

9

PSYCHOLOGICAL INOCULATION AGAINST FAKE NEWS

Sander van der Linden and Jon Roozenbeek

Acknowledgments

We would like to thank the Cambridge Social Decision-Making Lab, Ruurd Oosterwoud and DROG, Design Studio Gusmanson, the Economic and Social Research Council (ESRC), and the University of Cambridge for their support and for funding this research. We also thank the editors and reviewers for their helpful feedback on an earlier version of this chapter.

What is fake news?

Fake news appears everywhere. After gaining steam during the 2016 US presidential election, the phrase has become ubiquitous in popular media. US President Donald Trump uses it to lambast journalists and media outlets for what he sees as biased coverage; researchers build algorithms to detect false and misleading stories and document their spread; Facebook is regularly forced to explain how it intends to prevent fake news from going viral on its platform; and governments are taking steps to crack down on fake news stories circulating on the internet (Bremner, 2018; Shao et al., 2018; Wakabayashi, 2017).

Accordingly, “fake news” has rapidly become a catchall phrase that lacks an accepted working definition (Tandoc Jr, Lim, & Ling, 2018). Although some have attempted to explicate the “science of fake news” (e.g., Lazer et al., 2018), if parties with such diverse interests as the BBC and President Donald Trump are using the term and take it to mean entirely different things in different contexts, it becomes difficult to know what we talk about when we talk about fake news.

It is also quite clear that the term does not do a very good job at describing the full breadth of the problem. Perhaps a sensible definition of “fake news” could be “fabricated information that mimics news media content in form, but not in organizational process or intent” (Lazer et al., 2018, p. 1). Snopes is one of

the websites that keeps track of stories like this. Examples are not hard to find: headlines like “Australia to Forcibly Vaccinate Citizens via Chemtrails”, “Melania Trump Bans White House Staff from Taking Flu Shot” and “Muslim Doctor Refuses to Treat Christian Girl on Board a Flight” are just one Google search away (Adl-Tabatabai, 2016; Baxter, 2018; Patriot United, 2018).

But news stories do not have to be completely false to be misleading. It is easy to quote people out of context to make it look like they are saying something that they never said, or to add misleading context to a video or image. For example, see Figure 9.1. This was a commentary posted by the Facebook page “News World” on March 20, 2018. The video purports to show Muslim immigrants in France attacking a Catholic church during mass. It was viewed about 1.2 million times within a day after it was posted. Politicians, including Front National leader Marine le Pen, expressed their outrage on Twitter, writing that the church had been “desecrated” (Le Pen, 2018).

However, fact checkers were quick to point out a number of problems with these claims. There was no evidence of the protesters’ religion or the time of their arrival to France. Furthermore, the church was not “attacked”, at least not according to church members themselves. Instead, the people in the video were protesting a proposed bill that would make obtaining asylum in the country more difficult (Snopes, 2018). The demonstration remained nonviolent.



FIGURE 9.1 News World Facebook post (20 March 2018)

The problem with the Facebook post in Figure 9.1 is not that the information is completely false or that the events shown in the video never happened. Rather, it is the misleading context provided in the post that does most of the actual damage. In short, from a psychological perspective, intent matters, and *misleading* and *fake* are not entirely the same thing. It is therefore worthwhile to think of the different types of “fake news” along a spectrum. On one end of the spectrum we have *misinformation*, which is simply information that is false or incorrect (and can include human error). Next, we have *disinformation*, which involves misinformation coupled with a deliberate *intent* to deceive an audience. Compared to simple human error, the involvement of intent has important psychological connotations (van der Linden, 2017; van der Linden, Roozenbeek, Oosterwoud, Compton, & Lewandowsky, 2018). Propaganda is then defined as disinformation paired with an explicit or implicit political agenda (van der Linden, 2017). To keep things simple, we will be using the term *disinformation* instead of *fake news* in the current chapter, to ensure that we are not just talking about fake stories but about media manipulation more generally.

The disinformation problem

Although clearly not a new phenomenon, disinformation has become a much more serious issue with the advent of the internet. The possibility of instant and anonymous communication makes the internet an ideal vehicle for deception. Perhaps it is no surprise then that disinformation is commonly used by a variety of actors, including some governments, to influence public opinion. Social media platforms are a particularly fertile breeding ground. To give an example, around 47 million Twitter accounts (approximately 15%) are bots (Varol, Ferrara, Davis, Menczer, & Flammini, 2017) and many of these bots are used to purposefully spread political disinformation, especially during election campaigns (Ferrara, 2017).

Recent examples of influential disinformation include conspiracy theories about COVID-19, the MacronLeaks during the French presidential elections in 2017, the Pizzagate controversy during the 2016 US elections, the various “alternative” explanations surrounding the downing of Malaysia Airlines flight MH17 in July 2014, and rumors circulating in Sweden about the country’s cooperation with NATO (Kragh & Åsberg, 2017). This onslaught of online disinformation is taking its toll. For example, consider that a majority of Americans admit that fake news has left them feeling confused about basic facts (Barthel, Mitchell, & Holcomb, 2016), and 83% of Europeans think that fake news is a threat to democracy (Eurobarometer, 2018). Moreover, a recent British study by YouGov indicated that only 4% of participants were able to discern fake news from real news (Channel 4, 2017). In some ways, this is not surprising: people are bombarded with information as they scroll through their news feeds.

As any functioning democracy relies on a well-informed populace, the rise of disinformation is proving to be a real threat to the democratic process

(Lewandowsky, Ecker, & Cook, 2017; van der Linden et al., 2017). The political landscape in many countries is also getting more and more polarized. This level of polarization is reinforced by the emergence of ideologically homogeneous filter bubbles, where people are exposed to stories that are congenial to their ideological worldviews (Del Vicario et al., 2016). If vast amounts of people are in the dark about what to believe and whom to trust, this can have serious consequences for evidence-based decision making on a whole range of issues, from climate change and vaccinations to international relations (Lewandowsky et al., 2017; Poland & Spier, 2010; van der Linden, Leiserowitz, Rosenthal, & Maibach, 2017). In some cases, the viral spread of fake stories has led to injury and even death (BBC, 2018a).

We also know that effective disinformation campaigns are not easily reversed. Studies on the continued influence of misinformation consistently show that acquired beliefs are very difficult to correct, even when people acknowledge that their views are based on erroneous information (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). Thus, while disinformation is probably as old as the spread of rumor itself, it has become easier than ever to create and disseminate disinformation and more and more people are exposed to false content simply by virtue of their daily news consumption. Accordingly, governments, public officials, and media companies all have proposed a range of potential solutions to combat the growing disinformation problem.

Potential solutions

The solutions that are being proposed to solve the problem of disinformation can be divided into four broad categories: (1) algorithmic, (2) corrective, (3) legislative, and (4) psychological. In terms of the first category, Google and Facebook are discovering how to tweak their algorithms to disincentivize fake or unreliable news sites and prevent disinformation from appearing on people's newsfeeds in the same way as "reliable" news sites (Calfas, 2017; Elgin & Wang, 2018). However, algorithms are clearly imperfect at detecting misleading content and past attempts, such as by Facebook, have often backfired (Wakefield, 2017). The second category refers to the post-hoc correction of false stories through fact-checking tools. Fact-checking initiatives abound, and some (such as PolitiFact and Snopes) have even become household names. However, although fact-checking initiatives are laudable, evidence for their efficacy remains mixed (Nyhan, Porter, Reifler, & Wood, 2019). Moreover, it is impossible to debunk every fake or misleading story, as producing fake news requires less resources than debunking it, and the potential audience for fact-checking reports remains limited (Kurtzleben, 2016). Recent research also indicates that false stories spread more easily on social media than any other kind of news, even when controlling for stories spread by bots (Vosoughi, Roy, & Aral, 2018).

A more radical approach involves the introduction of new regulation and legislation to combat disinformation. A prominent example is France's "Fake

News Law”, which during election time would place tougher restrictions on media outlets as to what content they are allowed to put online (Bremner, 2018). Similar initiatives have been proposed in the United Kingdom (e.g., the “fake news unit” (BBC News, 2018b). Yet, granting any organization, governmental or not, the power to decide what information is “real” and what is “fake” can easily backfire. For example, a European Union–funded working group named “EUvsDisinfo” was heavily criticized for flagging a number of Dutch non-mainstream news sites and one local newspaper as “spreading Kremlin disinformation” (Pieters, 2018). Dutch parliamentarians expressed their concern that EUvsDisinfo was infringing on freedom of speech, and voted to lobby to scrap the working group altogether.

Insights from psychology: inoculation against disinformation

Accordingly, more attention is now being directed toward the role of psychology, education, and the behavioral sciences in combating fake news to help empower people at the individual level (European Commission, 2018). Of course, the basic idea that fostering critical and well-informed news consumers will make disinformation less effective in the long term is sensible. Yet, the problem with most traditional media literacy approaches lies in the fact that it is neither feasible nor possible to correct every false story. Another key problem is that developing better *debunking* techniques is unlikely to be sufficient by itself to stem the onslaught of fake news. In fact, even when corrections are issued, the damage has often already been done: once people have acquired a false belief, they are unlikely to update their views. Indeed, research on the “continued influence effect” suggests that corrections are often ineffective as people continue to rely on false (and debunked) information, especially in the face of politically motivated cognition (Lewandowsky et al., 2012). This raises the following question: is it possible to *prevent* false narratives from taking hold in the first place? To investigate this question, we turn to what Eagly and Chaiken (1993, p. 561) referred to as the “grandparent theory of resistance to persuasion”: *inoculation theory*. During World War II, the United States War Department had an experimental research branch in a unit called the “Department of Information and Education”. This division, led by the social psychologist, Carl Hovland, was tasked specifically with conducting research on political persuasion and propaganda campaigns (Hovland, Lumsdaine, & Sheffield, 1949). Their studies formed the basis of what later became known as one of the most foundational groups in social psychology: the Yale Attitude and Persuasion Program (Huddy, Sears, & Levy, 2013). Their mission was to uncover “the basic laws of persuasion” using controlled experimental methods. Yet, in the 1960s, a new concern arose; the potential brainwashing of captured American soldiers in the Far East. Accordingly, psychologist William McGuire shifted his focus toward a different question: how can we help people resist persuasion attempts? This ultimately led him to develop “inoculation theory”, which he described as a “vaccine for brainwash” (see Figure 9.2).

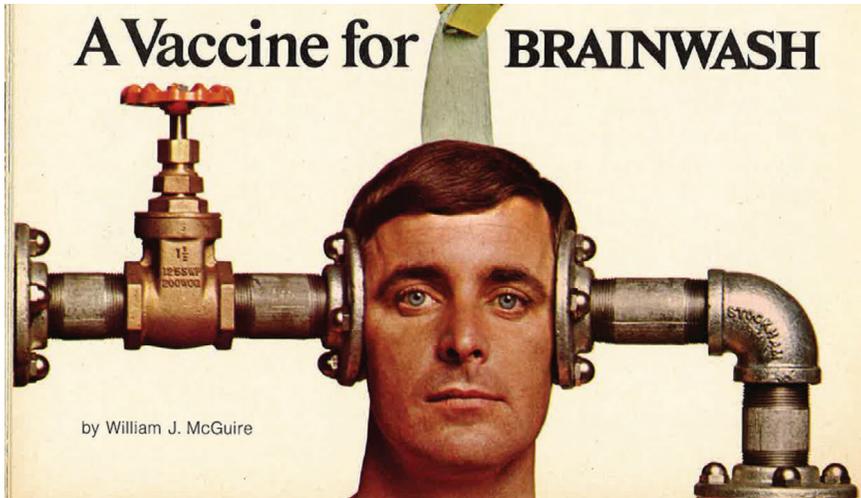


FIGURE 9.2 A vaccine for brainwash (McGuire, 1970)

Inoculation theory is based on an analogy from immunology (McGuire, 1970; McGuire & Papageorgis, 1961, 1962; Papageorgis & McGuire, 1961). Vaccines are weakened versions of pathogens (e.g., a virus) that, upon introduction to the body, trigger the production of antibodies. These antibodies become active once the real version of the pathogen enters the body thus conferring protection (immunity) against future infection. Inoculation theory postulates that the same can occur with information: by preemptively presenting someone with a weakened version of a misleading piece of information, a thought process is triggered that is analogous to the cultivation of “mental antibodies”, rendering the person immune to (undesirable) persuasion attempts (Compton, 2013; McGuire & Papageorgis, 1961). Over the years, a large body of evidence has been amassed showing that public attitudes can be inoculated across domains, including health (Compton, Jackson, & Dimmock, 2016) and politics (Pfau, Park, Holbert, & Cho, 2001). Meta-analyses also confirm that inoculation messages are effective at conferring resistance to persuasion (Banas & Rains, 2010). The inoculation process consists of two main components, namely: (1) a warning to elicit and activate threat in message recipients (the affective basis) and (2) refutational preemption (the cognitive basis). Forewarning people that they are about to be exposed to counter-attitudinal content is thought to elicit *threat* to motivate the protection of existing beliefs. In turn, two-sided refutational messages both inform and teach in the sense that they model the counterarguing process for people and provide specific content that can be used to resist persuasion attempts (McGuire, 1970; Compton, 2013).

Interestingly, a number of important open questions remain about the theory, particularly with regard to its application to fake news and disinformation. For example, inoculation theory has traditionally been applied to so-called cultural

truisms or widely held beliefs (e.g., the belief that brushing your teeth twice a day is good for your health, see McGuire, 1970). As such, a major open question has been how the inoculation process operates, theoretically, when people have divergent prior attitudes about an issue (as is often the case with fake news). When audiences do not already possess the desired attitude, the inoculation process is not prophylactic in the traditional sense, but rather takes on a “therapeutic” role – analogous to the emerging use of therapeutic vaccines (Compton, 2019). Second, from an intervention science perspective, it remains unclear how the inoculation process can be scaled at population level, as clearly, it is neither feasible nor possible to preemptively refute every fake news story specifically. Lastly, inoculation treatments have traditionally relied on a “passive” process where recipients read a persuasive message that forewarns and refutes potential counterarguments. However, McGuire theorized early on that a more active inoculation process could be more powerful by letting people generate their own pro- and counterarguments.¹ Accordingly, in three studies, we sought to provide initial answers to these important yet unresolved questions.

In the lab: inoculating the public against misinformation about climate change (study 1)

To answer the first question, we wanted to see whether exposing the public to a weakened version of a falsehood, and preemptively debunking that falsehood with scientific facts (a vaccine), could offer resistance against fake news about a highly polarized and contested issue: global warming. Our lab conducted two large online studies to test these hypotheses. In the first study, we used a national probability sample ($N = 1,000$) of the US population to evaluate what popular “falsehoods” about climate change people were most familiar with. The most commonly recognized source of misinformation was a real online petition (The Oregon Petition), which claims to have gathered over 31,000 signatures from scientists who disagree that human-caused global warming is real (fueling the most popular fake news story about climate change on social media in 2016, see Readfearn, 2016).

In the second study, we relied on a large and diverse sample ($N = 2,167$) of US adults from Amazon Mechanical Turk. We used a real screenshot of the petition website as the experimental “misinformation” treatment and the simple fact that over 97% of climate scientists have concluded that human-induced global warming is happening as the “factual” statement (Cook et al., 2016). We then randomly assigned participants to one of six conditions and asked about their judgments of the scientific consensus both before and after (see Figure 9.4). The six conditions were (1) simple facts (the 97% consensus), (2) real misinformation (the petition), (3) false balance (the consensus versus the petition), (4) partial inoculation (forewarning only), (5) full inoculation (forewarning plus preemptive refutation), and (6) a “pure” control group in which participants solved a neutral word puzzle.

In Figure 9.3, from left to right (panel a), in the “facts” condition, participants only read about the evidence and shifted their views (pre-post) on climate change in line with the scientific consensus (97%) by about 20 percentage points ($d = 1.23$). In the “misinformation” condition, subjects were only shown the misinformation and shifted their views down by about 10% ($d = 0.48$). In the “false balance” condition, participants viewed both treatments side by side; here, the presence of “sticky” misinformation completely neutralized the facts ($d = 0.04$). In the last two “inoculation” conditions, participants were informed of the facts with either a short warning (W) that politically motivated actors may try to influence public opinion using misleading tactics or additional arguments were used to *prebunk* the misinformation in detail (e.g., that signatories of this petition include Charles Darwin and the Spice Girls). The inoculation treatments (labeled “W” for warning only and “F” for full in Figure 9.3) proved effective, preserving about one-third ($d = 0.33$) and two-thirds ($d = 0.75$) of the factual message, respectively. Crucially, the same pattern replicated across the political spectrum as well as across participants’ prior attitudes about climate change (panel b).

In short, this study provided an important and compelling answer to our first question: even outside of the context of politically neutral “truisms”, inoculation can be effective. In fact, rather than backfire, inoculation appears to have important retroactive or “therapeutic” benefits, even among people who reject climate science for political reasons. An independent study conceptually replicated these findings with generally stronger results (see Cook, Lewandowsky, & Ecker, 2017).

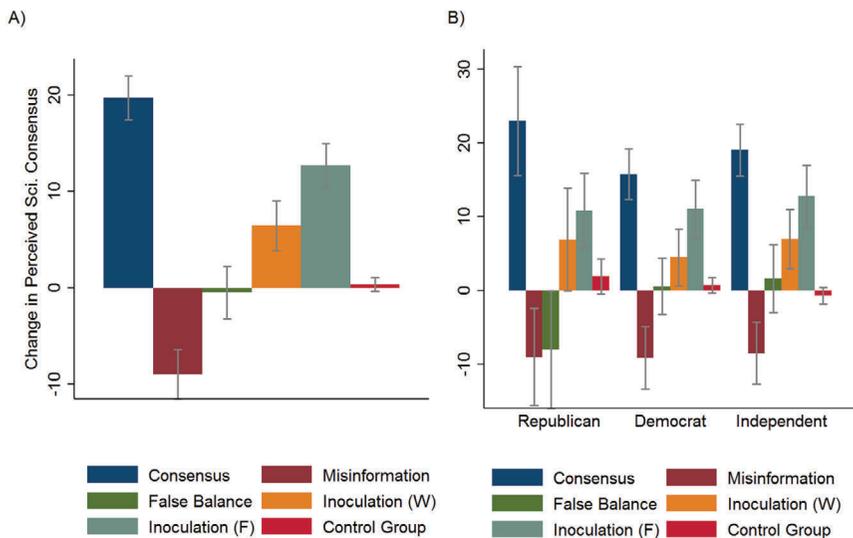


FIGURE 9.3 Inoculating against climate change misinformation (van der Linden et al., 2017)

In the field: actively inoculating against fake news (study 2)

Our initial research still left us with two important questions: (1) instead of passively reading articles is it possible to inoculate people in a more “active” and “experiential” manner to aid the learning process; and (2) can the “vaccine” be generalized and extended to other domains? For example, although it is possible to tailor inoculation messages to a particular issue by creating weakened doses of specific misinformation, what about the prospect of conferring general resistance against disinformation? To answer these questions, we established a partnership with the Dutch Media Collective “Bad News”, an organization that creates novel educational materials to combat disinformation. Jointly, we extended our initial work by translating the laboratory findings into an interactive educational experience: *The Fake News Game*. We theorized that taking on the role of someone who is actively trying to deceive you will be an effective way of conferring more general resistance to misinformation. Accordingly, the game lets players walk a mile in the shoes of a fake news producer. Initially, we produced a paper-based version of the game (see Figure 9.4) where students pick a specific character, such as a conspiracy theorist or a clickbait monger, and assemble structured news articles in a way that is consistent with their role.

We pilot tested the game with 95 senior students (aged 16–18) in a Dutch high school in the context of fake news about the Syrian refugee crisis (Roozenbeek & van der Linden, 2018). Classes were randomized into a treatment and a control group. After playing the game for about 30 minutes, students in the treatment



FIGURE 9.4 The Fake News card game (Roozenbeek & van der Linden, 2018)

group significantly downgraded the reliability of previously unseen fabricated news articles about the refugee crises compared to a control group who simply watched a video as part of the standard lesson plan. To evaluate whether threat was elicited in the process, a sentiment analysis on the open-ended responses revealed significantly higher negative affect levels in the treatment group. Although these results were encouraging, the power of the study was relatively low and students did not necessarily change their attitudes about immigration. This is consistent with other recent work which finds that although media interventions can reduce misperceptions, this doesn't necessarily reflect changes in political beliefs (Nyhan et al., 2019). To be fair, however, the intervention is not aimed at changing political beliefs, but simply to help people spot disinformation techniques. Crucially, in our field study, the “weakened” fake news article the students were “trained on” was different from the article they were tested on – providing preliminary evidence that the boundary conditions of the inoculation metaphor can be extended.

Into the wild: the bad news game (study 3)

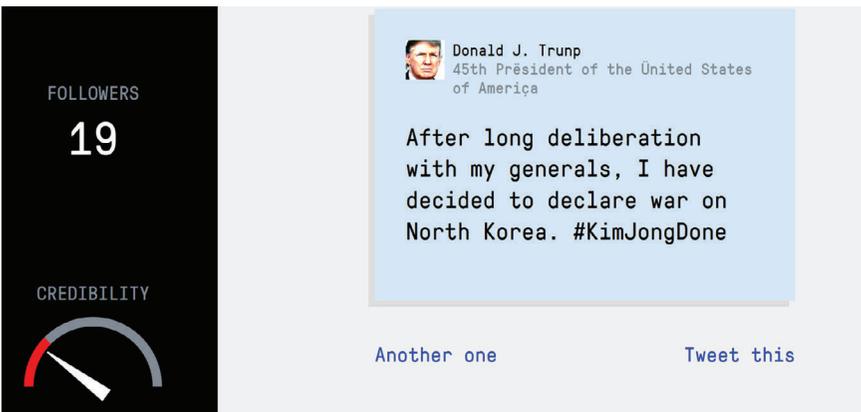
Based on these results, we designed a multiple award-winning online version of the Fake News game (FastCompany, 2018). We called it “*Bad News*” (Roozenbeek & van der Linden, 2019). The online game simulates a social media engine (Twitter) so that players have to attract followers by spreading fake news online. The interface of the game is user-friendly (see Figure 9.5a); players are shown a short text or image (such as a meme or headline) and can react to them in a variety of ways. In the game, scores are measured via a “followers” and “credibility” meter (panel b). The aim of the game is to gather as many followers as possible without losing credibility. Choosing an option that is in line with what a “real” producer of disinformation would choose gets players more followers and credibility. If, however, they lie too blatantly to their followers or act too much in line with journalistic best practices, the game either takes followers away or lowers their credibility score.

Disinformation strategies

Following the inoculation metaphor, the game exposes players to severely weakened doses of disinformation by *actively* letting them generate their own content. However, in contrast to issue-based inoculations, we hypothesized that it may be possible to “vaccinate” people against the very tactics that underlie the production of most fake news (analogous to a *broad-spectrum* vaccine). As it is impossible to cover all aspects of disinformation in detail in a 15-minute game, we chose to cover only the most common strategies. Over the course of six theory-driven “badges”, players learn about impersonating people online, using emotional language, group polarization, floating conspiracy theories, building echo chambers, discrediting opponents, trolling, and false amplification. These strategies are partially derived from the report “Digital Hydra” by NATO’s East Strategic



(a)



(b)

FIGURE 9.5 Screen captions of the Bad News game (www.getbadnews.com)

Command (East StratCom), which details the various forms that disinformation can take as well as academic work on deception strategies (Bertolin, Agarwal, Bandeli, Biteniece, & Sedova, 2017). The following sections offer a quick summary of the scenarios and theoretical background of each badge specifically.

Impersonation

It is no longer difficult to start a website and publish content that looks entirely legitimate. Since there is almost no entry barrier in terms of costs, pretty much anyone can become a content producer (Goga, Venkatadri, & Gummadi, 2015; Orita & Hada, 2009; Reznik, 2013). The purpose of this badge is to show how

easy this process is and how a professional look or name does not necessarily imply legitimacy. “Impersonation” has two main components: (1) impersonating a real person or organization by mimicking their appearance, for example, by using a slightly different username, for example when a hoaxer impersonated billionaire investor Warren Buffett on Twitter in late 2018 (BBC News, 2018c), and (2) posing as a legitimate news website or blog without the usual journalistic norms and credentials.

In the game, players first post a tweet about something that frustrates them, which can be anything from a failing government to the Flat Earth Society. This gets them their first followers, and the game explains how the follower counter and credibility meter functions. Players then impersonate the official account of either Donald Trump (who declares war on North Korea), NASA (which announces that a massive meteor is about to hit the earth), or Nickelodeon (declaring the impending cancellation of *SpongeBob SquarePants*). Players are subsequently shown tweets by Twitter users who fell for the impersonation hoax. The game then prompts them to go professional and start their own news site. They pick a website name, title, and slogan.

Emotional content

Emotional content is content that is not necessarily “fake” or “real” but deliberately plays into people’s basic emotions, such as fear, anger, or empathy (Aday, 2010; Bakir & McStay, 2017; Gross & D’Ambrosio, 2004; Konijn, 2013). The aim of this badge is to show how players can manipulate basic emotions in order to rile up their followers and make their content shared more readily.

This is the first badge where players produce content for their fictional news site. They are prompted to browse news headlines for a topic that they can publish about on their site, with a choice between climate change and genetically modified organisms (GMOs). Players are then asked for their opinion on their topic of choice. The game prompts them to say that their topic will either bring about the apocalypse (in the case of GMOs) or is a complete hoax (in the case of climate change), as this is the easiest way to gain followers. The game asks them to choose an approach to the topic at hand: attack scientists, post an emotional story, or talk about the science. The latter option returns a negative response, as players are encouraged to use reactionary content to rile up their followers. They can then either create a meme (a humorous piece of media, usually an image or GIF, that spreads from person to person on the internet) or write an article that reflects their choice. Each choice comes with numerous options, of which one is always bad (because it misses the point). Some of their followers will react to their post on Twitter in an emotional, angry way, which is exactly the player’s goal.

Polarization

Polarization involves deliberate attempts to expand the gap between the political left and the political right and drive people away from the political center (Bessi

et al., 2016; Groenendyk, 2018; Melki & Pickering, 2014; Prior, 2013; Twenge, Honeycutt, Prislin, & Sherman, 2016). In order to gain followers, young news sites often use polarization as a way to stake out a niche in the online media landscape. This badge also covers the concept of “false amplification” or the idea that it is not necessary to create a completely fake story in order to get a point across. Instead, one can also amplify existing grievances and make them look more popular than they really are (Bertolin et al., 2017).

At the start of this badge, players are asked if they want to publish something fake or something real. Choosing “fake” tells them that they do not always have to invent fake news in order to make headlines, but that they can also find a real story and blow it out of proportion. They can then drive left and right further apart by choosing between three local news stories as reported by random citizens on Twitter: a chemical spill, a small-town bribery scandal, or the heavy-handed arrest of a criminal. Players first pick a target: in two cases, they can attack either big corporations or the government, and in one case either the police or violent criminals. They try to give voice to the story by talking about it on their news site’s Twitter account from their chosen perspective, but this fails. They are asked to make the story look bigger than it is by writing an article about it or by posting a meme. This gets them more followers, as people are beginning to pick up on the story. Next, the game asks players if they want to purchase Twitter bots that can amplify the story for them. If they repeatedly refuse, the game ends, but if they accept, they gain 4,000 robot followers. They are shown examples of bots amplifying their chosen story. Their target determines if they are polarizing their chosen topic toward the left (by focusing on big corporations or police brutality) or the right (by focusing on the government or crime-related issues). The key lesson is that it doesn’t matter what side they ultimately choose: the aim is simply to polarize.

Conspiracy

Conspiracy theories are part and parcel of fringe online news sites. Conspiracies can be defined as the belief that unexplained events are orchestrated by a covert group or organization with sinister intentions (Goertzel, 1994; Lewandowsky, Gignac, & Oberauer, 2013; van der Linden, 2015).

In this badge, players are first encouraged to come up with an interesting new theory and post it on their news site. However, since all options are overtly ridiculous (e.g., public schools no longer teach cursive writing so that people stop reading the Communist Manifesto), their theory is too far removed from reality to be believable. Some followers call the player out for their strange theory. To save their credibility, players then look for a more believable conspiracy. They can either choose between Agenda 21, a non-binding United Nations treaty on sustainable development, or the so-called vaccine conspiracy (the idea that the World Health Organization uses vaccinations to indoctrinate people). Players score points if they cast doubt on the official narrative and ask questions that point people in the direction of conspiratorial thinking, and lose points for going

off the rails by publishing content that is considered too weird. Followers react more positively this time, and the player is encouraged to write a serious news article about their topic of choice. If they do well, they gain a cult following, with people trusting their news site more and more and becoming more skeptical of the so-called mainstream media.

Discrediting opponents

When misleading news sites are accused of bad journalism, they can deflect attention away from the accusation by attacking the source of the criticism (“you are fake news!”; see van der Linden, Panagopoulos, & Roozenbeek, 2020) or denying that the problem exists (A’Beckett, 2013; Lischka, 2017). In this badge, players are confronted with a fact checker who debunks the conspiracy theory from the previous badge. They are given three options: either apologize, do nothing, or take revenge. The first option costs them points, and the game moderator explains that apologizing is never a good idea. “Do nothing” prompts a response from one of their news site’s followers asking why they are not responding to the fact check. Eventually, all three choices lead to the same point where players have to choose between either denying the allegations or attacking the fact checker. Their vehement denial or ruthless ad hominem attack on the fact checker triggers a supportive response in the player’s followers, and their reputation remains intact.

Trolling

Trolling is a fishing term, originally referring to the process of slowly dragging a lure or baited hook from the back of a fishing boat. On the internet, it means deliberately evoking a response by using bait (Griffiths, 2014; McCosker, 2014; Thacker & Griffiths, 2012).

In this badge, players put together the techniques they learned in the other five badges. This time, they can only choose one topic. At the beginning of the badge, they are asked to talk about one of three topics (the 25 most romantic cities in Europe; a passenger plane crash; and a newly discovered species of starfish), of which only the second one leads to the next stage. Choosing one of the other two provokes a scolding from the game’s moderator. After this, players are given two options: either pay respects to the victims of the plane crash or start sowing doubt about its cause. The first option prompts a response from their followers asking why they are not investigating the story in more detail. Both options eventually lead to the player to ask whether the crash was a cover-up. Due to their higher credibility and number of followers, their post attracts the attention of other news sites as well, and the story begins to escalate. Players can then throw fuel onto the fire by either impersonating a crash victim’s family member or photoshopping evidence of a cover-up. Both choices then lead to even more emotional responses, and now the mainstream media is also beginning to weigh

in on the story. Players are instructed to keep increasing the pressure, either by discrediting the investigation further or by using another army of Twitter bots to spread the hashtag #InvestigateNow. Depending on their initial choice between impersonating a victim's family member or photoshopping evidence, they can then deliver the final blow by fabricating another news article about the crash. The Aviation Disaster Committee, the (fictional) agency responsible for the investigation, then responds to the manufactured controversy on Twitter. Players then attack this response either by calling for the resignation of the chairman of the Committee or by using the Twitter bot army again. The game ends with the Committee chairman resigning over the handling of the investigation.

Launch and survey results

Following its launch in February 2018, international media around the world covered the game's release both online and in print (BBC, 2017; Reuters, 2018; CNN, 2019). The game included a voluntary pre-post survey, which tested people's ability to recognize disinformation techniques. Over the course of two months, hundreds of thousands of people played the game (mostly from the UK and US). In total, about $N = 15,000$ people opted in for scientific research and completed all of the (pre-post) survey tests. Each survey question came in the form of a fabricated tweet that represented a specific disinformation strategy. Specifically, as an initial evaluation, we tested participants' ability to recognize impersonation (by way of an account impersonating HBO saying that "The 8th season of Game of Thrones will be postponed due to a salary dispute"), conspiracy (a tweet stating that "The Bitcoin exchange rate is being manipulated by a small group of rich bankers"), discrediting opponents (another tweet claiming that "The mainstream media has been caught in so many lies that it can't be trusted as a reliable news source") and polarization (we showed participants an invented news headline that was randomized to state either that a "New study shows that left-wing people lie far more than right-wing people" or the reverse "New study shows that right-wing people lie far more than left-wing people"). Participants were asked to rate the reliability of each of these tweets on a scale between one and seven, plus two "real" control tweets that did not contain any disinformation strategies (e.g., #Brexit, the United Kingdom's exit from the European Union, will officially happen in 2019). We used fictional headlines inspired by "real" fake news for two key reasons; namely (1) to be able to isolate the specific disinformation techniques and (2) to avoid familiarity confounds with real "fake" content (Roozenbeek & van der Linden, 2019). An example of the testing environment is provided in Figure 9.6.

The results are displayed in Figures 9.7 and 9.8. In Figure 9.7, the leftmost bar of the pair is the response people gave before playing (light gray), and the bar on the right is from after playing (dark gray). Some initial concerns about the game were that it could simply make players more skeptical about news media across the board. This is not what we found. "Control_1" and "Control_2"

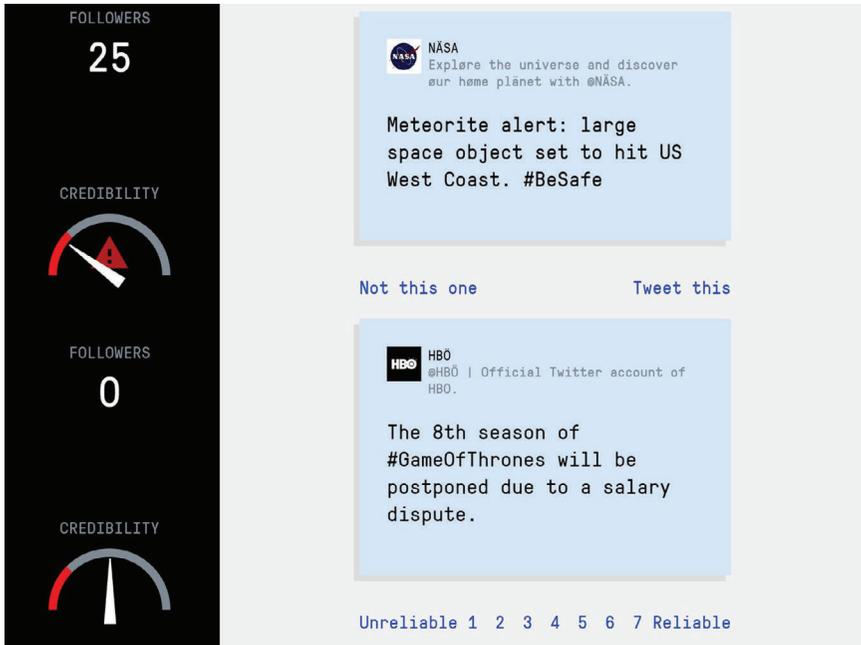


FIGURE 9.6 Bad News game survey

Notes: The top panel illustrates how a technique [impersonation] is used in the game, and the bottom panel shows how the same technique is used in a different example on which participants were evaluated before and after playing.

represent the control questions. We expected these to be rated quite reliably by participants, both before and after playing. This is also borne out by the results: in both cases, the pre- and the post-measure are almost identical and rated highly reliable. The other questions represent techniques commonly used in disinformation that also appear in the game: impersonation, conspiracy, and discrediting opponents. The figure shows that participants rated the “fake news” questions significantly lower after playing the game, indicating a significant inoculation effect for impersonation ($d = .36$), conspiracy ($d = .35$), deflection ($d = .30$), and polarization ($d = 0.16$). These effect sizes are in line with resistance to persuasion research (Walter & Murphy, 2018) and can be considered meaningful, especially when scaled across a population (Funder & Ozer, 2019). Importantly, the learning effects did not differ significantly by political ideology, age, gender, or education (Figure 9.8) and were greater for those who proved most susceptible on the pre-test, bolstering the potential for broad-spectrum immunization. In other words, by actively inoculating people against the strategies that underpin the creation of fake news – through weakened exposure – broad-scale resistance against misinformation can be conferred.

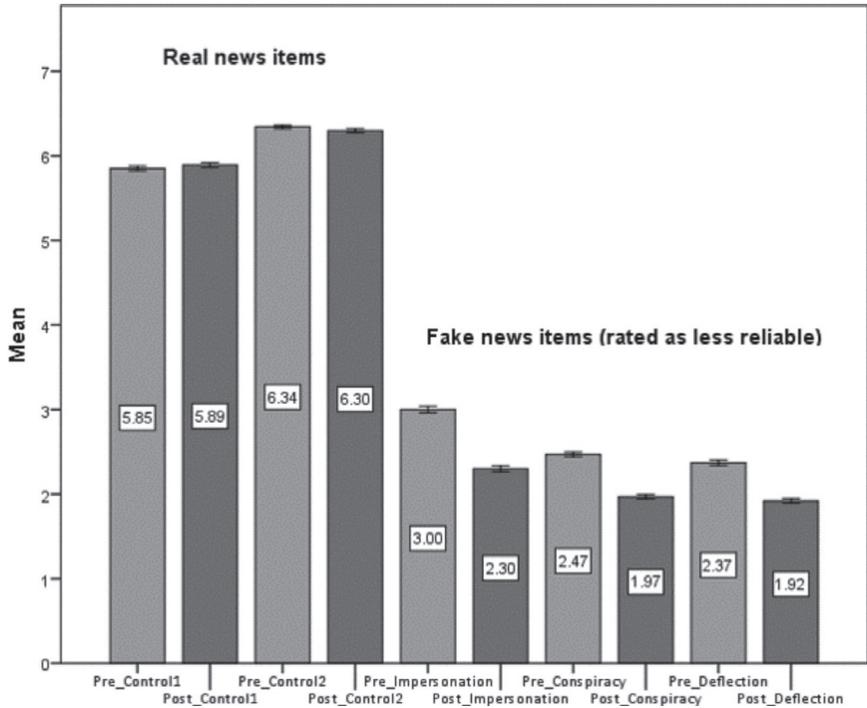


FIGURE 9.7 Results from the Bad News game

Note: Paired bars represent judgments before (light grey) and after (dark grey) playing the game. Participants rated the reliability of all real (control) and fake news items on a 1–7 scale

Limitations and conclusion

In this chapter, we have looked at why disinformation is a problem and what types of solutions are being explored to combat it. In our view, a large part of the solution lies in empowering individuals with evidence-based tools from psychology and behavioral science. We have argued that it is especially important to focus on *preventing* disinformation from going viral in the first place. In fact, the spread of fake news can be modelled much like the spread of a viral contagion. As such, inoculation theory offers an intuitive framework to help develop broad-spectrum immunization against fake news.

In three studies, we have shown how the inoculation metaphor can be extended from the realm of cultural truisms to contested issues, and how we can move from narrow-spectrum vaccines targeting single instances of fake news to a broad-spectrum approach that inoculates people against the very strategies that underlie the production of most disinformation. Of course, these studies are not without their limitations. First, although the samples were diverse, they were either based on students, Mturkers, or a large but self-selected convenience

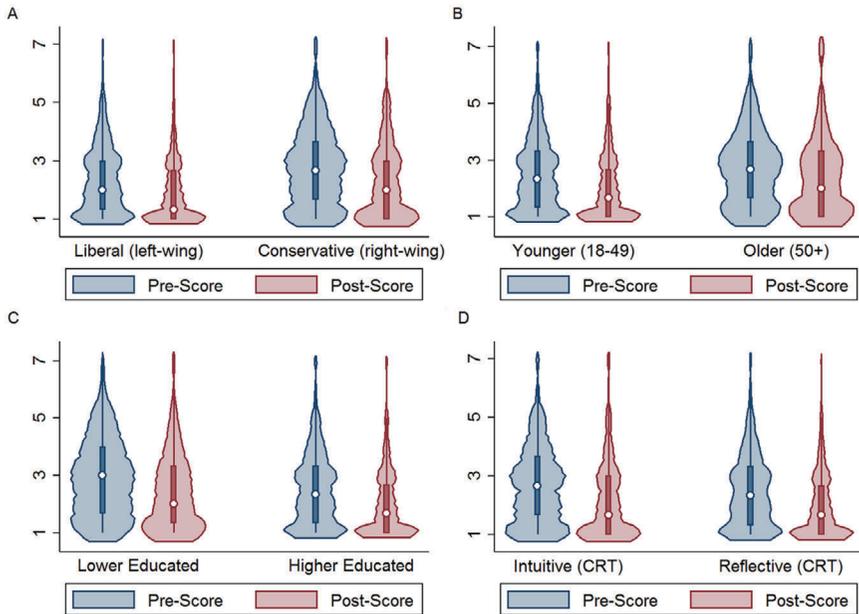


FIGURE 9.8 Violin plots adopted from Roozenbeek & van der Linden (2019) displaying mean pre-post fake news reliability judgments (aggregated) by political ideology (A), age (B), education (C), and a 1-item cognitive reflection measure (D)

sample. Accordingly, these results are not representative of the population. In addition, study 3 used single-item measures and lacked a randomized control group. As such, its causal effects should therefore be interpreted with caution. In addition, although the inoculation treatments proved effective across the ideological spectrum, participants in the game could still branch scenarios in a manner relatively congenial to their ideology. Nonetheless, by documenting the translational process of how to move from a theoretical finding in the lab to a real-world intervention, we highlight the educational potential of novel interactive game-based interventions. The value of intervention science can also be seen in many of its applications. For example, in partnership with the UK government, we have translated the game into 12 new languages, which allows for large-scale cross-cultural evaluations of the game's effectiveness (Roozenbeek, van der Linden, & Nygren, 2020). We have also partnered with WhatsApp to create a special version of the game to help inoculate people against the spread of fake news on direct messaging platforms. We are working with the U.S. government and Behavioral Insights Team in the UK and Lebanon to conduct larger randomized trials to continue to evaluate and improve the success of the intervention. In short, if the new science of *prebunking* is as effective as it appears, there are many open and important questions that future social and behavioral science research can answer to help cultivate societal immunity against fake news.

Note

- 1 Though later research has not always found this (see Banas & Rains, 2010), possibly due to the higher cognitive load associated with participants' having to generate their own counterarguments.

References

- A'Beckett, L. (2013). Strategies to discredit opponents: Russian representations of events in countries of the former Soviet Union. *Psychology of Language and Communication*, 17(2). <https://doi.org/10.2478/plc-2013-0009>
- Aday, S. (2010). Leading the charge: Media, elites, and the use of emotion in stimulating rally effects in wartime. *Journal of Communication*, 60(3), 440–465. <https://doi.org/10.1111/j.1460-2466.2010.01489.x>
- Adl-Tabatabai, S. (2016). Australia to forcibly vaccinate citizens via Chemtrails. Retrieved August 8, 2018, from <https://yournewswire.com/australia-to-forcibly-vaccinate-citizens-via-chemtrails/>
- Bakir, V., & McStay, A. (2017). Fake news and the economy of emotions: Problems, causes, solutions. *Digital Journalism*, 1–22. <https://doi.org/10.1080/21670811.2017.1345645>
- 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>
- Barthel, M., Mitchell, A., & Holcomb, J. (2016, December). Many Americans believe fake news is sowing confusion. *Pew Research Center*. Retrieved from www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/
- Baxter, D. (2018). Melania Trump bans white house staff from taking flu shot. Retrieved August 8, 2018, from <http://archive.is/F0vor>
- BBC. (2017). Cambridge scientists consider fake news “vaccine”. Retrieved August 29, 2017, from www.bbc.co.uk/news/uk-38714404
- BBC News. (2018a, July 19). How WhatsApp helped turn an Indian village into a lynch mob. Retrieved from www.bbc.co.uk/news/world-asia-india-44856910
- BBC News. (2018b). Government announces anti-fake news unit. Retrieved August 8, 2018, from www.bbc.co.uk/news/uk-politics-42791218
- BBC News. (2018c, August 28). A fake billionaire is fooling people on Twitter. Retrieved from www.bbc.co.uk/news/world-us-canada-45331781
- Bertolin, G., Agarwal, N., Bandeli, K., Biteniece, N., & Sedova, K. (2017). Digital Hydra: Security implications of false information online. Retrieved from www.stratcomcoe.org/digital-hydra-security-implications-false-information-online
- Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., . . . Quattrociocchi, W. (2016). Users polarization on Facebook and Youtube. *PLoS One*, 11(8), 1–24. <https://doi.org/10.1371/journal.pone.0159641>
- Bremner, C. (2018). France aims to ban fake news at election time. Retrieved from www.thetimes.co.uk/article/france-aims-to-ban-fake-news-at-election-time-jwspzjx83
- Calfas, J. (2017). Google is changing its search algorithm to combat fake news. Retrieved August 8, 2018, from <http://fortune.com/2017/04/25/google-search-algorithm-fake-news/>
- Channel 4. (2017). C4 study reveals only 4% surveyed can identify true or fake news. Retrieved August 29, 2017, from www.channel4.com/info/press/news/c4-study-reveals-only-4-surveyed-can-identify-true-or-fake-news

- CNN. (2019). Researchers have created a “vaccine” for fake news: It’s a game. Retrieved from https://edition.cnn.com/2019/07/04/media/fake-news-game-vaccine/index.html?utm_medium=social&utm_content=2019-07-04T13%3A46%3A03&utm_source=twCNN&utm_term=link
- Compton, J. (2013). Inoculation theory. In J. P. Dillard & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (2nd ed., pp. 220–236). Thousand Oaks: Sage Publications. <https://doi.org/10.4135/9781452218410>
- Compton, J. (2019). Prophylactic versus therapeutic inoculation treatments for resistance to influence. *Communication Theory*, qtz004.
- Compton, J., Jackson, B., & Dimmock, J. A. (2016). Persuading others to avoid persuasion: Inoculation theory and resistant health attitudes. *Frontiers in Psychology*, 7, 122.
- Cook, J., Lewandowsky, S., & Ecker, U. K. (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>
- Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., . . . Rice, K. (2016). Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, 11(4), 48002. Retrieved from <http://stacks.iop.org/1748-9326/11/i=4/a=048002>
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., . . . Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554–559. <https://doi.org/10.1073/pnas.1517441113>
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt Brace Jovanovich.
- Elgin, B., & Wang, S. (2018). Facebook’s battle against fake news Notches an Uneven scorecard. Retrieved August 8, 2018, from www.bloomberg.com/news/articles/2018-04-24/facebook-s-battle-against-fake-news-notches-an-uneven-scorecard
- Eurobarometer (2018). *Final results of the Eurobarometer on fake news and online disinformation*. Directorate-General CONNECT, European Commission. Retrieved from <https://ec.europa.eu/digital-single-market/en/news/final-results-eurobarometer-fake-news-and-online-disinformation>
- European Commission. (2018). Final report of the high level expert group on fake news and online disinformation. Retrieved from <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>
- FastCompany. (2018). Innovation by design awards: University of Cambridge/DROG. Retrieved from www.fastcompany.com/innovation-by-design/2018/company/university-of-cambridge-drog
- Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*, 22(8). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2995809
- 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.
- Goertzel, T. (1994). Belief in conspiracy theories. *Political Psychology*, 15(4), 731–742. <https://doi.org/10.2307/3791630>
- Goga, O., Venkatadri, G., & Gummedi, K. P. (2015). The doppelgänger bot attack: Exploring identity impersonation in online social networks. In *Proceedings of the 2015 internet measurement conference* (pp. 141–153). New York, NY, USA: ACM. <https://doi.org/10.1145/2815675.2815699>
- Griffiths, M. D. (2014). Adolescent trolling in online environments: A brief overview. *Education and Health*, 32(3), 85–87. Retrieved from <http://irep.ntu.ac.uk/id/eprint/25950/>

- Groenendyk, E. (2018). Competing motives in a polarized electorate: Political responsiveness, identity defensiveness, and the rise of partisan antipathy. *Political Psychology*, 39, 159–171. <https://doi.org/10.1111/pops.12481>
- Gross, K., & D'Ambrosio, L. (2004). Framing emotional response. *Political Psychology*, 25(1), 1–29. Retrieved from www.jstor.org/stable/3792521
- Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. (1949). *Experiments on mass communication*. (Studies in social psychology in World War II) (Vol. 3). Princeton, NJ: Princeton University Press.
- Huddy, L., Sears, D. O., & Levy, J. S. (2013). *The Oxford handbook of political psychology*. Oxford: Oxford University Press.
- Konijn, E. A. (2013). The role of emotion in media use and effects. *The Oxford Handbook of Media Psychology*, 186–211.
- Kragh, M., & Åsberg, S. (2017). Russia's strategy for influence through public diplomacy and active measures: The Swedish case. *Journal of Strategic Studies*, 40(6), 773–816. <https://doi.org/10.1080/01402390.2016.1273830>
- Kurtzleben, D. (2016). Do fact checks matter? Retrieved August 8, 2018, from www.npr.org/2016/09/27/495233627/do-fact-checks-matter
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., . . . Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Le Pen, M. (2018, March 19). Twitter post. Retrieved August 30, 2018, from https://twitter.com/mlp_officiel/status/975808872802856960?lang=en
- 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*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). The role of conspiracist ideation and worldviews in predicting rejection of science. *PLoS One*, 8(10), 1–11. <https://doi.org/10.1371/journal.pone.0075637>
- Lischka, J. A. (2017). A badge of honor? How the New York Times discredits President Trump's fake news accusations. *Journalism Studies*, 1–18. <https://doi.org/10.1080/1461670X.2017.1375385>
- McCosker, A. (2014). Trolling as provocation: YouTube's agonistic publics. *Convergence*, 20(2), 201–217. <https://doi.org/10.1177/1354856513501413>
- McGuire, W. J. (1970). A vaccine for brainwash. *Psychology Today*, 3(9), 36–64.
- McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *Journal of Abnormal and Social Psychology*, 62, 327–337.
- 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>
- Melki, M., & Pickering, A. (2014). Ideological polarization and the media. *Economics Letters*, 125(1), 36–39. <https://doi.org/10.1016/j.econlet.2014.08.008>
- Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2019). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability. *Political Behavior*, 1–22.
- Orita, A., & Hada, H. (2009). Is that really you? An approach to assure identity without revealing real-name online. In *Proceedings of the 5th ACM workshop on*

- digital identity management* (pp. 17–20). New York, NY, USA: ACM. <https://doi.org/10.1145/1655028.1655034>
- Papageorgis, D., & McGuire, W. J. (1961). The generality of immunity to persuasion produced by pre-exposure to weakened counterarguments. *Journal of Abnormal and Social Psychology*, 62, 475–481.
- Patriot United. (2018). Muslim doctor refuses to save a Christian man's life on flight from New York to Las Vegas. Retrieved August 8, 2018, from <https://web.archive.org/web/20180108183412/http://patriotunited.club/2018/01/03/muslim-doctor-refuses-to-save-a-christian-mans-life-on-flight-from-new-york-to-las-vegas/>
- Pfau, M., Park, D., Holbert, R. L., & Cho, J. (2001). The effects of party- and PAC-sponsored issue advertising and the potential of inoculation to combat its impact on the democratic process. *American Behavioral Scientist*, 44(12), 2379–2397.
- Pieters, J. (2018). Dutch politicians want EU anti-fake news watchdog scrapped. Retrieved August 8, 2018, from <https://nltimes.nl/2018/03/06/dutch-politicians-want-eu-anti-fake-news-watchdog-scrapped>
- Poland, G. A., & Spier, R. (2010). Fear, misinformation, and innumerates: How the Wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine*, 28(12), 2361–2362.
- Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16(1), 101–127. <https://doi.org/10.1146/annurev-polisci-100711-135242>
- Readfearn, G. (2016). Revealed: Most popular climate story on social media told half a million people the science was a Hoax. Retrieved August 8, 2018, from www.desmogblog.com/2016/11/29/revealed-most-popular-climate-story-social-media-told-half-million-people-science-was-hoax
- Reuters. (2018). Fake news “vaccine” teaches you to spot disinformation. Retrieved January 15, 2019, from <https://uk.reuters.com/video/2018/03/20/fake-news-vaccine-teaches-you-to-spot-di?videoId=410596269>
- Reznik, M. (2013). Identity theft on social networking sites: Developing issues of internet impersonation. *Touro Law Review*, 29(2), 12.
- Roozenbeek, J., & van der Linden, S. (2018). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*, 22(5), 570–580. <https://doi.org/10.1080/13669877.2018.1443491>
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Nature Palgrave Communications*, 5(65). <https://doi.org/10.1057/s41599-019-027>
- Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on “inoculation” theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*, 1(2). <https://doi.org/10.37016/mr-2020-008>
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, 9(1), 4787.
- Snopes. (2018). Did “Muslim migrants” attack a Catholic Church during mass in France? Retrieved August 30, 2018, from www.snopes.com/fact-check/muslim-migrants-attack-catholic-church-mass-france/
- Tandoc, E. C., Jr., Lim, Z. W., & Ling, R. (2018). Defining “fake news” a typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153.
- Thacker, S., & Griffiths, M. D. (2012). An exploratory study of trolling in online video gaming. *International Journal of Cyber Behavior, Psychology and Learning*, 2(4). <https://doi.org/10.4018/ijcbpl.2012100102>

- Twenge, J. M., Honeycutt, N., Prislin, R., & Sherman, R. A. (2016). More polarized but more independent: Political party identification and ideological self-categorization among U.S. adults, college students, and late adolescents, 1970–2015. *Personality and Social Psychology Bulletin*, 42(10), 1364–1383. <https://doi.org/10.1177/0146167216660058>
- van der Linden, S. (2015). The conspiracy-effect: Exposure to conspiracy theories (about global warming) decreases pro-social behavior and science acceptance. *Personality and Individual Differences*, 87, 171–173. Retrieved from www.sciencedirect.com/science/article/pii/S0191886915005024
- van der Linden, S. (2017). Beating the hell out of fake news. *Ethical Record: Proceedings of the Conway Hall Ethical Society*, 122(6), 4–7. Retrieved from <https://ssrn.com/abstract=3089590>
- 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., Maibach, E., Cook, J., Leiserowitz, A., & Lewandowsky, S. (2017). Inoculating against misinformation. *Science*, 358(6367), 1141–1142. <https://doi.org/10.1126/science.aar4533>
- 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.
- van der Linden, S., Roozenbeek, J., Oosterwoud, R., Compton, J., & Lewandowsky, S. (2018). *The science of prebunking: Inoculating the public against fake news*. Written evidence submitted to the Parliamentary Inquiry on Fake News. House of Commons, UK Parliament, London: UK. Retrieved from <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/fake-news/written/79482.html>
- Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). *Online human-bot interactions: Detection, estimation, and characterization*. Proceedings of the Eleventh International AAAI Conference on Web and Social Media. Retrieved from www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/viewPaper/15587
- 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>
- Wakabayashi, D. (2017). As Google fights fake news, voices on the margins raise alarm. Retrieved February 15, 2018, from www.nytimes.com/2017/09/26/technology/google-search-bias-claims.html
- Wakefield, J. (2017). Facebook's fake news experiment backfires. Retrieved August 8, 2018, from www.bbc.co.uk/news/technology-41900877
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423–441.