.-P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546. <https://doi.org/10.1177/0956797617714579>
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. <https://doi.org/10.1037/0033-295X.82.6.407>
- Compton, J. (2009). Threat explication . *STAM Journal*, 39, 1–18.
- Compton, J. (2013). Inoculation theory. *The Sage Handbook of Persuasion: Developments in Theory and Practice*, 2, 220–237. <https://doi.org/10.4135/9781452218410>
- Compton, J. (2019). Prophylactic versus therapeutic inoculation treatments for resistance to influence. *Communication Theory*. <https://doi.org/10.1093/ct/qtz004>
- Compton, J., & Ivanov, B. (2012). Untangling threat during inoculation-conferred resistance to influence. *Communication Reports*, 25(1), 1–18. <https://doi.org/10.1080/08934215.2012.661018>
- 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. *Annals of the International Communication Association*, 29(1), 97–146. <https://doi.org/10.1080/23808985.2005.11679045>
- Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PloS One*, 12(5), e0175799. <https://doi.org/10.1371/journal.pone.0175799>
- Coppock, A. (2019). Generalizing from survey experiments conducted on Mechanical Turk: A

replication approach. *Political Science Research and Methods*, 7(3), 613–628.

<https://doi.org/10.1017/psrm.2018.10>

De keersmaecker, J., & Roets, A. (2017). ‘Fake news’: Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence*, 65, 107–110. <https://doi.org/10.1016/j.intell.2017.10.005>

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.

Ebbinghaus, H. (1885). *Über das gedächtnis: Untersuchungen zur experimentellen psychologie*. Duncker & Humblot.

Ecker, U. K. H., & Ang, L. C. (2019). Political attitudes and the processing of misinformation corrections. *Political Psychology*. <https://doi.org/10.1111/pops.12494>

Ecker, U. K. H., Lewandowsky, S., Chang, E. P., & Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology. Applied*, 20(4), 323–335. <https://doi.org/10.1037/xap0000028>

Ecker, U. K. H., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, 42(2), 292–304. <https://doi.org/10.3758/s13421-013-0358-x>

Farrell, J., McConnell, K., & Brulle, R. (2019). Evidence-based strategies to combat scientific misinformation. *Nature Climate Change*. <https://doi.org/10.1038/s41558-018-0368-6>

Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38, 127–150. <https://doi.org/10.1111/pops.12394>

- Forgas, J. P. (2001). *Feeling and thinking: The role of affect in social cognition*. Cambridge University Press.
- Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews. Neuroscience*, 6(2), 119–130. <https://doi.org/10.1038/nrn1607>
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- Frenkel, S., Alba, D., & Zhong, R. (2020). Surge of virus misinformation stumps Facebook and Twitter. *The New York Times*. https://www.bridgeportedu.net/cms/lib/CT02210097/Centricity/Domain/3754/Costello_Journalism_11_3_23.3_31.pdf
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Giner-Sorolla, R. (2018). *Powering your interaction*. <https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/>
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, 363(6425), 374–378. <https://doi.org/10.1126/science.aau2706>
- Guess, A., & Coppock, A. (2018). Does counter-attitudinal information cause backlash? Results from three large survey experiments. *British Journal of Political Science*, 1–19. <https://doi.org/10.1017/S0007123418000327>

- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), eaau4586.
<https://doi.org/10.1126/sciadv.aau4586>
- Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends in Cognitive Sciences*, 17(3), 111–120.
<https://doi.org/10.1016/j.tics.2013.01.001>
- Hornsey, M. J., & Fielding, K. S. (2017). Attitude roots and Jiu Jitsu persuasion: Understanding and overcoming the motivated rejection of science. *The American Psychologist*, 72(5), 459–473. <https://doi.org/10.1037/a0040437>
- Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: a potentially confusing task. *Psychological Bulletin*, 82(4), 511. <https://doi.org/10.1037/h0076767>
- Ivanov, B. (2012). Designing inoculation messages for health communication campaigns. In H. Cho (Ed.), *Health communication message design: Theory and practice* (pp. 73–93). Sage.
- Ivanov, B. (2017). Inoculation theory applied in health and risk messaging. In *The Oxford Encyclopedia of Health and Risk Message Design and Processing*. Oxford University Press.
<https://doi.org/10.1093/acrefore/9780190228613.013.254>
- Ivanov, B., Parker, K. A., & Dillingham, L. L. (2018). Testing the limits of inoculation-generated resistance. *Western Journal of Speech Communication: WJSC*, 82(5), 648–665.
<https://doi.org/10.1080/10570314.2018.1454600>
- Jolley, D., & Paterson, J. L. (2020). Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence. *British Journal of Social Psychology*, 59(3),

628–640. <https://doi.org/10.1111/bjso.12394>

- Jost, J. T., van der Linden, S., Panagopoulos, C., & Hardin, C. D. (2018). Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current Opinion in Psychology*, 23, 77–83. <https://doi.org/10.1016/j.copsyc.2018.01.003>
- Karpicke, J. D., & Roediger, H. L. III. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Laato, S., Islam, A. N., Islam, M. N., & Whelan, E. (2020). What drives unverified information sharing and cyberchondria during the COVID-19 pandemic? *European Journal of Information Systems*, 1-18. <https://doi.org/10.1080/0960085X.2020.1770632>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the ‘post-truth’ era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Linton, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. Freeman.

- Maertens, R., Anseel, F., van der Linden, S. (2020). Combatting climate change misinformation: Evidence for longevity of inoculation and consensus messaging effects. *Journal of Environmental Psychology*, *70*, 101455. <https://doi.org/10.1016/j.jenvp.2020.101455>
- McGuire, W. J. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *Journal of Abnormal and Social Psychology*, *63*(2), 326. <https://doi.org/10.1037/h0048344>
- McGuire, W. J. (1964). Inducing resistance to persuasion: Some contemporary approaches. *Advances in Experimental Social Psychology*, *1*, 191–229. [https://doi.org/10.1016/S0065-2601\(08\)60052-0](https://doi.org/10.1016/S0065-2601(08)60052-0)
- McGuire, W. J. (1970). Vaccine for brainwash. *Psychology Today*, *3*(9), 36–39.
- McGuire, W. J. (1973). Persuasion, resistance, and attitude change. *Handbook of Communication*, *1*.
- McGuire, W. J., & Papageorgis, D. (1962). Effectiveness of forewarning in developing resistance to persuasion. *Public Opinion Quarterly*, *26*(1), 24–34. <https://doi.org/10.1086/267068>
- Miller, C. H., Ivanov, B., Sims, J., Compton, J., Harrison, K. J., Parker, K. A., Parker, J. L., & Averbeck, J. M. (2013). Boosting the potency of resistance: Combining the motivational forces of inoculation and psychological reactance. *Human Communication Research*, *39*(1), 127–155. <https://doi.org/10.1111/j.1468-2958.2012.01438.x>
- Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PloS One*, *10*(7), e0120644. <https://doi.org/10.1371/journal.pone.0120644>
- Nader, K., & Hardt, O. (2009). A single standard for memory: the case for reconsolidation.

Nature Reviews Neuroscience, 10(3), 224–234. <https://doi.org/10.1038/nrn2590>

Niederdeppe, J., Heley, K., & Barry, C. L. (2015). Inoculation and narrative strategies in competitive framing of three health policy issues. *The Journal of Communication*, 65(5), 838–862. <https://doi.org/10.1111/jcom.12162>

Pennycook, G., Cannon, T., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12). <https://doi.org/10.1037/xge0000465>

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*. <https://doi.org/10.1111/jopy.12476>

Petersen, A. M., Vincent, E. M., & Westerling, A. L. (2019). Discrepancy in scientific authority and media visibility of climate change scientists and contrarians. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-09959-4>

Petty, R. E., Priester, J. R., & Wegener, D. T. (1994). Cognitive processes in attitude change. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (pp. 69–142). Lawrence Erlbaum Associates, Inc.

Pfau, M., Compton, J., Parker, K. A., An, C., Wittenberg, E. M., Ferguson, M., Horton, H., & Malyshev, Y. (2006). The conundrum of the timing of counterarguing effects in resistance: Strategies to boost the persistence of counterarguing output. *Communication Quarterly*,

54(2), 143–156. <https://doi.org/10.1080/01463370600650845>

Pfau, M., Compton, J., Parker, K. A., Wittenberg, E. M., An, C., Ferguson, M., Horton, H., &

Malyshev, Y. (2004). The traditional explanation for resistance versus attitude accessibility:

Do they trigger distinct or overlapping processes of resistance? *Human Communication*

Research, 30(3), 329–360. <https://doi.org/10.1111/j.1468-2958.2004.tb00735.x>

Pfau, M., Ivanov, B., Houston, B., Haigh, M., Sims, J., Gilchrist, E., Russell, J., Wigley, S.,

Eckstein, J., & Richert, N. (2005). Inoculation and mental processing: The instrumental role

of associative networks in the process of resistance to counterattitudinal influence.

Communication Monographs, 72(4), 414–441. <https://doi.org/10.1080/03637750500322578>

Pfau, M., Szabo, A., Anderson, J., Morrill, J., Zubric, J., & Wan, H. H. (2001). The role and

impact of affect in the process of resistance to persuasion. *Human Communication*

Research, 27(2), 216–252. <https://doi.org/10.1111/j.1468-2958.2001.tb00781.x>

Pfau, M., Tusing, K. J., Koerner, A. F., Lee, W., Godbold, L. C., Penaloza, L. J., Yang, V. S.-H.,

& Hong, Y.-H. (1997). Enriching the inoculation construct: The role of critical components

in the process of resistance. *Human Communication Research*, 24(2), 187–215.

<https://doi.org/10.1111/j.1468-2958.1997.tb00413.x>

Pfau, M., & Van Bockern, S. (1994). The persistence of inoculation in conferring resistance to

smoking initiation among adolescents. *Human Communication Research*, 20(3), 413–430.

<https://doi.org/10.1111/j.1468-2958.1994.tb00329.x>

Pfau, M., Van Bockern, S., Kang, J.G. (1992). Use of inoculation to promote resistance to

smoking initiation among adolescents. *Communication Monographs*, 59(3), 213–230.

<https://doi.org/10.1080/03637759209376266>

- Pryor, B., & Steinfatt, T. M. (1978). The effects of initial belief level on inoculation theory and its proposed mechanisms. *Human Communication Research, 4*(3), 217–230.
<https://doi.org/10.1111/j.1468-2958.1978.tb00611.x>
- Richards, A. S., & Banas, J. A. (2018). The opposing mediational effects of apprehensive threat and motivational threat when inoculating against reactance to health promotion. *Southern Communication Journal, 83*(4), 245–255. <https://doi.org/10.1080/1041794X.2018.1498909>
- Roediger, H. L. III, & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.
<https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Roediger, H. L. III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210.
<https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Rogers, R. W., & Thistlethwaite, D. L. (1969). An analysis of active and passive defenses in inducing resistance to persuasion. *Journal of Personality and Social Psychology, 11*(4), 301–308. <https://doi.org/10.1037/h0027354>
- Roozenbeek, J., Maertens, R., McClanahan, W., & van der Linden, S. (2020). Disentangling item and testing effects in inoculation research on online misinformation: Solomon revisited. *Educational and Psychological Measurement, 1*–23.
<https://doi.org/10.1177/0013164420940378>
- Roozenbeek, J., & van der Linden, S. (2018). The fake news game: actively inoculating against the risk of misinformation. *Journal of Risk Research, 1*–11.
<https://doi.org/10.1080/13669877.2018.1443491>

- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1), 65.
<https://doi.org/10.1057/s41599-019-0279-9>
- Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on the psychological theory of ‘inoculation’ can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-008>
- Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America*.
<https://doi.org/10.1073/pnas.1805871115>
- Smith, E. R. (1998). Mental representation and memory. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 391–445). McGraw-Hill.
- Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, 46(2), 137–150. <https://doi.org/10.1037/h0062958>
- Song, M.-K., & Ward, S. E. (2015). Assessment effects in educational and psychosocial intervention trials: An important but often-overlooked problem. *Research in Nursing & Health*, 38(3), 241–247. <https://doi.org/10.1002/nur.21651>
- Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: comprehending the Trump phenomenon. *Royal Society Open Science*, 4(3), 160802. <https://doi.org/10.1098/rsos.160802>
- Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology. Learning, Memory, and*

Cognition, 43(12), 1948–1961. <https://doi.org/10.1037/xlm0000422>

Tannenbaum, P. H., Macauley, J. R., & Norris, E. L. (1966). Principle of congruity and reduction of persuasion. *Journal of Personality and Social Psychology*, 3(2), 233–238.

<https://doi.org/10.1037/h0022893>

Thorndike, E. L. (1913). *The psychology of learning*. Teachers College, Columbia University.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, 64(1), 49–60.

<https://doi.org/10.1037/h0044616>

van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2), 1600008.

<https://doi.org/10.1002/gch2.201600008>

van der Linden, S., Panagopoulos, C., & Roozenbeek, J. (2020). You are fake news: political bias in perceptions of fake news. *Media Culture & Society*, 42(3), 460–470.

<https://doi.org/10.1177/0163443720906992>

van der Linden, S., & Roozenbeek, J. (2020). Psychological inoculation against fake news. In R. Greifeneder, M. Jaffé, E. J. Newman, & N. Schwarz (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation*. Psychology Press.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>

Weber, R., & Popova, L. (2012). Testing equivalence in communication research: Theory and application. *Communication Methods and Measures*, 6(3), 190–213.

<https://doi.org/10.1080/19312458.2012.703834>

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual

adherence. *Political Behavior*, 41(1), 135–163. <https://doi.org/10.1007/s11109-018-9443-y>

Zerback, T., Töpfl, F., & Knöpfle, M. (2020). The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them. *New Media & Society*, 1461444820908530. <https://doi.org/10.1177/1461444820908530>